D'Aloisio, V, Dognini, P, Hutcheon, GA and Coxon, CR

PepTherDia: database and structural composition analysis of approved peptide therapeutics and diagnostics.

Article

PepTherDia: database and structural composition analysis of approved peptide therapeutics and diagnostics (2020)

Vera D'Aloisio[1], Paolo Dognini[1], Gillian A. Hutcheon[1], and Christopher R. Coxon[2*]

[1]School of Pharmacy and Biomolecular Sciences, Faculty of Science, Liverpool John Moores University, Liverpool, United Kingdom, L3 3AF

[2]Institute of Chemical Sciences, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, United Kingdom, EH14 4AS

*** Corresponding author:**

Christopher R. Coxon (email: c.coxon@hw.ac.uk; Tel: +44 131 451 8029)

Keywords: peptide therapeutics and diagnostics, peptide drug, peptide market, structural analysis, approved peptides database

Teaser: We describe a freely-accessible database of approved peptide therapeutics and diagnostics, providing an overview of some key structural and compositional trends to help guide the design of future peptide medicines.

Abstract

As of 2020, there are over 100 approved peptides with therapeutic or diagnostic applications. However, a complete database providing information on marketed peptides is not freely available, making the peptide chemists' job of designing future peptide drug candidates challenging. Unlike the rules for small molecule drugs, there is no general set of guidelines for designing a successful peptide-based drug. This foundation review, together with our freely available database (PepTherDia, http://peptherdia.herokuapp.com), provides some insights into what a successful peptide therapeutic or diagnostic agent looks like and will lay the foundation for establishing a set of rules to help future medicinal chemists to design peptide candidates with increased approval rates.

Introduction

Following the approval of the first peptide therapeutic agent, the 51-amino acid hormone insulin in 1923 [1,2], drug discovery has been progressively expanding into the chemical space between small molecules and large proteins. Subsequently, a significant number of peptides (and peptidomimetics) have received regulatory approval. Recently, peptides have emerged as novel modalities for various applications in the therapeutic and diagnostic markets, providing new opportunities for the modulation of difficult targets. As shown in Figure 1, since the second half of the last century, the number of peptides on the therapeutics and diagnostics market has steadily increased, reaching the milestone of over 100 approved peptide drugs in 2020. These drugs represent a unique class of chemical compounds that bridges the gap between small molecules (typically molar mass < 500 g/mol) and large biologics (typically molar mass > 5000 g/mol). Occupying an intermediate region of complexity and molar mass, they combine many of the benefits of the two abovementioned categories. The main disease areas presently treated with peptide drugs are metabolic disorders, cancer and cardiovascular disease, but other emerging therapeutic applications are in the areas of infectious diseases, pain, urinary tract, gastrointestinal and respiratory disorders [1,3]. As the incidence of metabolic disorders, cancer and cardiovascular diseases in the Western world is increasing alarmingly [4–6] and the need for new effective medicines to treat emerging health problems (e.g., SARS-CoV-2/COVID-19) is growing, it is highly likely that the demand for peptide drugs will continue to grow.

Current state of peptide therapeutics and diagnostics

According to Transparency Market Research's latest report, the worldwide market for peptide pharmaceuticals has been growing at a Compound Annual Growth Rate (CAGR) of roughly 8%; this is expected to increase over time with the same trend, reaching a value of 50 billion USD in 2027 from 25 billion USD in 2018 [7].

As of November 2020, according to our database analysis, there are 105 peptide pharmaceutical products (see definitions below) with regulatory approval in the main pharmaceutical markets - North America, Europe and Japan – of which 89 are peptide drugs and 16 are diagnostic agents. Moreover, nowadays, a great number of clinical studies involve peptide agents – 4,859 in total, 468 of which are in phase 3 [8]. This suggests that the pharmaceutical industry is committed to exploring the role of peptide therapeutics on modulating previously 'undruggable' targets and addressing unmet medical needs. In Table 1, the main advantages of peptides over small molecules and proteins are illustrated. Despite these important benefits, the drug development process for future therapeutic peptides from laboratory to approval has traditionally presented many obstacles. In fact, unique challenges, such as chemical and physical instability, short circulating half-life, high proteolytic

degradation, rapid renal clearance, poor membrane permeability, poor oral bioavailability, and low solubility must be addressed in order to bring a peptide to clinical use [1,9]. Nonetheless, lately, advances in drug delivery and emerging medicinal chemistry strategies have brought peptides to a significant renaissance, by overcoming their issues and eventually improving pharmacokinetic profiles and oral bioavailability.

In the last decade, several research groups have tried to reach a better understanding of approved peptides and their properties. In 2010, Vlieghe *et al* reviewed and listed the synthetic therapeutic peptides that have reached the main pharmaceutical markets (US, EU, and Japan) [10]. Six years later, Raghava and collaborators filed a repository (PEPlife) to provide the scientific community with data on peptide half-lives [11], followed, one year later, by a database containing 852 FDA-approved biologics, among which 28 were peptides [12]. Recently, an extensive review on approved peptide therapeutics targeting G protein-coupled receptors (GPCRs) have been published by Davenport and collaborators [13], demonstrating their dominant presence in the market.

Nevertheless, to the best of the authors' knowledge, up to the present moment (November 2020), a complete database with structural analysis, production methods, pharmacokinetic properties (i.e., terminal half-life and protein binding), indications, and routes of administration of regulatory-approved peptides is not freely available online. In fact, the information is scattered throughout the scientific literature and in various websites, making the search for approved peptides very challenging. In addition to this, there is a critical lack of rules that makes the medicinal chemists' job of designing entirely new potential peptide drugs very difficult. In fact, as already pointed out by Tyagi and co. [14], the "drug-likeness" criteria used for small molecules – principally, *Lipinski's Rule of Five, Ro5* [15], but also models such as the central nervous system multiparameter optimization (CNS MPO) [16] - are not applicable to peptides due to their entirely different intrinsic properties and applications.

PepTherDia

In this context, we have developed and made accessible online **PepTherDia** (**Pep**tide **Ther**apeutics and **Dia**gnostics: http://peptherdia.herokuapp.com), a manually curated database containing a searchable list of approved peptide drugs and diagnostic agents, with information on their physicochemical and pharmacokinetic properties, as well as their routes of administration and indications. Its purpose is to provide assistance to medicinal chemists and scientists in the field of peptide drug discovery. On the compounds enumerated and described in PepTherDia, we have performed a detailed analysis of the structural features and collected information on their terminal half-life, plasma protein-binding, indication, route of administration, production methodologies, marketing authorisation (year and agency of first approval) and origin of their design. The information contained on this database will be updated on a regular basis (e.g., yearly) with new approvals as well as new properties investigated. For the first time, this review highlights some important trends in peptide approvals, together with providing the reader with an insight into the features and characteristics that are common in approved peptide agents. We envisage that this information will aid the scientific community to more successfully design or pre-screen candidates at an early stage of the peptide drug discovery process to increase longitudinal approval rates.

How is a "peptide" defined?

A challenge in designing this study was the breadth and diversity in the so-called 'peptide' molecule due to differences in structure, size and composition. Therefore, the first thing the authors felt necessary to clarify was: what is a peptide? The International Union of Pure and Applied Chemistry (IUPAC) defines peptides as "amides derived from two or more amino carboxylic acid molecules (the same or different) by formation of a covalent bond from the carbonyl carbon of one to the nitrogen atom of another with formal loss of water" [17]. Whereas the recommendations of the IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) defines a peptide as a chemical entity presenting from 2 to 50 amino acid residues [18], the currently used regulatory FDA definition delineates a peptide as "any polymer composed of 40 or fewer amino acids" and regardless their production method [19,20]. Finally, the European Medicines Agency (EMA), instead of describing peptides based on their size, considers them as small molecules if chemically synthesised, while treats them as biological entities if they are extracted from natural sources or produced with recombinant methodologies [21], which underlines that the definition of peptide is still ambiguous.

At this stage, the reader might state that this definition can be considered as a philosophical debate that may elicit controversial answers. Indeed, the scientific community differs greatly on where to stop using the term peptide and start using the term protein. In the glossary section of the "Glossary and Key Sources Box", the definition of peptide that led our research, together with other terminologies are explained.

Data collection and calculations

A repository of 105 compounds was obtained by searching in DrugBank [22], FDA and EMA web pages [19,23], Pharmaceutical and Medical Devices Agency website [24], and Drug Central website [25]. The key inclusion and exclusion criteria used in the curation of PepTherDia are listed in Table 2. Examples of peptides not included and the reason for their exclusion are shown in Table 3. With the aim of providing each peptide with a complete profile comprising relevant information regarding terminal half-life, protein binding, therapeutic indications, and routes of administration, specific searches were carried out in DrugBank [22], National Centre for Advancing Translational Sciences web page [26], Drugs.com [27], and pharmaceutical companies' websites, using the generic name of the individual peptide. References specific to each approved peptide as well as SMILES codes used to calculate the peptide molar mass values can be found on our website PepTherDia.

Structural composition analysis of approved peptide agents

As discussed and defined earlier, peptide medicines are generally composed of natural amino acids, unnatural amino acids and non-amino acidic modifications. Figure 2 provides an example of how the peptide daptomycin may be divided into the above components. However, in some cases, e.g., the glycopeptide antibiotics (dalbavancin, telavancin, oritavancin, and teicoplanin), this is a very complex (if not impossible) task. Similarly, the complexity of some multicyclic peptides does not allow the unambiguous identification of a defined single macrocycle and, therefore, the members of each macrocycle were not counted here.

Each constitutional member can be further classified as polar, acidic, basic, non-polar aliphatic or aromatic based on its structural and physicochemical characteristics. For the natural amino acidic residues, the designations polar, acidic, basic, non-polar aliphatic or aromatic, derived from literature precedent [28], are generally ascribed by the nature of the

side chain. Since non-natural amino acid members form an amide backbone in the same way as natural amino acids, they may be classified following the same principles used for natural amino acids. In contrast, peptide modifications are a broad and varied structural class, and their classification requires consideration of their complete structure and the way they are conjugated to the peptide. For example, in daptomycin (Figure 2), decanoic acid can be classified as a non-polar aliphatic modification as the carboxylic acid moiety becomes part of an amide and its contribution to the final polarity predominantly increases lipophilicity.

The complete list of non-natural amino acids and modifications together with their polarity classification can be found on PepTherDia website.

Molar mass distribution and origin of peptide design

Thanks to the improvements in synthetic and manufacturing technologies, nowadays it is possible to synthesise ever-larger peptides in a short time, yielding high purities and quantities. Nonetheless, from our study it emerged that the majority of approved peptides (68%) are relatively 'small' peptides, composed of 2-16 constitutional members, with a second minor cluster (27%) of larger size peptides with around 28-37 members (Figure 3A). This is mirrored in a bimodal molar mass distribution in the ranges 300-1750 g/mol (major - 71%) and 2750-4250 g/mol (minor - 22%), with a remarkable lack of mid-length approved peptides (Figure 3B). Hence, the data suggest that there are two main groups of peptides: low molar mass and high molar mass with only a few examples in between (e.g., sinapultide and ziconotide, with molar mass of 2469.45 and 2639.14 g/mol, respectively). To evaluate the possibility of bias in the peptide design, which may have led to the deliberate development of peptides of certain sizes to mimic specific biomolecules, it is necessary to analyse the origin of the peptide design case by case (Figure 3E). In this context, it emerges that *natural peptides* account for 30% of our sample. However, *peptide analogues* account for 54%. Finally, *heterologous peptides* account for 16% of our sample. This underlines that heterologous peptides are difficult to design *a priori* and are mostly discovered by library screening. In light of these findings, we can state that, unsurprisingly, there is a clear trend (in 84% of the cases) towards following the route of inspiration from nature as a greater promise of success and that the bimodal distribution may be attributed to the characteristics of the natural molecules that have inspired the design. Examples that demonstrate this are the natural nonapeptide oxytocin and the 32-membered peptide calcitonin. Here, in both cases, their length is due to the size of the natural molecule of origin.

Natural side chain amino acid occurrence

Amongst all the constitutional members, the large majority (around 81%) is represented by natural L-amino acids. The residual 19% comprises non-natural AAs and modifications. A careful analysis of the amino acid residues contained in each approved peptide (Figure 3C) showed that the most common amino acids found in the sequences are the non-polar aliphatic leucine (L) and glycine (G), followed by the polar serine (S). On the contrary, the least common residues are methionine (M), histidine (H) and isoleucine (I). This is largely in agreement with the occurrence of natural amino acids in proteins: in nature, leucine (L) accounts for 9.1%, serine (S) for 6.8%, glycine (G) for 7.2%, and alanine (A) for 7.8%, being the most common amino acids. In contrast, methionine (M) and histidine (H) account each only for 2.3%, cysteine (C) for 1.9%, and tryptophan (W) for 1.4%, being the least common amino acid residues found in proteins [28]. The occurrence of cysteine (C) in pharmaceutical peptides is higher than in proteins, which is due to the frequent use of disulfide bonds as a tool for macrocyclisation (see section '*Conformational and shape properties'*). In general, it is important to emphasise that the amino acid composition of proteins and peptides is highly

variable; some amino acids may occur only once or not at all in a given peptide and may be repeated several times in another peptide sequence. An example of where this is seen is the repetition of the moiety KL_4 in the peptide sinapultide, designed to mimic the C-terminal domain of the surfactant protein B [29].

D-amino acids with natural side chains account only for a small percentage (around 4% of the total amino acids with natural side chain) and are mainly represented by phenylalanine, alanine tryptophan, and arginine. The selective replacement of L-amino acids by their enantiomers (D-AAs) can protect the molecule from protease degradation [30]. This is a common technique to obtain proteolytic stabilisation by backbone modification, even if this causes conformational changes that might affect biological activity. An example is the somatostatin-like peptide octreotide, in which natural phenylalanine and tryptophan are replaced with their mirror-image forms, leading to a 100-fold increase in the terminal half-life [31,32].

Non-natural amino acids occurrence

In the field of peptide drug discovery, the use of non-natural amino acids as well as the conjugation with non-amino acidic members are common techniques to overcome peptide limitations and have been widely explored in both protein and peptide design [33,34]. In Figure 3D, the most common non-natural amino acids and structural modifications in approved peptides are reported.

Position-specific incorporation of non-natural amino acids bearing a variety of bespoke side chains can provide improvements in peptide properties, activity and function [35]. In this study, non-natural amino acids have emerged to be present in heterologous peptides (e.g., ACE inhibitors, macimorelin, argatroban) and peptide analogues (e.g., GnRH agonists, buserelin, carbetocin, icatibant, pasireotide, carfilzomib) but also in one third of the natural peptides. In fact, from our analysis, the percentage of proteinogenic amino acids in natural peptide antibiotics (e.g., capreomycin, vancomycin, bleomycin, oritavancin), has been estimated to be between 0 and 10% of all the constitutional members, a value that is significantly lower than the average percentage of natural amino acids in the pool of 105 approved peptides (81%).

Ornithine (Orn), 2,4-Diaminobutyric acid (Dab), and 2,3-diaminopropionic acid (Dap) are homologues of lysine (Lys), where the variability in structures is due to the difference in the number of side chain carbons. They all contain an amino group on the side chain, which exhibits a basic/ionisable contribution and, at the same time, allows opportunities for cyclisation or conjugation. Moreover, it has been reported that the use of Dap and Dab in antimicrobial peptides (AMPs) can prevent the hemolytic activity of positively charged natural amino acids (i.e., Arg and Lys), solving a characteristic issue of AMPs [36]. AMPs have also shown improved stability to trypsin while retaining their biological activity, when Lys and Arg residues are replaced by Dab, Dap, or homoarginine [37], demonstrating that Lys homologues are not suitable substrate of this hydrolytic enzyme. Likewise, lower susceptibility to tryptic hydrolysis has been reported after Lys replacement with Orn [38]. In our pool of approved peptides, Dap and Dab are only found in natural peptides – Dap in capreomycin, viomycin, enviomycin, while Dab in colistin and polymyxin B. On the other hand, Orn is used in a variety of peptide analogues (e.g., ornipressin, atosiban, anidulafungin, caspofungin, and micafungin) as well as in natural peptides (e.g., daptomycin and bacitracin). Overall, amino acid side chains bearing a primary amine, such as lysine and its homologues, represent one of the most frequently used amino acid moieties (around 7% of the total amino

acids). Specifically, the Dab residue was encountered 12 times in total, but this amino acid appears only in two peptide structures (colistin and polymyxin B), while Orn appears only four times, each in different peptides (Daptomycin, Ornipressin, Atosiban, Bacitracin).

4-hydroxyproline (Hyp) is a proline containing a hydroxyl group on the pyrrolidine ring. In general, the relatively frequent use of hydroxyproline derivatives (9 in total in the pool of approved peptides, e.g., caspofungin, icatibant, voxilaprevir, to name but a few) may be ascribable to the polarity enhancement, additional hydrogen-bonding ability and the possibility of further conjugation gained with their introduction. Moreover, it has been demonstrated that the presence of Hyp stabilises the triple-helical structure of collagen [39]. Similarly, this stabilisation may occur in peptide secondary structures.

Naphthyl-alanine (Nal) is often used to mimic tryptophan (Trp) and to explore potential improvements in peptide pharmacological profiles [40–42]; however, it is not clear if 1-Nal or 2-Nal adequately replicate the effects of Trp aromatic interactions. In fact, as with any modification, the consequence of these replacements upon the peptide potency needs careful assessment, because it has been demonstrated that substitution of Trp with 1-Nal or 2-Nal decreases the potency of cholecystokinin analogues [43]. Nonetheless, this strategy has been successfully used in the development of GnRH receptor (GnRH-R) antagonists (abarelix, ganirelix, degarelix, and cetrorelix) and other peptides (lanreotide, nafarelin, pralmorelin). Specifically, in GnRH-R blockers, the His – Trp motif of the natural hormone GnRH, has been replaced with the three-amino acid motif 2-Nal – (4-Cl) Phe – 3-Pal, suggesting an intention to improve the aromatic contribution at the peptide N-terminus, which may be important for the binding and the antagonistic activity at the receptor.

Non-amino acidic modifications

An additional common strategy used to improve peptide drug-likeness is the introduction of non-amino acid appendages to tune the pharmacokinetic properties. These are usually linked to the main chain by only one functional group (e.g., -COOH for fatty acids or -OH for sugars) and are generally attached to amino acids containing polar functional groups (e.g., OH, NH_2, COOH) or to the N- or C-termini. Enhancement of stability, protein binding, and membrane permeability can be obtained through peptide lipid acylation, while improved solubility and bioavailability can be achieved through glycosylation [44,45]. Indeed, lipid acylation and glycosylation are the most common modifications encountered in the pool of approved peptides.

Lipid acylation is a post-translational modification of proteins that has found applications in peptide design, in order to improve pharmacokinetic and pharmacodynamic properties while retaining the ability to bind the target receptor [44,46,47]. 13% of marketed peptides present a lipophilic carbon chain attached to their structure and, in some cases, they demonstrated prolonged terminal half-life (e.g., in oritavancin and dalbavancin that present a terminal half-life of 245 and 346 hours, respectively,) and high protein binding (>90% in the majority of lipidated approved peptides). The length of the carbon chain may influence the half-life duration but, at the moment, we do not have enough data around approved lipidated peptides to state so. Other examples of approved peptides presenting a lipophilic carbon chain, include the popular diabetes medicines and GLP-1 receptor agonists, liraglutide (conjugated with palmitic acid) and semaglutide (conjugated with an octadecanedioic acid). They both bind with high affinity to plasma proteins (98-99% of the peptide bound), promoting greater peptide stability that results in significantly extended half-lives of 13 and 168 hours, respectively, compared with the parent GLP-1.

Carbohydrate groups are less frequent but still significant modifications, being found in 8 out of 105 peptides on the market (e.g., bleomycin and vancomycin). Given the synthetic challenges that glycochemistry presents, these are typically found in peptides of natural origin. In these peptides, glycosyl units are attached to the main structure via an N-terminal amine group or hydroxyl group on the side chain, similar to recombinant glycoprotein therapeutics, in which carbohydrates are commonly N-linked to asparagine or O-linked to serine and threonine [48]. It is probably not by chance that these 8 examples comprise a large aromatic core or aliphatic chain, in which hydrophobicity is balanced by one or more sugars. In fact, glycosylation improves the physicochemical and pharmacokinetics properties of peptide drugs through an enhancement of solubility and an increase in bioavailability and oral absorption [21,45].

Other modifications can include metal cation-chelating agents (DOTA in dotatate and DTPA in pentetreotide), typically found in diagnostic agents or linkers (2-amino-4,6-dimethyl-3-oxo-3H-phenoxazine-1,9-dicarbonyl in dactinomycin). Finally, pyroglutamic acid – cyclic lactam of glutamic acid – is naturally found at the N-terminus of many neuronal peptides and hormones but its function in living cells is still unclear [49]. In drug design, N-pyroglutamyl formation is a common modification used to cap the N-terminus in order to modulate peptide activity and increase resistance to degradation [50]. Indeed, in some cases, Pyr has shown to be essential in order to achieve full biological activity [51]. This modification is found in 8 approved peptides, such as leuprolide and all GnRH agonists.

Notably, voxilaprevir is the first example of fluorinated peptide on the market, bearing four fluorine atoms: a difluoromethylene adjacent to a benzopyrazine modification involved in forming a macrocyclic structure, and a difluoromethyl group on an aliphatic non-natural amino acid – 1-amino-2-(difluoromethyl)cyclopropane-1-carboxylic acid. We predict that the exploitation of fluorine in peptides will follow what has already happened in small molecules, by becoming a key medicinal chemistry tool, in which the judicious addition of a small and highly electron withdrawing atom such as fluorine has been shown to play a key role in improving pharmacokinetic and physicochemical properties [52]. Hence, in the current peptide drug discovery pipeline, fluorination has been found to increase thermal stability and proteolytic stability, without affecting biological activity [53]. This has been applied to glucagon-like peptide-1 (GLP-1), in which the substitution with hexafluoroleucine in different positions has been shown to improve both binding affinity and protease (DPP IV) stability [54].

Polarity trends

Further examination of the peptide structures reveals that there is on average a balance between polar and hydrophobic residues, when the polar contribution is derived by the summation of the polar, basic and acidic constitutional members and the hydrophobic contribution is the summation of the aromatic and aliphatic constitutional members. Figure 4A shows that the majority of the approved peptides contain from 35 to 75% of polar residues, indicating that these molecules do not present a high excess of either hydrophilic or lipophilic components. This perhaps should not be a surprise, given that both polar and hydrophobic components are generally required for drugs with good pharmacokinetic profiles. A small number of outliers that comprise 100% polar or 100% hydrophobic building blocks are present. However, these exceptions are generally represented by a very small number of building blocks (2, 3 or 5) and their small size makes these peptides more similar to small molecules. Indeed, the hydrophilic dipeptide spaglumic acid as well as the hydrophobic tripeptides ACE inhibitors (enalapril, perindopril, ramipril, quinapril, and

trandolapril) respect the Lipinski's rule of five tailored for small molecules (computed by ChemAxon, Chemicalize [55]). Other outliers are represented by the growth hormone secretagogue receptor agonist macimorelin composed of only 3 building blocks, and the hydrophobic antiviral peptides telaprevir, boceprevir, and ombitasvir, composed of 4 or 5 members. Finally, cyclosporine is the only case in which a marketed peptide composed by more than 5 members (11 in this case) completely lacks the polar component. This is reflected in its very long half-life (19 hours) [56] and the fact that, after administration, 90% is found to be bound to serum proteins, mainly lipoproteins [57]. In this respect, it must be underlined that the classification of natural and non-natural amino acids is solely based on the nature of their side chain. Assuming that the backbone impact on the final properties is largely consistent with the size of the peptide, its polar contribution has not been considered. This is the clearest way to distinguish between hydrophilic and hydrophobic amino acids, without overshadowing the contribution of the side chain with that of the backbone. To provide the reader with a visual overview of the relative physical property balance, the colour map in Figure 4B shows the % composition of each approved peptide when the constitutional members are colour-coded according to their side chain properties: polar, acidic, basic, non-polar aliphatic, and aromatic.

Conformational properties

Another important aspect of structural composition is the proportion of peptide drugs and diagnostics that are either linear or contain a macrocycle. In nature, cyclic peptides of various sizes (from 8 to 50 AAs) occur in all the kingdoms of life. Their enhanced stability and advantageous biopharmaceutical properties make their application in drug design very common [58]. Indeed, as shown in Figure 3F, 53% of the marketed peptides are linear, while 47% present one or more macrocycles in their structure. Among the approved cyclic peptides, 39% are of natural origin, 55% are analogues and only 6% are heterologous – perhaps nature has once again demonstrated how to develop stable, biocompatible peptides, and is consequently a rich source of inspiration for candidates with optimal drug-like properties. Interestingly, our analysis highlighted that, overwhelmingly, peptide macrocycles consist of 5 to 7 residues, with only a few exceptions (e.g., nesiritide and carperitide comprise a 17-membered ring), as shown in Figure 3G. Examples of peptides with 5 to 7-membered macrocycles include oxytocin, desmopressin, lanreotide, and eptifibatide, to name but a few. In general, smaller macrocycles tend to have greater conformational stability due to physical restraints and fewer rotatable bonds. As such, macrocyclisation is a very common medicinal chemistry technique used to enhance peptide conformational stability and restrict the usual peptide chain flexibility [59,60]. This may stabilise peptide conformation for optimal receptor complexation and confer a protein-like secondary and tertiary structure [60–62].

Depending on the desired site of cyclisation, there are various strategies to generate cyclic peptides [63]. These can involve the peptide head (peptide C-terminal moiety), the peptide tail (peptide N-terminal moiety), or amino acid side chains. According to our findings, summarised in Fig. 3H, the most common technique is side chain-to-side chain cyclisation (58% of all the marketed cyclic peptides), with 25 out of 26 side chain-to-side chain macrocycles formed by a disulfide bond between cysteine thiol pairs. The exception is bremelanotide, in which a lactam is formed between the amine side chain of a lysine residue and the carboxylic acid side chain of an aspartic acid residue. However, disulfide bridges are not always metabolically stable *in vivo*, limiting their application [64]. Macrocycles within a peptide can be formed also by head-to-side chain cyclisation (24%), mainly *via* lactamisation between the C-terminal carboxylic acid and a side chain amine (e.g., lysine). The head-to-tail cyclisation between the N- and C-termini (7%), generates an all-amide end-to-end cyclic lactam, thus abrogating exopeptidase

hydrolysis. Finally, another cyclisation strategy encountered in 2 out of 47 cyclic peptides (i.e., grazoprevir and elcatonin) is side chain-to-tail macrocyclisation. When the head and/or the tail of the peptide are involved in the macrocycle, protease access to the backbone is reduced as the cyclisation removes the free N- and C-termini that are targeted by amino- and carboxy-peptidases, respectively [65]. This explanation bears out in the approved peptides, with the mean experimental terminal half-life of cyclic peptides compared to linear peptides, being 27 h and 12 h, respectively.

Amongst marketed peptide therapeutics and diagnostics, 38% present an amidated C-terminus, while 10% present an acetyl group at the N-terminus (Figures 3I and 3J). Modifications of peptide N-terminus include also the addition of pyroglutamic acid (7%) or deamination of the last AA (4%). Similarly to the head-to-tail cyclisation, this abrogates exopeptidase hydrolysis by masking amino- and carboxyl- termini [66]. Moreover, N-terminal acetylation or C-amidation precludes ionisation and hydrogen-bonding of NH_2 and COOH groups, respectively [67,68], thus better mimicking natural proteins.

Conclusions and future prospects

For the first time, this foundation review together with the database PepTherDia, offers the possibility of exploring common trends in approved peptide drugs and diagnostics. We highlighted some of the strategies most commonly used in peptide drug design, which successfully brought these peptides to the market. The trends underlined cannot be ascribed to luck or coincidence. The majority of approved peptides (84%) follows the rules "established" by nature over several millennia of evolution as they are naturally-derived or analogues of natural compounds. It is open to debate among drug discovery scientists whether it is important to follow the route of nature to obtain a successful peptide lead or whether the exploration of completely new chemical space and compounds through rational design and library screening is a fruitful solution. We believe that strategies provided by nature have to be properly comprehended and taken into account when designing new drug candidates.

On account of this meta-analysis, we can conclude that, at present, a peptide most likely to become a drug will have a molar mass lower than 2000 g/mol and will present a balance between hydrophobic and polar contributions. Furthermore, careful evaluation of the C- and N-terminal modifications will be key for the pharmacokinetic properties as well as for the desired activity. Moreover, if a cyclisation strategy is under evaluation, it is important for a peptide chemist to consider a more stable small-size macrocycle (5 to 7 members). In addition to the most frequently encountered natural amino acids, we have provided a list of commonly employed non-natural amino acids that can effectively replace the former to introduce enhanced properties or opportunities for further diversification. Indeed, strategies to modify a peptide through conjugation with fatty acids or sugars are well represented in various approvals, suggesting that they are promising tools to modify peptidic structures while retaining favourable pharmacological properties. The highlighted common threads that bring together the approved peptides does not have the presumption of being strict rules to follow, but rather they should be seen as a means of lending a hand to peptide designers to avoid certain pitfalls during early drug discovery stages. Indeed, some of the above trends may be particularly subjective towards certain disease areas or routes of administration.

Forecasting a starring role for peptides in the next decades, we anticipate that new clear trends – and perhaps rules – in structural composition will emerge beyond our observations, leading to improved rational peptide drug design. Whilst the mainstay of bioactive peptide

discovery remains an analogue-based approach – modifying naturally-derived peptides or protein epitopes – new medicinal chemistry strategies e.g., peptide-stapling, multicyclisation, novel amino acid synthesis, selective-fluorination, peptoids and improved *in silico* design, should accelerate the future approval rate of new designed or heterologous peptide pharmaceutical agents. In the medium term, a combination of both naturally derived and rationally designed strategies is likely to be the most successful route to fulfilling the increasing need for chemical entities to treat new or previously untreatable diseases, from cancer to infections, cardiovascular and neurodegenerative diseases.

Acknowledgements

The authors gratefully acknowledge Daniele Tomasi for the development of the website PepTherDia.

This work was supported by a Liverpool John Moores University Vice Chancellor's PhD Scholarship (VDA) and a University Alliance Doctoral Training Alliance COFUND/Marie Skłodowska-Curie (European Union's Horizon 2020 research and innovation programme, grant agreement No 801604) PhD Fellowship Programme in Applied Biosciences for Health (PD). The funding sources had no involvement in the study design.

Conflict of interest

Dr Coxon is a Director of Pepmotec Ltd, a peptide synthesis spin-out company from Durham University, UK.

Supplementary Material

The website PepTherDia contains the full list and classification of non-natural amino acids and modifications together with the detailed methodologies used for data collection, computational analysis and structural analysis.

References

1 Henninot, A., Collins, J.C. and Nuss, J.M. (2018) The Current State of Peptide Drug Discovery: Back to the Future? *Journal of Medicinal Chemistry*, **61**, 1382–1414. https://doi.org/10.1021/acs.jmedchem.7b00318.

2 Lau, J.L. and Dunn, M.K. (2018) Therapeutic Peptides: Historical Perspectives, Current Development Trends, and Future Directions. *Bioorganic and Medicinal Chemistry*, **26**, 2700–2707. https://doi.org/10.1016/j.bmc.2017.06.052.

3 Tsomaia, N. (2015) European Journal of Medicinal Chemistry Peptide Therapeutics: Targeting the Undruggable Space. *European Journal of Medicinal Chemistry*, **94**, 459–470.

4 Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D. and Bray, F. (2015) Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012. *International Journal of Cancer*, Wiley-Liss Inc., **136**, E359–E386. https://doi.org/10.1002/ijc.29210.

5 Daviglus, M.L., Lloyd-Jones, D.M. and Pirzada, A. (2006, September 20) Preventing Cardiovascular Disease in the 21st Century: Therapeutic and Preventive Implications

of Current Evidence. American Journal of Cardiovascular Drugs, Springer, 87–101. https://doi.org/10.2165/00129784-200606020-00003.

6 Zimmet, P.Z., Magliano, D.J., Herman, W.H. and Shaw, J.E. (2014, January 1) Diabetes: A 21st Century Challenge. The Lancet Diabetes and Endocrinology, Elsevier, 56–64. https://doi.org/10.1016/S2213-8587(13)70112-8.

7 Peptide Therapeutics Market to Expand at a CAGR of 7.9% from 2019 to 2027. Transparency Market Research, (2019) www.transparencymarketresearch.com/peptide-therapeutics-market.html [Accessed August 4, 2020]

8 ClinicalTrials.Gov https://www.clinicaltrials.gov/ [Accessed November 20, 2020]

9 Craik, D.J., Fairlie, D.P., Liras, S. and Price, D. (2013) The Future of Peptide-Based Drugs. *Chemical Biology and Drug Design*, **81**, 136–147. https://doi.org/10.1111/cbdd.12055.

10 Vlieghe, P., Lisowski, V., Martinez, J. and Khrestchatisky, M. (2010) Synthetic Therapeutic Peptides: Science and Market. *Drug Discovery Today*, **15**, 40–56. https://doi.org/10.1016/j.drudis.2009.10.009.

11 Mathur, D., Prakash, S., Anand, P., Kaur, H., Agrawal, P., Mehta, A., Kumar, R., Singh, S. and Raghava, G.P.S. (2016) PEPlife : A Repository of the Half- Life of Peptides. *Scientific Reports*, 1–7.

12 Usmani, S.S., Bedi, G., Samuel, J.S., Singh, S., Kalra, S., Kumar, P., Ahuja, A.A., Sharma, M., Gautam, A. and Raghava, G.P.S. (2017) THPdb: Database of FDA-Approved Peptide and Protein Therapeutics. *PLoS ONE*, **12**, 1–12. https://doi.org/10.1371/journal.pone.0181748.

13 Davenport, A.P., Scully, C.C.G., de Graaf, C., Brown, A.J.H. and Maguire, J.J. (2020) Advances in Therapeutic Peptides Targeting G Protein-Coupled Receptors. *Nature Reviews Drug Discovery*, **19**, 389–413. https://doi.org/10.1038/s41573-020-0062-z.

14 Tyagi, M., Begnini, F., Poongavanam, V., Doak, B.C. and Kihlberg, J. (2020) Drug Syntheses Beyond the Rule of 5. *Chemistry - A European Journal*, **26**, 49–88. https://doi.org/10.1002/chem.201902716.

15 Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (1997) Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Advanced drug delivery reviews*, 3–25.

16 Wager, T.T., Hou, X., Verhoest, P.R. and Villalobos, A. (2016) Central Nervous System Multiparameter Optimization Desirability: Application in Drug Discovery. *ACS Chemical Neuroscience*, **7**, 767–775. https://doi.org/10.1021/acschemneuro.6b00029.

17 McNaught, A.D. and A. Wilkinson. (1997) IUPAC. Compendium of Chemical Terminology, 2nd Ed. (the “Gold Book”). . . Blackwell Scientific Publications,

Oxford. https://goldbook.iupac.org/terms/view/P04479.

18 IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). (1984) Nomenclature and Symbolism for Amino Acids and Peptides: Recommendations 1983. *European Journal of Biochemistry*, **138**, 9–37. https://doi.org/10.1111/j.1432-1033.1984.tb07877.x.

19 FDA.Gov www.fda.gov/drugs/regulatory-science-action/impact-story-developing-tools-evaluate-complex-drug-products-peptides [Accessed August 10, 2020]

20 Wu, L.C. (2013) Regulatory Considerations for Peptide Drug Products. 1–30.

21 Uhlig, T., Kyprianou, T., Martinelli, F.G., Oppici, C.A., Heiligers, D., Hills, D., Calvo, X.R. and Verhaert, P. (2014) The Emergence of Peptides in the Pharmaceutical Business: From Exploration to Exploitation. *EuPA Open Proteomics*, **4**, 58–69. https://doi.org/10.1016/j.euprot.2014.05.003.

22 Go.DrugBank https://go.drugbank.com/ [Accessed May 20, 2020]

23 EMA-European Medicines Agency www.ema.europa.eu/en [Accessed May 20, 2020]

24 PMDA - Pharmaceuticals and Medical Devices Agency www.pmda.go.jp/english/index.html [Accessed May 10, 2020]

25 Drug Central http://drugcentral.org [Accessed August 20, 2020]

26 National Center for Advancing Translational Sciences NIH https://drugs.ncats.io [Accessed August 20, 2020]

27 Drugs.Com https://www.drugs.com [Accessed September 1, 2020]

28 Lehninger, A.L., Cox, M.M. and Nelson, D.L. (2000) Lehninger Principles of Biochemistry. 6th ed., New York: Worth Publisher.

29 Walther, F.J., Gordon, L.M. and Waring, A.J. (2016) Design of Surfactant Protein B Peptide Mimics Based on the Saposin Fold for Synthetic Lung Surfactants. *Biomedicine Hub*, **1**, 1–21. https://doi.org/10.1159/000451076.

30 Yin, N., Brimble, M.A., Harris, P.W. and Wen, J. (2014) Enhancing the Oral Bioavailability of Peptide Drugs by Using Chemical Modification and Other

Approaches. *Medicinal Chemistry*, **4**, 763–769. https://doi.org/10.4172/2161-0444.1000227.

31 Chanson, P., Tiit, J. and Harris, A.G. (1993) Clinical Pharmacokinetics of Octreotide: Therapeutic Applications in Patients with Pituitary Tumours. Clinical Pharmacokinetics, Clin Pharmacokinet, 375–391. https://doi.org/10.2165/00003088-199325050-00004.

32 O'Toole, T.J. and Sharma, S. (2019) Physiology, Somatostatin. StatPearls, StatPearls Publishing. http://www.ncbi.nlm.nih.gov/pubmed/30855911.

33 Hodgson, D.R.W. and Sanderson, J.M. (2004) The Synthesis of Peptides and Proteins Containing Non-Natural Amino Acids. *Chemical Society Reviews*, Royal Society of Chemistry, **33**, 422–430. https://doi.org/10.1039/b312953p.

34 Hohsaka, T. and Sisido, M. (2002, December 1) Incorporation of Non-Natural Amino Acids into Proteins. Current Opinion in Chemical Biology, Elsevier Ltd, 809–815. https://doi.org/10.1016/S1367-5931(02)00376-9.

35 Pollegioni, L. and Servi, S. (2012) Unnatural Amino Acids. Methods in Molecular Biology, Press, Humana.

36 Mant, C.T., Jiang, Z., Gera, L., Davis, T., Nelson, K.L., Bevers, S. and Hodges, R.S. (2019) De Novo Designed Amphipathic α-Helical Antimicrobial Peptides Incorporating Dab and Dap Residues on the Polar Face to Treat the Gram-Negative Pathogen, Acinetobacter Baumannii. *Journal of Medicinal Chemistry*, American Chemical Society, **62**, 3354–3366. https://doi.org/10.1021/acs.jmedchem.8b01785.

37 Lu, J., Xu, H., Xia, J., Ma, J., Xu, J., Li, Y. and Feng, J. (2020) D- and Unnatural Amino Acid Substituted Antimicrobial Peptides With Improved Proteolytic Resistance and Their Proteolytic Degradation Characteristics. *Frontiers in Microbiology*, Frontiers Media S.A., **11**, 563030. https://doi.org/10.3389/fmicb.2020.563030.

38 Witkowska, E., Orłowska, A., Sagan, B., Smoluch, M. and Izdebski, J. (2001) Tryptic Hydrolysis of HGH-RH(1-29)-NH2 Analogues Containing Lys or Orn in Positions 12 and 21. *Journal of Peptide Science*, J Pept Sci, **7**, 166–172. https://doi.org/10.1002/psc.316.

39 Inouy, K., Sakakibara, S. and Prockop, D.J. (1976) Effects of the Stereo-Configuration of the Hydroxyl Group in 4-Hydroxyproline on the Triple-Helical Structures Formed by Homogeneous Peptides Resembling Collagen. *BBA - Protein Structure*, Elsevier, **420**, 133–141. https://doi.org/10.1016/0005-2795(76)90352-4.

40 Meyer, D., Mutschler, C., Robertson, I., Batt, A. and Tatko, C. (2013) Aromatic Interactions with Naphthylalanine in a *β* -Hairpin Peptide. *Journal of Peptide Science*, John Wiley & Sons, Ltd, **19**, 277–282. https://doi.org/10.1002/psc.2496.

41 Sidransky, H., Verney, E. and Kurl, R. (1990) Comparison of Effects of L-Tryptophan and a Tryptophan Analog, D,L-β-(1-Naphthyl)Alanine, on Processes Relating to Hepatic Protein Synthesis in Rats. *Journal of Nutrition*, Oxford Academic, **120**, 1157–1162. https://doi.org/10.1093/jn/120.10.1157.

42 Sidransky, H. and Verney, E. (1996) Influence of L-Alanine on Effects Induced by L-Tryptophan on Rat Liver. *Journal of Nutritional Biochemistry*, Elsevier Inc., **7**, 200–206. https://doi.org/10.1016/0955-2863(96)00010-1.

43 Rodriguez, M., Bernad, N., Galas, M., Lignon, M., Laur, J., Aumelas, A. and Martinez, J. (1991) Synthesis and Biological Activities of Cholecystokinin Analogues Substituted in Position 30 by 3-(1-Naphthyl)-l-Alanine [Nal(1)] or 3-(2-Naphthyl)-l-Alanine [Nal(2)]. *European Journal of Medicinal Chemistry*, Elsevier Masson, **26**, 245–253. https://doi.org/10.1016/0223-5234(91)90056-S.

44 Zhang, L. and Bulaj, G. (2012) Converting Peptides into Drug Leads by Lipidation. *Current Medicinal Chemistry*, Bentham Science Publishers Ltd., **19**, 1602–1618. https://doi.org/10.2174/092986712799945003.

45 Moradi, S.V., Hussein, W.M., Varamini, P., Simerska, P. and Toth, I. (2016, April 1) Glycosylation, an Effective Synthetic Strategy to Improve the Bioavailability of Therapeutic Peptides. Chemical Science, Royal Society of Chemistry, 2492–2500. https://doi.org/10.1039/c5sc04392a.

46 Kang, R., Wan, J., Arstikaitis, P., Takahashi, H., Huang, K., Bailey, A.O., Thompson, J.X., Roth, A.F., Drisdel, R.C., Mastro, R., Green, W.N., Yates 3rd, J.R., Davis, N.G. and El-Husseini, A. (2008) Neural Palmitoyl-Proteomics Reveals Dynamic Synaptic Palmitoylation HHS Public Access. *Nature*, **456**, 904–909. https://doi.org/10.1038/nature07605.

47 Ward, B.P., Ottaway, N.L., Perez-Tilve, D., Ma, D., Gelfanov, V.M., Tschöp, M.H. and DiMarchi, R.D. (2013) Peptide Lipidation Stabilizes Structure to Enhance Biological Function. *Molecular Metabolism*, **2**, 468–479. https://doi.org/10.1016/j.molmet.2013.08.008.

48 DeFrees, S. (2007, July 23) Glycosylation of Peptides via O-Linked Glycosylation Sequences.

49 Akhilesh Kumar and Anand K. Bachhawat. (2012) Pyroglutamic Acid: Throwing Light on a Lightly Studied Metabolite on JSTOR. Current Science, 288–297. https://www.jstor.org/stable/24083854?seq=1.

50 Garden, R.W., Moroz, T.P., Gleeson, J.M., Floyd, P.D., Li, L., Rubakhin, S.S. and Sweedler, J. V. (1999) Formation of N-Pyroglutamyl Peptides from N-Glu and N-Gln Precursors in Aplysia Neurons. *Journal of Neurochemistry*, John Wiley & Sons, Ltd, **72**, 676–681. https://doi.org/10.1046/j.1471-4159.1999.0720676.x.

51 Kehlen, A., Haegele, M., Böhme, L., Cynis, H., Hoffmann, T. and Demuth, H.U. (2017) N-Terminal Pyroglutamate Formation in CX3CL1 Is Essential for Its Full Biologic Activity. *Bioscience Reports*, Portland Press Ltd, **37**, 20170712. https://doi.org/10.1042/BSR20170712.

52 Shah, P. and Westwell, A.D. (2007) The Role of Fluorine in Medicinal Chemistry. *Journal of Enzyme Inhibition and Medicinal Chemistry*, Taylor & Francis, **22**, 527–540. https://doi.org/10.1080/14756360701425014.

53 Marsh, E.N.G., Buer, B.C. and Ramamoorthy, A. (2009) Fluorine - A New Element in

the Design of Membrane-Active Peptides. *Molecular BioSystems*, **5**, 1143–1147. https://doi.org/10.1039/b909864j.

54 Meng, H., Krishnaji, S.T., Beinborn, M. and Kumar, K. (2008) Influence of Selective Fluorination on the Biological Activity and Proteolytic Stability of Glucagon-like Peptide-1. *Journal of Medicinal Chemistry*, **51**, 7303–7307. https://doi.org/10.1021/jm8008579.

55 Chemicalize https://chemicalize.com/ [Accessed September 4, 2020]

56 Faulds, D., Goa, K.L. and Benfield, P. (1993) Cyclosporin: A Review of Its Pharmacodynamic and Pharmacokinetic Properties, and Therapeutic Use in Immunoregulatory Disorders. *Drugs*, **45**, 953–1040. https://doi.org/10.2165/00003495-199345060-00007.

57 Novartis. (1988) Sandimmune. Intensive Therapy and Clinical Monitoring, 172.

58 Thorstholm, L. and Craik, D.J. (2012) Discovery and Applications of Naturally Occurring Cyclic Peptides. *Drug Discovery Today: Technologies*, Elsevier Ltd, **9**, e13–e21. https://doi.org/10.1016/j.ddtec.2011.07.005.

59 Qvit, N., Rubin, S.J.S., Urban, T.J., Mochly-Rosen, D. and Gross, E.R. (2017) Peptidomimetic Therapeutics: Scientific Approaches and Opportunities. *Drug Discovery Today*. https://doi.org/10.1016/j.drudis.2016.11.003.

60 Hill, T.A., Shepherd, N.E., Diness, F. and Fairlie, D.P. (2014) Constraining Cyclic Peptides to Mimic Protein Structure Motifs. *Angewandte Chemie - International Edition*, **53**, 13020–13041. https://doi.org/10.1002/anie.201401058.

61 Lenci, E. and Trabocchi, A. (2020) Peptidomimetic Toolbox for Drug Discovery. *Chemical Society Reviews*, **49**, 3262–3277. https://doi.org/10.1039/d0cs00102c.

62 Lambert, J.N., Mitchell, J.P. and Roberts, K.D. (2001) The Synthesis of Cyclic Peptides. *Journal of the Chemical Society. Perkin Transactions 1*, 471–484. https://doi.org/10.1039/b001942i.

63 Tapeinou, A., Matsoukas, M.-T., Simal, C. and Tselios, T. (2015) Review Cyclic Peptides on a Merry-Go-Round; towards Drug Design. *Biopolymers*, John Wiley & Sons, Ltd, **104**, 453–461. https://doi.org/10.1002/bip.22669.

64 Botti, P., Pallin, T.D. and Tam, J.P. (1996) Cyclic Peptides from Linear Unprotected Peptide Precursors through Thiazolidine Formation. *Journal of the American Chemical Society*, **118**, 10018–10024. https://doi.org/10.1021/ja954278g.

65 Jing, X. and Jin, K. (2020) A Gold Mine for Drug Discovery: Strategies to Develop Cyclic Peptides into Therapies. *Medicinal Research Reviews*, **40**, 753–810. https://doi.org/10.1002/med.21639.

66 Wallace, R.J. (1992) Acetylation of Peptides Inhibits Their Degradation by Rumen Micro-Organisms. *British Journal of Nutrition*, Cambridge University Press (CUP),

68, 365–372. https://doi.org/10.1079/bjn19920095.

67 Lee, A.C.L., Harris, J.L., Khanna, K.K. and Hong, J.H. (2019) A Comprehensive Review on Current Advances in Peptide Drug Development and Design. *International Journal of Molecular Sciences*, **20**. https://doi.org/10.3390/ijms20102383.

68 Kim, K. and Seong, B.L. (2001) Peptide Amidation: Production of Peptide Hormones in Vivo and in Vitro. *Biotechnol. Bioprocess Eng.*, **6**, 244–251.

Figure Legends

[FIGURE 1]

Figure 1. Cumulative frequency distribution-year of approval; withdrawn peptides have not been considered. Data to create this graph have been collected by searching the year in which each peptide has been firstly approved by one of the main agencies (FDA, EMA, PMDA).

[FIGURE 2]

Figure 2. Daptomycin and its constitutional members. Kyn (kynurenine), Orn (ornithine), 3-Me-Glu (3-methyl-glutamic acid).

[FIGURE 3]

Figure 3. Structural trends in the pool of approved peptide therapeutics and diagnostics. A. Number of constitutional members distributions; B. Molar mass (g/mol) distribution; C. L-AAs (light blue) and D-AAs (green) occurrence; D. Most frequent encountered non-natural AAs, in pink, and modifications, in purple; E. Peptide origin; F. Peptide structure, divided in linear, monocyclic and multicyclic; G. Macrocycle size, shown as number of constitutional members per cycle; H. Type of bond to form the cycle within the peptide structure; I. C-terminal modifications; J. N-terminal modifications.

[FIGURE 4]

Figure 4. Peptide polarity evaluation. A. Polarity distribution within the pool of approved peptides; B. Colour-coded plot to show the aliphatic, aromatic, polar, basic and acidic contributions in each peptide under evaluation.