Bioorganic & Medicinal Chemistry 20 (2012) 3565–3574

Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Automated generation of turn mimetics: Proof of concept study for the MC4 receptor

J. Christian Baber [†], Richard Lowe [‡], John Saunders [§], Miklos Feher [*,¶]

Neurocrine Biosciences, 12790 El Camino Real, San Diego, CA 92130, USA

ARTICLE INFO

Article history:
Received 23 January 2012
Revised 29 March 2012
Accepted 3 April 2012
Available online 13 April 2012

Keywords:
MC4 receptor
Melanocortin 4 receptor
Turn mimetics
β Turns
De novo design
Flexible alignments

ABSTRACT

An algorithm has been devised for the automatic design of peptide turn mimetics, particularly applicable to peptide-activated GPCRs. The method is based on flexible alignments using a new design paradigm and scoring system that aims to reduce the molecular weight of the compound and preferentially lead to drug like molecules. The process can be applied either as a de novo design or a virtual screening tool. Its use has been demonstrated by the design of novel double digit nanomolar ligands for the melanocortin 4 receptor (MC4). The method is, in principle, applicable to any type of receptor, including orphan receptors.

Introduction

Mimicking peptides has been an attractive path for generating novel lead structures for drug discovery during the last few decades.[1–3] Peptides are capable of complex interactions with their targets but generally possess properties (such as rapid metabolism and low oral bioavailability) that prevent them from becoming oral drugs.[4] Peptidomimetic design has been most successful in mimicking peptide binding, especially in characteristic conformations such as α-helices, β-strands, β-sheets and various peptide turns.[5–8]

In this work our goal was to develop an automated procedure for the de novo design of drug-like small molecules for peptide-activated GPCRs (such as SSRT, MC4, MCH, CRF, GnRH). It is well known that different turn structures are often crucial in the binding of peptides to their GPCR targets.[9] We wanted to test whether turn mimetics could be designed without any experimental structural information using a novel design paradigm. Our hope was that such a method could then be generalized to find new lead structures for any receptor, potentially including orphan receptors.

The melanocortin 4 (MC4) receptor was selected for the initial testing of our concept.[10] This receptor is associated with weight homeostatis[11] and ligands for this receptor, both agonists and antagonists, are being pursued by a number of pharmaceutical companies for this purpose and other indications.[12] Different kinds of MC4 inhibitors have been tested at Neurocrine Biosciences[13–15] and the de novo design of turn mimetics was viewed as a novel way to generate ideas for alternative starting points. One natural peptide ligand of this receptor is α-MSH[16] (Ac-Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-Val-NH_2). It is known from the literature that the minimal sequence required for efficacy is the Ac-His-Phe-Arg-Trp-NH_2 section[17] but this tetrapeptide is 200,000-fold less active than the full-length peptide. Smaller fragments of this sequence (Ac-His-Phe-NH_2, Ac-Phe-Arg-NH_2, Ac-His-Phe-Arg-NH_2) are devoid of melanotropic activity at concentrations as high as 0.1 mM, similarly to the tetrapeptide, Ac-Phe-Arg-Trp-Gly-NH_2. On the other hand, *d*-Phe is thought to enhance the Type-I β-turn conformation and hence certain cyclic peptide structures, such *n*-C_4H_9CO-cyclo(Asp-APC-*d*-Phe-Arg-Trp-Lys)-NH_2 (where APC indicates 1-amino-4-phenylcyclohexane-1-carboxylic acids) or cyclo($COCH_2CH_2CO$-His-*d*-Phe-Arg-Trp-Lys)-NH_2 have nanomolar binding affinities and are being developed by pharmaceutical companies as MC4 agonists.[18] From this it was concluded that the Phe-Arg-Trp sequence is the essential 'messenger' region in the peptide.[19] It was established by NMR studies that this is a turn region in the peptide and the turn type

* Corresponding author. Tel.: +1 416 581 7611.
E-mail address: mfeher@uhnres.utoronto.ca (M. Feher).
† Present address: Cubist Pharmaceuticals, Inc., 65 Hayden Avenue, Lexington, MA 02421, USA.
‡ Present address: Venture Strategies Innovations, 2401 East Katella Avenue, Anaheim, CA 92806, USA.
§ Present address: Vertex Pharmaceuticals, 11010 Torreyana Road, San Diego, CA 92121, USA.
¶ Present address: Campbell Family Institute for Breast Cancer Research, University Health Network, Toronto Medical Discovery Tower, 101 College Street, Suite 5-361, Toronto, ON, Canada M5G 1L7.

0968-0896/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bmc.2012.04.001

varies by the type of peptide substituents, although it is a type I β-turn in the wild-type sequence[20] Thus we concentrated on this region and turn type (shown in Fig. 1a) with our automated mimetic design.

Classical peptide turn mimetics have typically been based on the idea that strategically placed sidechains off of a central template were capable of mimicking three or all of the principal residues of the turn.[21] This design principle is shown for our concrete case in Figure 1b, which, as can be seen, puts a substantial bulk in the central region where it is not likely to have any direct interaction with the target. Examples for this principle include the design of non-peptidic endothelin antagonists[22] or cyclic LDV peptide analogues.[23] Our initial expectation was that the molecular weight of the compound could be decreased without reducing affinity by shifting the bulk of the template to overlay with one of the residues (see Fig. 1c). In this way, the template itself participates in 'useful' interactions, mimicking hydrophobic or hydrogen bonding interactions of the peptide. The expectation was that this would lead to hits with lower molecular weight and higher novelty compared to other approaches.

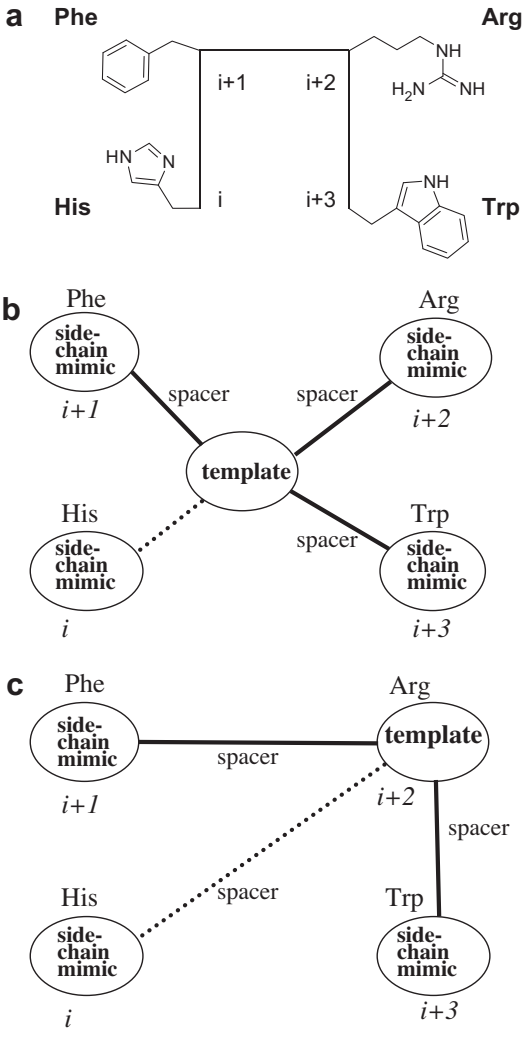


Figure 1. Design hypotheses for generating peptide turn mimetics. (a) Schematic structure of the His-Phe-Arg-Trp β-turn of the α-MSH peptide, showing the sidechains targeted by the mimetic design. (b) 'Traditional' design using a central template to hold the substituents and attached substituents to mimic the interaction of at least 3 peptide sidechains. (c) Our proposed design idea, with the template directly mimicking one of the peptide sidechains and substitutions to mimic the interactions of at least 2 further peptide sidechains. It was hoped that by placing the template at one of the sidechains, the obtained designs will be novel and have somewhat lower molecular weights.

2. Computational chemistry

2.1. The computational process

The computational process was designed to be general and iterative, as shown in Figure 2. The process starts with turn generation, in which low energy conformations of the turn residues of interest are generated. Although this part of the process was initially integrated with the rest of the script, this component is typically used less often and hence was later separated into its own module.

Two different workflows were applied in the design of turn mimetics: virtual screening and de novo design. In virtual screening mode, a virtual library of ligands was screened for their ability to mimic the turn. This was achieved by aligning ligands to the peptide turn using automated flexible alignment, followed by scoring, with the molecules with better than a specified quality of fit being stored in a molecular database. In the de novo design mode, the raw template was modified iteratively on the fly using an evolutionary algorithm to have increasingly better substituents. Each design was aligned to the template and scored with the quality of fit used to drive the evolutionary algorithm towards improved designs. Molecules that achieved sufficiently high scores were added to a molecular database of accepted solutions. Regardless of which workflow was applied, once the turn of interest was fully evaluated by the program, the resulting molecular database was manually inspected and molecules were selected for synthesis.

All molecular alignment, scoring and most of the other computational steps were performed within the MOE software using a program written in MOE's svl language.[24] De novo design was done using the EA Inventor (EAI) program.[25] Each population of structures was transferred from EAI to MOE as a set of SMILES strings via a Python wrapper. Next, the computational processes underlying each step will be discussed.

2.2. Turn generation

The process of turn generation was designed to be entirely general, starting from the primary sequence of the peptide and the turn type of interest to be mimicked, as expressed by the Ψ and Φ angles of the turn. The process starts at a predefined point on the primary sequence (e.g., at the N-terminus or in an area where the turn is known or suspected). The algorithm then proceeds along the peptide towards the C-terminus, examining each amino acid in turn until a known turn-initiating sequence (or other specified sequence) is identified. The peptide is then truncated to a sequence including just the turn and one residue on each side and the turn backbone conformation generated using predefined turn (Ψ and Φ) angles. Multiple low energy protein sidechain conformations are then generated using the stochastic conformational search in MOE, with the entire backbone conformation fixed as defined by the turn geometry. The additional residues at the end of the sequence help to restrain the conformations of the first and last turn residues which tend to behave undesirably if there is nothing adjacent to them. Since the stochastic conformational search in MOE is relatively efficient and usually finds the lowest energy conformations within the first 10–50 steps, the search is run with the process terminating if either no new conformation is generated in 100 sequential attempts or when a total of 200 conformations are reached. Next the backbones of the key turn residues are left fixed while the rest of the structure (sidechains and neighboring residues) minimized using the Merck forcefield[26] with GBSA continuum solvation.[27,28] Although there are protein-specific forcefields

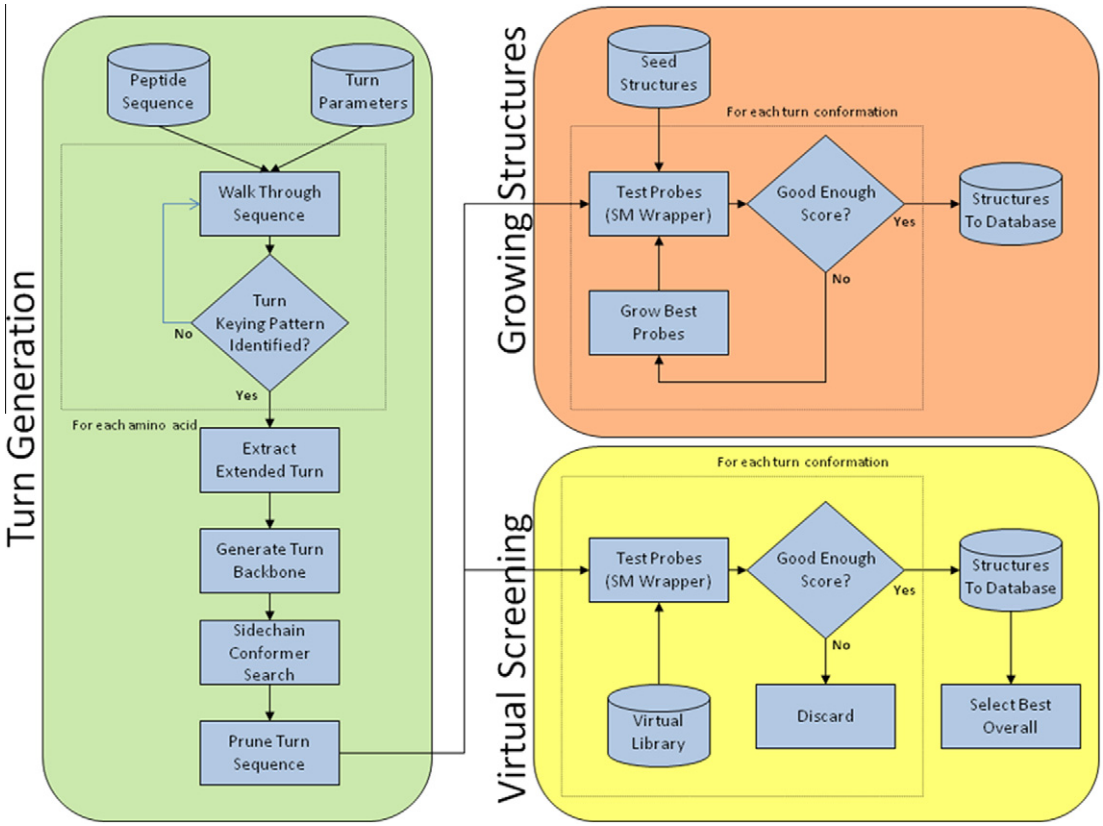


Figure 2. Schematic diagram of the design process.

available (e.g., CHARmM, OPLS, Amber), the Merck forcefield was selected primarily for compatibility with small molecule optimizations and flexible alignments in later stages. The minimization is carried out in two stages. Since it is possible that the initial turn conformation might contain bond lengths and bond angles far from optimal, the structures are first pre-minimized in vacuum to a gradient of 20 to avoid negative Born radii terms in the continuum solvation model. The continuum solvation is then turned on and each turn conformation further minimized to a gradient of 0.01 with high energy conformations (50 kcal/mol above minimum) discarded. The additional residues at each end of the turn are then deleted and a diverse subset of conformers (typically up to 5) retained. Finally the entire backbone of each of the retained turn conformations is deleted to leave a number of different low-energy sidechain conformations.

In theory, the process can handle any type of turn, as defined by the backbone Ψ and Φ angles. The backbone angles and keying sequences for different turn types are stored in the program, hence only the turn type needs be defined as input. However, if additional information is available—either on the turn location or structural information that may further restrain the conformation—then this may be used either as a post-processing filter or included directly in the process. This was the case in the validation example described in this work where there is evidence that the interaction of interest is with a Type-I β-turn at the specific residue quartet His-Phe-Arg-Trp and therefore only this region was used for turn generation. It is also possible to restrict the alignment so that specified parts of the structure are aligned to selected residues in the peptide. In this study the core was designed to contain a basic nitrogen which was fixed to align with the basic Arg sidechain in the turn. In the remainder of this manuscript, only this specific case will be described. Next, each of the generated conformations was considered in turn in the virtual screening or de novo design process.

2.3. Computational testing of probes and compound scoring

In our process, molecules that were tested against the peptide turn were called probes. The concept of probes included any molecule that was computationally checked for similarity against the peptide turn at various stages of the process. Probes were tested using the flexible alignment process implemented in MOE[29] against each of the selected peptide turn sidechain conformations. As mentioned above, the different peptide conformations were separately considered, with the given conformation of the peptide turn sidechains was kept fixed for each individual run. The process is designed to be iterative with all conformations used initially but only those that seem to be the most biologically relevant (as shown by the accuracy of the predictions compared to synthesis and biological testing) used in later optimization runs. The flexible alignment score, S,[29] was applied to optimize and rank different alignment conformations and poses. These scores were developed to reproduce X-ray alignments[29] and consider hydrogen bonding and hydrophobic interactions in a distance-dependent manner. They are determined as the overlap functions of Gaussian densities. These Gaussians describe property distributions and are defined so that any number of properties can be included; in this work we used the optimized properties (such as volume, donors and acceptors and aromaticity) and optimized weights.[29]

The flexible alignment score contains a sum of distance dependent terms that, with the default settings, treats all atoms as being equivalent, instead of emphasizing the proper alignment of critical residue atoms (such as charges, planes of aromatic rings, etc.). This is a particularly serious issue with the linker regions of the peptide mimetic and the backbone of the peptide. Their role is to properly hold end-groups in place, they are not generally meant to interact with the target but, if present, the

flexible alignment score treats them equally to critical interacting atoms. In the case of the peptide backbone this was addressed by deleting it at the end of the turn conformation generation process and using only the sidechain conformations fixed in space for the alignment and scoring. However, such an approach was not possible for the flexibly aligned probes themselves, necessitating the development of a relevant additional scoring function for the purposes of ranking.

In order to improve the chances and speed of identifying active ligands, an empirical scoring function was introduced. The empirical scoring system was initially designed to quickly include properties in addition to the quality of the alignment in order to ensure that generated structures were drug-like. However, it was observed that a great deal of time was spent in the relatively slow flexible alignment part of the process attempting to align structures that clearly stood no chance of aligning well—either through not having been extended far enough or through lacking the appropriate pharmacophoric groups. For this reason, when dealing with structure growth, the scoring system was modified to also include a quick assessment of whether a good alignment was a priori possible. If the quick empirical scoring indicates that the structure cannot possibly align well with the target turn sidechains then the time-consuming full flexible alignment and scoring stage is skipped. This empirical scoring system is used as a wrapper around the main flexible alignment part of the scoring algorithm both to include physical properties and improve the speed of the system. The wrapper (Fig. 3) ensures that if either the properties or the potential of achieving a good alignment are very poor, the full flexible alignment is never run saving a substantial amount of time. If the flexible alignment is run, the properties and fit are still included in the final score. This substantially improves the speed of the process—particularly at the start of a de novo design run when an extension is often not large enough to reach the space occupied by the target residue—while also allowing specific weight and other property ranges to be targeted.

When testing the relationship between the scores generated by the system and activity of synthesized compounds as potential predictive tools for binding affinity, it was discovered that the scores did not correlate well with actual experimental data (not presented). This is not too surprising given that that the alignment score was developed for reproducing the relative placement of ligands in binding pockets and not for scoring ligands according to their activities, and the additional factors added in the wrapper relate more to drug likeness than binding affinity. The score generated by the system was therefore used to rank order compounds and provide guidance on which to make along with other factors such as synthetic accessibility.

2.4. Virtual screening and de novo design

The turn generation and alignment process can be used to assess the potential of virtual compounds from any source. In this study two different sources were considered; virtual screening and de novo design.

In the virtual screening workflow, the alignments of molecules from a virtual library were tested consecutively against each of the peptide turn sidechain conformations. In a general case the alignment of the molecule against all conformations of all turn residues is considered and the results ranked and scored. However, in the examples described here, virtual libraries were built around a specific template. In this case, only the known turn residues of interest were considered although multiple conformations of the turn were included. Flexible alignments of the potential structures with the peptide turn sidechains were generated and scored as described above. All results scoring above a specified level were stored in a MOE database and ranked at the end of the process according to

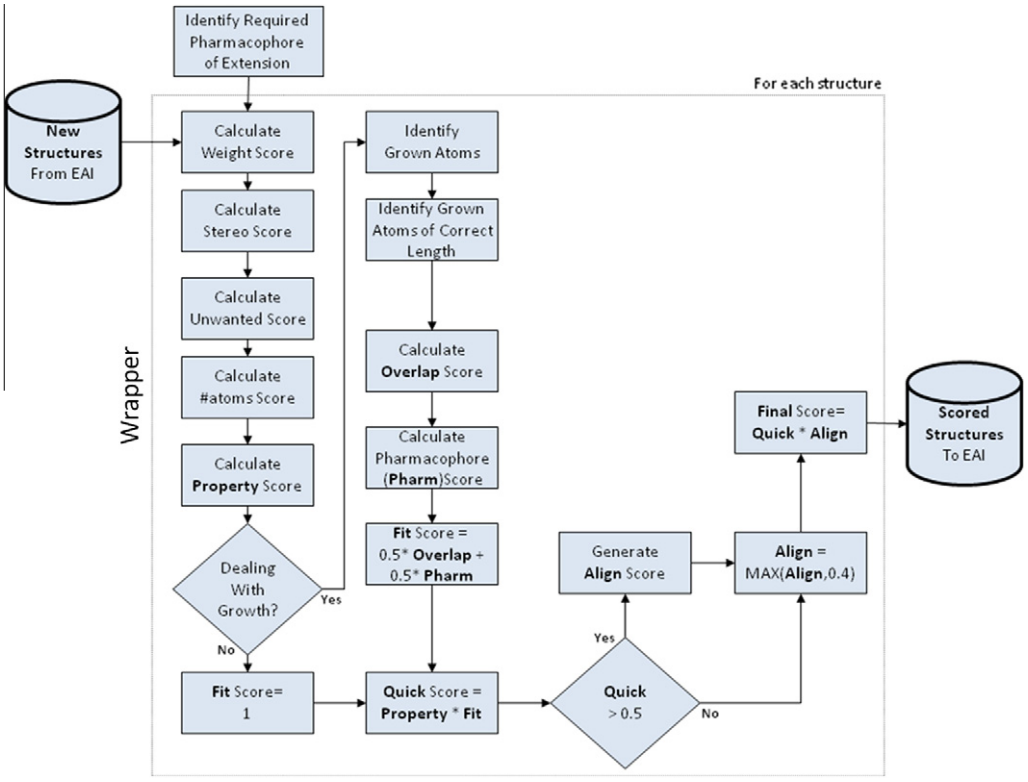


Figure 3. The wrapper used to quickly establish the 'empirical score'. See text for further details.

the empirical alignment score. The top-scoring solutions were then manually inspected, with molecules with the most convincing alignments selected for synthesis.

While the virtual screening process pre-supposes that a virtual library of appropriate molecules is supplied to the program, the de novo process generates interesting substitutions from scratch. In this study, the process was started with a template designed to mimic specific turn residues, This template was again aligned to the respective turn residues and anchored to them. Substitutions were then built at one of the selected vectors on the template using the Evolutionary Inventor (EA Inventor). Possible growth vectors on the template were selected based on the synthetic route used for its synthesis in order to increase the synthetic accessibility of generated compounds. Evolutionary algorithms evolve improved populations of structures by performing modifications on the members of previous generations ('mutations') and, to a lesser extent, swapping successful substructures among different molecules ('crossovers'). The evolution of new molecules is via 'survival of the fittest' to optimize the applied empirical score. In reality, this meant that increasingly better fitting substituents were added to the template until a structure with both a good alignment with the turn and acceptable properties was generated. There is a natural tendency for the process to keep building in extra bulk and this was prevented by an additional pruning step every 10 iterations which culled the structure set based on size and a measure of complexity. In the described example, substitutions were added along two vectors of the template. Although in theory the two substitutions could have been built by EA Inventor simultaneously, this was found to be very inefficient with a strong tendency of the program to select one position and repeatedly extend there while ignoring the other point of diversity. This is thought to be due to the random nature of the growth process in EA Inventor with a modification point selected at random from the list of allowed atoms. Initially, with two points of diversity there are two allowed extension points. However, once one of those points is extended, every atom that is added is also eligible for further extension. Since the selection of where to extend is on an atom-by-atom basis rather than by substituent, whichever point is selected first is far more likely to be selected again and therefore mono-substituted compounds tend to be generated. Hence at first one of the substitutions was 'optimized' (i.e., 10–20 solutions were generated by the process and the top scoring ones chosen after manual selection). The process was then re-run for each of the identified solutions with substitutions allowed only on the second point of diversity. This fitted well with synthetic plans as it allowed for compounds with one 'optimal' substitution to be synthesized (these were called 'supertemplates') and used as intermediates when synthesizing final products.

The time required by the computational process depended on the number of generations of molecules that were produced. In a typical benchmark run taking 3 h, around 50 generations of 50 molecules were produced on a 2.8 GHz Xenon PC with 2 GB of memory, running under the RedHat Enterprise 3.0 operating system. This translated to around 30 molecules passing the initial screen, 10 passing the quick score and 2 molecules scoring highly enough in the full flexible alignment scoring to be of further interest. Increasing the number of generations to 80 lead to a run taking 8 h with 40 molecules passing the initial screen, 30 passing the quick score and 3 molecules scoring highly enough in the full flexible alignment scoring to be of further interest. Consequently the process was generally carried out using multiple short (50 generations) runs rather than fewer longer runs which both resulted in more high-scoring molecules being generated and a greater diversity in the generated molecules due to the inherent tendency of each individual de novo run to optimize towards multiple similar high-scoring molecules.

3. Chemistry

The selection of the initial template was considered from several aspects. Based on the design idea, one essential requirement was that it had to have at least two sites of diversity and these vectors had to point in directions that allowed the mimicking of the peptide turn. Also, the template had to be accessible by known chemistry to allow relatively straightforward synthesis although initial searches on product structures had to indicate that there is freedom to operate. Several different templates were considered based on these criteria, and the pyrrolidine-based templates (Schemes 1 and 2) were selected as the first examples given that both cis and trans geometries were readily available.

Synthesis of the *cis*-pyrrolidine compounds is shown in Scheme 1 and proceeded as follows: reaction of the anhydride **1**[30] with an amine in acetonitrile at room temperature gave the acid amide **2**. Amide coupling with a second amine using the polymer-supported coupling reagents, PS-DCC and PS-DIPAM, followed by removal of the Boc-group with TFA in methanol and purification by HPLC gave the bis-amide **3**. Reductive amination with a carboxaldehyde (R_3CHO) followed by HPLC purification gave the tertiary amines **4**.

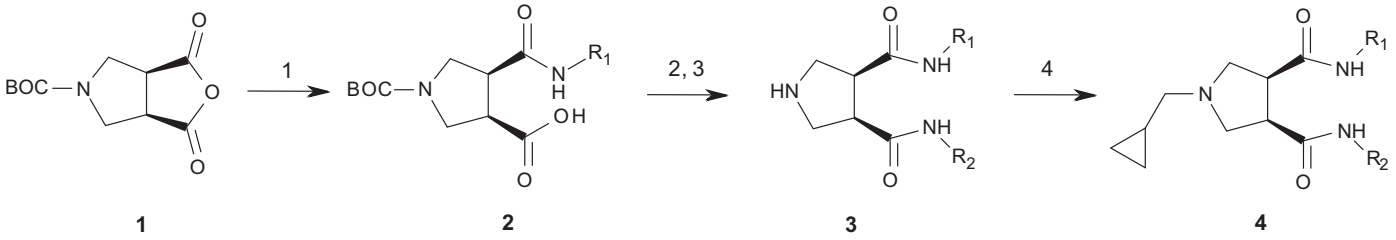
Synthesis of the *trans*-pyrrolidine compounds was carried out using similar chemistry (Scheme 2) and proceeded as follows: reaction of ethyl benzyl fumarate, *N*-benzyl glycine and paraformaldehyde gave the N-benzylated *trans*-diester **5**.[30] In a one-pot process, the benzyl-groups were removed by hydrogenation and the free amine was Boc-protected. To assist isolation and purification, the resulting free acid was converted to the sodium salt **6** using sodium ethoxide. Neutralization back to the free acid followed by amide coupling gave the amide-ester **7**. Hydrolysis of the ethyl ester with lithium hydroxide gave the free acid ready for a second amide coupling using polymer-supported reagents. Removal of the Boc-group with TFA and purification by HPLC gave the *trans*-bis-amides **8**.

Compounds from both libraries were assessed for purity by HPLC (>95%) and correct mass by LC–MS (see Supplementary data). In the cis-series, four compounds **3**(**a**–**d**) were initially identified as having sub-10 μM activity (Table 1). In the trans-series, four compounds **8**(**a**–**d**) with sub-10 μM activity were also identified (Table 2). Based on preliminary activity data and ease of synthesis, it was felt that further preliminary optimization should take place in the cis-series. This included variations to the substituent on the basic pyrrolidine nitrogen atom and halogen substitution on the benzyl or phenethyl group of the amide (Scheme 1 and Table 3). Optimization of both cis and trans-series was performed using standard medicinal chemical strategies and is not described here.

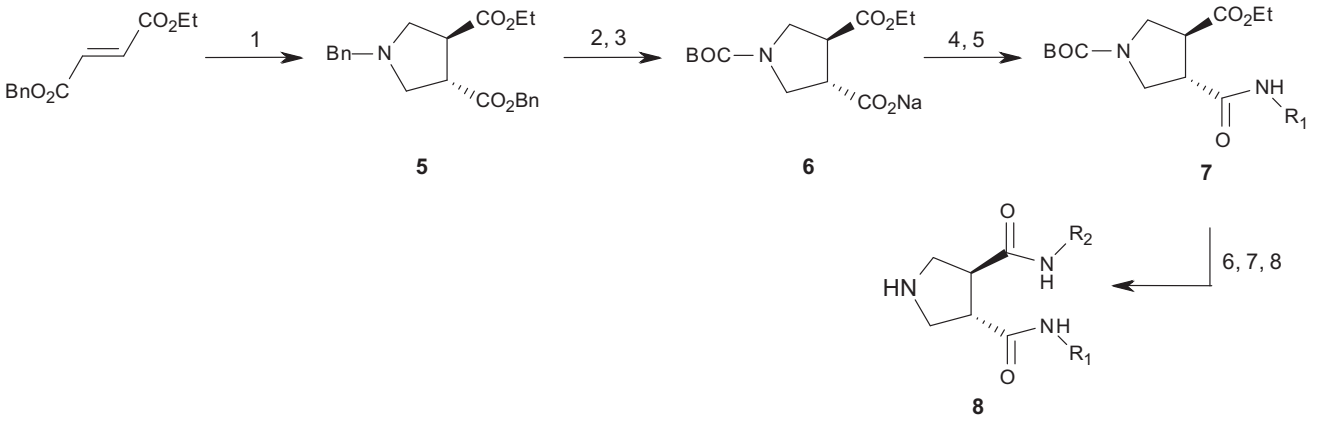
4. Results and discussion

As described above, we selected the His-Phe-Arg-Trp β-turn of the α-MSH peptide for the current validation and chose the pyrrolidine template, which could act as an arginine (lysine) mimic due to its positive charge. For computational, ease the Arg residue of the peptide was replaced in the calculations with a Lys residue, which had been shown to be an adequate replacement in peptide SAR. In the first step of the automated process, a constrained conformational search of the type-I β-turn was performed, leading to 3 significantly different conformations. The 3,4-disubstituted pyrrolidine in both the cis- and trans-configurations was considered as a template for the purposes of de novo design, since the exit vectors from these configurations lead to different potential substitution patterns.

Compound design for the cis-template proceeded in the following manner. The charged nitrogen was anchored to the charged nitrogen of the lysine (arginine) residue, and the de novo engine

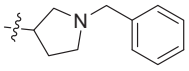
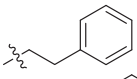
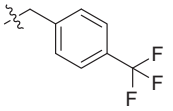
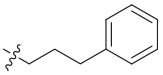
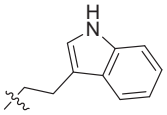
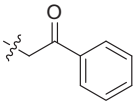
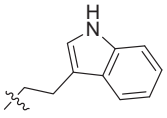
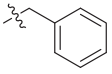


Scheme 1. Synthetic route for preparation of *cis*-pyrrolidine compounds. (1) R_1NH_2, MeCN, rt, 2 h; (2) R_2NH_2, polymer-supported dicyclohexylcarbodiimide (PS-DCC), polymer-supported diisopropylamine (PS-DIPAM), CH_2Cl_2, DMF, rt, 24–48 h; (3) TFA, MeOH rt, 3 h; (4) cyclopropyl carboxaldehyde, $NaCNBH_3$, MeOH, 2 h.



Scheme 2. Synthetic route for preparation of *trans*-pyrrolidine compounds. (1) *N*-benzyl glycine hydrochloride, paraformaldehyde, toluene, 110 °C, 2 h, 56%; (2) hydrogen, $Pd(OH)_2$, di-*tert*-butyl dicarbonate, diisopropylethylamine, 50 atm, rt, 24 h; (3) Na, EtOH, rt 0.5 h, 86% from 5; (4) HCl aq, THF, rt; (5) R_1NH_2, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, 1-hydroxy-benzotriazole, MeCN, rt 16 h; (6) LiOH, MeOH, rt, 2 h; (7) R_2NH_2, polymer-supported dicyclohexylcarbodiimide, 1-hydroxy-benzotriazole, Et_3N, CH_2Cl_2–DMF, rt, 24 h; (8) TFA, MeOH, 3 h.

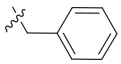
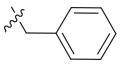
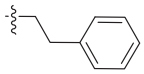
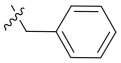
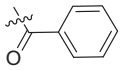
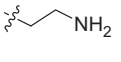
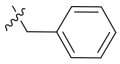
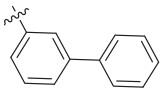
Table 1
Initial hits for the MC4 receptor from de novo design to mimic the His-Phe-Arg-Trp Type-I β-turn of the α-MSH peptide, based on a *cis*-pyrrolidine library

Compound	R_1	R_2	MC4R K_i (μM)	Amino acid residues
3a			5.0	KFH
3b			5.8	KFW
3c			5.5	KFH/KFW
3d			5.5	KFH/KFW

was first applied to grow substitutions at the 3-position with good interactions to any of the other three residues. In order to increase the diversity of the generated compounds, the de novo generation process was run three times with EAI populations of 100 and the top 10 solutions stored from each run. Following this, 15 structures, so-called supertemplates, were selected by manual inspection from the perspective of both alignment and chemistry (ease of synthesis, availability of starting materials) and those shown in Figure 4 were synthesized. Starting from these 6 supertemplates, de novo design process was again applied to find the most promising 4-position substitutions. As before, the de novo process was run multiple times with the results reduced down to a manageable level by a combination of alignment score and manual inspection before attempting synthesis.

We typically prefer to think about de novo design as an idea generator, rather than a source of final molecules.[31] To improve synthetic access for these de novo designed structures, virtual libraries were built for the second amidation step using a short-list of 40 commercially available amines selected from reliable suppliers in ACD,[32] and nearest neighbor similarity analysis to the de novo design solutions was performed to identify the most promising solutions from these libraries. This was followed by running the

Table 2
Initial hits for the MC4 receptor from de novo design to mimic the His-Phe-Arg-Trp Type-I β-turn of the α-MSH peptide, based on a *trans*-pyrrolidine library

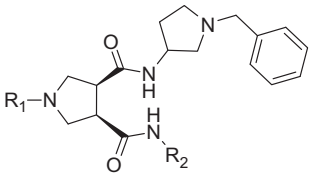
Compound	R_1	R_2	MC4R K_i (μM)
8a			0.9
8b			1.0
8c			5.3
8d			6.6

program in 'virtual screening mode', that is, performing automated flexible alignments, to identify those synthetically feasible solutions that still have sufficiently high scores. Based on these considerations, 51 molecules were proposed for synthesis a selection of which are displayed in Table 1 together with their activity in the MC4R binding assay.

Using a process similar to that described for the *cis*-pyrrolidine library, a library based on *trans*-pyrrolidines was also designed and selected molecules reported in Table 2. No further optimization of this library was attempted but this example demonstrates that the process was able to identify biologically active molecules using a geometrically different starting point and resulting in different preferred substituents given the change in topology.

The next step was meant to check if these novel solutions indeed represent optimizable starting points for chemistry. For reasons given above, we decided to develop the cis-series (**3**, Scheme 1) with an *N*-benzyl-pyrrolidine amide group in one of the two amidic positions. Thus compound **3** was reacted with a range of aldehydes under reductive alkylation conditions being careful to keep the total molecular weight below 550 and ensuring that the nitrogen remains charged. Before synthesis, each of these molecules was aligned to the peptide turn sidechains to ensure that they would still satisfy the initial hypothesis, that is, they still have high scores when overlaid with the peptide turn. This library also included molecules with halogen substitutions on the phenyl group, a medicinal chemistry strategy that was not derived from this process but one that is routinely employed for compound optimization. A total of 47 molecules were synthesized and selected examples are shown in Table 3. The cyclopropyl methyl group seemed to provide a

Table 3
Optimization of the *cis*-pyrrolidine series designed to mimic the His-Phe-Arg-Trp Type-I β-turn of the α-MSH peptide



Compound	R_1	R_2	MC4R K_i (μM)
4a	Cyclopropyl methyl	3-Fluoro phenethyl	1.8
4b	Cyclopropyl methyl	3-Chloro phenethyl	0.67
4c	Cyclopropyl methyl	4-Chloro phenethyl	0.38
4d	Cyclopropyl methyl	2,4-Dichloro phenethyl	0.16
4e	Cyclopropyl methyl	3-Chlorobenzyl	0.40
4f	Cyclopropyl methyl	2,4-Dichloro benzyl	0.07
4g	H	Phenethyl	5.0
4h	Cyclopropyl methyl	Phenethyl	2.6
4i	Hydroxyl ethyl	Phenethyl	6.2
4j	1*H*-Pyrrolo methyl	Phenethyl	1.2
4k	Phenoxyethyl	Phenethyl	1.0
4l	4-Methyl-1*H*-imidazo methyl	Phenethyl	0.9

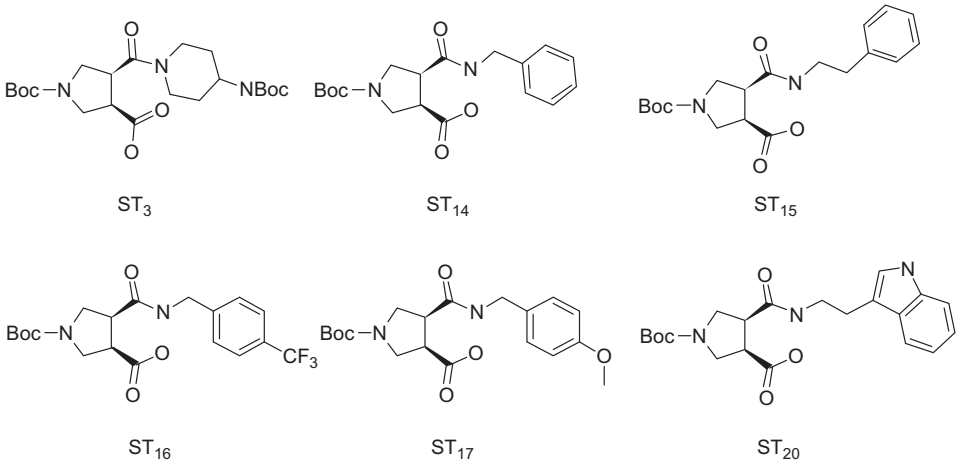


Figure 4. Supertemplates selected as inputs in the second stage of design for the *cis*-pyrrolidine library.

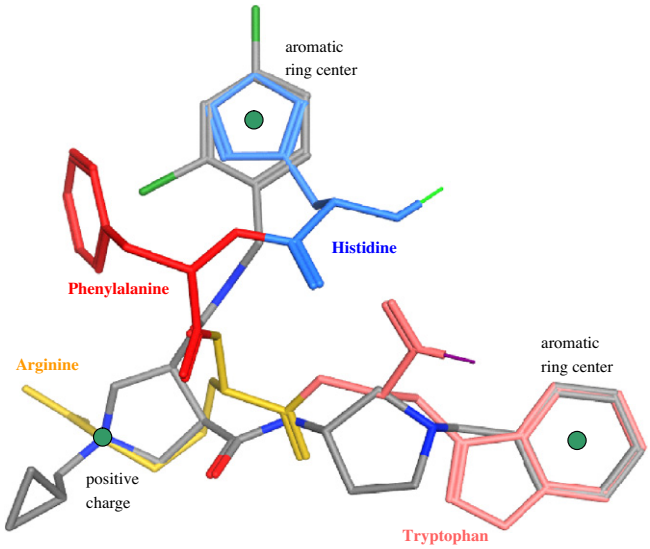


Figure 5. Alignment of the most potent described molecule (K_i = 70 nM) with the modeled peptide β-turn.

balance of promising activity and reasonable properties and molecules containing this group were optimized further. Substitution on the pyrrolidine nitrogen with a small hydrophobic group improved the potency of the compounds two to fivefold, and additional halogen substitution on the benzyl or phenethyl group increased potency further. The cyclopropyl methyl group was chosen in order to generate compounds with the most suitable physico-chemical properties, and when combined with halogenated benzyl or phenethyl groups gave a number of active compounds, the most active of which, **4f**, had an affinity of 0.07 μM.

As mentioned above, all of the described molecules are mimics of the His-Phe-Arg-Trp β-turn of the α-MSH peptide. The alignment of the most active molecule with this peptide turn is displayed in Figure 5. We can see from this alignment that key features of the peptide are in near perfect alignment with the small molecule and this latter follows the peptide turn structure well. Note that having this alignment provides an ideal starting point for ligand-based optimization, however no further attempts were made to improve the binding affinity of these molecules, instead the generation of further derivatives was handed over to the medicinal chemists at this point. Nonetheless, this example still provides a proof of concept that the de novo building process works well enough to generate a feasible starting point for chemistry, and from this starting point the biochemical activity can easily be optimized down to double digit nanomolar K_i values.

5. Conclusions

In this work we described a proof-of-concept validation of an automated process to design peptide turn mimetics. The process involves elements of de novo design and virtual screening. The central design idea is using a probe to overlay with one of the sidechains of the peptide and then grow the ligand to overlay with the sidechains of the other two residues with the hope that it would lead to novel types of structures with a lower molecular weight. In the described validation case, we generated single digit micromolar ligands directly by de novo design, which were rapidly optimized to the double digit nanomolar range, also with the help of the process. This process also helped us generate submicromolar ligands for the GnRH and MCH receptors (details not given).

The described process could potentially be applied in drug discovery to identify small molecule starting points for chemistry by only utilizing known peptide ligands. If the turn position is unknown, the process could also be useful for generating hypotheses by placing the turn sequentially at different positions of the peptide primary sequence. Also, the system was designed so that there is nothing inherent in the process that restricts it to peptide turns, and indeed could be applied to other peptide secondary structures with defined backbone conformations (such as α-helices), making it a potential tool for finding novel small molecule ligands for protein–protein interactions.

6. Experimental

6.1. General experimental

Solvents and reagents were purchased from commercial sources and used without purification. Chromatography was performed on silica gel using the solvent systems indicated. Anhydrous magnesium sulfate ($MgSO_4$) was used to dry the organic layer from extractions and evaporation was achieved in vacuo using a rotary evaporator followed by evacuation with a vacuum pump. HPLC mass spectra (HPLC MS) were recorded on a Hewlett–Packard Series 1100 MSD using an ODS-AQ column and electrospray ionization (CI). Conditions employed were 100% 0.05% TFA/water to 90% MeCN/0.05% TFA/water over 3 min with a total run time of 3.5 min. Purities quoted below represent UV purities recorded at 220 and 254 nm, respectively. Proton NMR spectra were recorded on a Varian Mercury 300 instrument at 300 MHz; chemical shifts were reported in ppm (δ) from an internal tetramethylsilane standard in deuterochloroform. Coupling constants (J) were reported in hertz (Hz) and signals are expressed as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet) or br (broad). Elemental analyses were carried out by NuMega Resonance Laboratories, San Diego, CA.

6.2. HPLC methods

The molecules in this work were purified and analyzed using a range of chromatographic methods and equipment, which are denoted by different names in the text and Supplementary data. These methods are listed below.

SFCPYR4: Agilent Technologies 1100 series chromatograph equipped with atmospheric pressure chemical ionization with mass selective detector (APCI + MSD), diode array detector and a Gilson 215 autosampler. Column: Berger instruments 2-ethyl amino pyridine, 4.6 × 100 mm column, using a gradient of 10% methanol + 0.025% ethane sulfonic acid:90% supercritical carbon dioxide to 60% methanol/supercritical carbon dioxide at 4 mL per minute over 2 min.

LCPurity: Agilent Technologies 1100 series chromatograph with APCI + MSD and diode array detector and autosampler. Column: Phenomenex Synergi MAX-RP 2 × 50 mm 4 mM, 80 Å, using a gradient of 5% acetonitrile + 0.025% trifluoroacetic acid/95% water + 0.025% trifluoroacetic acid to 95% acetonitrile/water at 1 mL per minute over 13 min.

PurityMSQ4: Thermo Scientific MSQ APCI+ single quadrupole mass spectrometer with an Agilent 1100 series diode array detector and a Gilson 215 autosampler. Column: Phenomenex Gemeni-NX C18, 3 × 150 mm column, using a gradient of 5% acetonitrile + 0.2% ammonium hydroxide: 95% water + 0.2% ammonium hydroxide to 95% acetonitrile/water over 9.86 min.

AQA1: ThermoScientific AQA ESI+ single quadrupole mass spectrometer with a Dionex dual wavelength UV–VIS detector and Gilson 215 autosampler, Column: Phenomenex Synergi MAX-RP 2 × 50 mm 4 mM, 80 Å column using a gradient of 5% acetonitrile + 0.025% trifluoroacetic acid/95% water + 0.025% trifluoroacetic acid to 95% acetonitrile/water at 2.5 mL per minute over 5.5 min.

6.3. General methods

6.3.1. Procedure for compounds in the *cis*-pyrrolidine series

A solution of *N*-Boc-pyrrolidine anhydride **1** (1.0 g, 4.1 mmol) and the amine (4.1 mmol) was stirred at room temperature for 2 h. The mixture was evaporated to dryness to give the crude carboxylic acid amide **2** as an off-white solid. A mixture of **2** (0.20 g, 0.5 mmol) and polymer-supported dicyclohexylcarbodiimide (PS-DCC) (0.64 g, 1.0 mmol) in CH_2Cl_2 (15 mL) and DMF (5 mL) was shaken for 30 min at ambient temperature and then the appropriate amine (0.11 g, 0.65 mmol) and polymer supported diisopropylethyl amine (PS-DIPAM) (0.33 g, 1.0 mmol) were added. The mixture was shaken until LC–MS showed the reaction to be complete. The mixture was then filtered and the filtrate evaporated and the residue was dissolved in CH_2Cl_2 (3 mL) and TFA (0.5 mL) and shaken for 3 h. The mixture was diluted with methanol (6 mL) and DMF (2 mL) and purified by automated HPLC (AQA 1 method) to give the TFA salt of the free base **3**. This material was dissolved in methanol (0.5 mL) and Et_3N (0.026 mL) was added, followed by the desired carboxaldehyde (0.015 mL) and a solution of sodium cyanoborohydride (0.010 g) in methanol (0.5 mL). After shaking for 2 h, the resulting mixture was evaporated and the residue purified on HPLC MS MSQ1 to give compound **4**. All compounds reported here are supported by LC–MS and show the correct molecular ion and are >95% pure as witnessed by HPLC with uv detection at two distinct wavelengths.

6.3.2. Procedure for compounds in the *trans*-pyrrolidine series

A mixture of *N*-benzyl glycine hydrochloride (12 g, 60 mmol) and paraformaldehyde (3.6 g, 120 mol) was added in four portions over 1 h to a solution of ethyl benzyl fumarate (9.36 g, 40 mmol) and diisopropyl ethylamine (14.1 mL, 80 mmol) in toluene (150 mL) at reflux. The mixture was heated in a Dean Stark apparatus for 4 h. After cooling, the mixture was filtered through celite and the plug washed with ethyl acetate. The filtrate was washed with water, dried over $MgSO_4$, filtered and evaporated. The residue was purified by silica gel chromatography (7:1 ethyl acetate/hexanes) to give *N*-benzyl-pyrrolidine *trans*-ethyl-benzyl ester **5** (8.2 g, 56%) as a yellow oil.

A mixture of *N*-benzyl-pyrrolidine *trans*-ethyl-benzyl ester **5** (8.0 g, 22.7 mmol), di-*tert*-butyl dicarbonate (6.4 g, 29.5 mmol), palladium hydroxide (1 g) and diisopropyl ethylamine (4.1 mL, 25 mmol) in methanol (70 mL) was shaken under an atmosphere of hydrogen gas at 55 psi in a Parr hydrogenator for 24 h. The mixture was filtered through celite and evaporated and the residue dissolved in ethyl acetate, washed with 10% aqueous citric acid (2×), water and brine then dried ($MgSO_4$), filtered and evaporated to afford the *N*-benzyl-pyrrolidine *trans*-ethyl ester-carboxylic acid (6.3 g) as a pink oil which could not be crystallized. The crude acid was dissolved in THF (50 mL) and a solution of sodium metal (0.5 g, 22 mmol) in ethanol (20 mL) was added slowly via a cannula. The mixture was evaporated to dryness, the residue triturated with diethyl ether and hexanes and filtered to give the sodium salt of the acid **6** (5.8 g, as a yellow solid).

The *N*-benzyl-pyrrolidine *trans*-ethyl ester-carboxylic acid sodium salt **6** was converted to its free acid mixing a suspension of the salt in THF with aqueous 1 M HCl. The mixture was extracted with ethyl acetate, dried ($MgSO_4$) and evaporated.

A solution of *N*-benzyl-pyrrolidine *trans*-ethyl ester-carboxylic acid (6.3 mmol), the amine (7.0 mmol), 1-ethyl-2-(3-dimethylaminopropyl)-carbodiimide (EDC) (8.2 mmol) and *N*-hydroxybenzotriazole (HOBT) (8.2 mmol) was stirred in acetonitrile (20 mL) for 16 h at room temperature. The solvent was evaporated and the residue dissolved in ethyl acetate, washed with water, dried ($MgSO_4$) and evaporated. The residue was purified by silica-gel chromatography to (ethyl acetate/hexanes 3:1 to 2:1) to give the amide ethyl ester **7**.

A solution of the amide ethyl ester (1 mmol) in methanol (5 mL) and 1 M aqueous lithium hydroxide (3 mmol) was shaken at room temperature for 2 h. The methanol was evaporated and the residue diluted with aqueous HCl (1 M) and extracted with ethyl acetate. The ethyl acetate layer was dried with sodium sulfate and evaporated to give the crude acid which was used without further purification

Amine (0.42 mmol) was added to solution of the amide carboxylic acid (0.21 mmol), amine (0.42 mmol), PS-DCC (0.42 mmol), HOBT (0.32 mmol), triethylamine (0.63 mmol) in a mixture of CH_2Cl_2 (3 mL) and DMF (1 mL) and the mixture shaken for 24 h. The mixture was filtered and evaporated and the residue dissolved in CH_2Cl_2 (1 mL). TFA (0.5 mL) was added and the solution shaken for 3 h and then evaporated. The residue was purified by HPLC (MSQ4) to give the bis-amide **8**. All compounds reported here are supported by LC–MS and show the correct molecular ion and are >95% pure as witnessed by HPLC with uv detection at two distinct wavelengths.

6.4. Procedure for the synthesis of *rac*-(3*R*, 4*S*)-N^3-(1-benzyl pyrrolidine-3-yl)-1-(cyclopropylmethyl-N^4-(2,4-dichlorobenzyl) pyrrolidine-3,4-dicarboxamide 4f

A solution of *N*-Boc-pyrrolidine anhydride (1.0 g, 4.1 mmol) and benzyl-3-amino-pyrrolidine (0.73 g, 4.1 mmol) was stirred at room temperature for 2 h. The mixture was evaporated to dryness to give the crude carboxylic acid amide **2a** (1.7 g, 100%) as an off-white solid which was used directly without further purification.

A mixture of **2a** (0.031 g, 0.075 mmol), 1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride (EDC) (0.029 g, 0.15 mmol), *N*-hydroxybenzotriazole (HOBT) and 2,4-dichlorobenzylamine (30 μL, 0.20 mmol) in CH_2Cl_2 (1 mL) was shaken at room temperature when LC–MS showed the reaction to be complete (5 h). Saturated sodium hydrogen carbonate solution (1 mL) was added and the mixture shaken and the organic layer was filtered through a plug of sodium sulfate the solvent evaporated. The residue was dissolved in CH_2Cl_2 (0.3 mL) and TFA (0.3 mL), shaken for 8 h and then the solvent evaporated. The residue was dissolved in methanol (1 mL) and triethylamine (30 mL, 0.225 mmol), cyclopropane carboxaldehyde (30 mL, 0.3 mmol) and sodium cyanoborohydride (10 mg, 0.15 mmol) added. The mixture was shaken at room temperature for 5 h and then purified by HPLC to give compound **4f** as a clear oil.

LC–MS 1 Purity (MSQ4) 220 (100%); 254 (100%); TIC 100% (529.2).

^{1}H NMR (DMSO-d_6) δ = 10.33 (m, 1H), 8,74 (m, 1H), 7.60 (s, 1H), 7.57-7.30 (m, 7H), 4.31 (m, 2H), 3.77 (m, 1H), 3.50 (m, 4H), 3.42 (m, 2H), 3.36 (m, 2H), 3.16 (m, 2H), 3.03 (m, 2H), 2.29 (m, 2H), 1.51 (m, 1H), 1.08 (m, 1H), 0.85 (br t, J 6.7 Hz, 1H), 0.60 (m, 2H), 0,38 (m, 2H).

HRMS (ESI$^+$) *m*/*z* Calcd for $C_{28}H_{35}O_2N_4Cl_2$ $(M+H)^+$ 529.2132; found 529.2117.

6.5. Biological assay

Receptor binding determinations were performed in HEK293 cells stably expressing the human MC4 receptor using [^{125}I]-NDP-MSH as the radiolabeled ligand. Typically, data for compounds of

interest (K_i <1 μM) are the average of two independent experiments. Details are reported elsewhere.[33]

Acknowledgements

Dimitri Grigoriadis and John Williams (Neurocrine Biosciences) are gratefully acknowledged for allowing us to retrieve synthesis details from NBI archives.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bmc.2012.04.001.

References and notes

1. Grauer, A.; König, B. *Eur. J. Org. Chem.* **2009**, *30*, 5099.
2. Vagner, J.; Qu, H.; Hruby, V. J. *Curr. Opin. Chem. Biol.* **2008**, *12*, 292.
3. Kieber-Emmons, T.; Murali, R.; Greeneieber-Emmons, M. I. *Curr. Opin. Biotech.* **1997**, *8*, 435.
4. Olson, G. L.; Bolin, D. R.; Bonner, M. P.; Bös, M.; Cook, C. M.; Fry, D. C.; Graves, B. J.; Hatada, M.; Hill, D. E.; Kahn, M.; Madison, V. S.; Rusiecki, V. K.; Sarabur; Sepinwall, J.; Vincent, G. P.; Voss, M. E. *J. Med. Chem.* **1993**, *36*, 3039.
5. Peczuh, M. W.; Hamilton, A. D. *Chem. Rev.* **2000**, *100*, 2479.
6. Loughlin, W. A.; Tyndal, J. D. A.; Glenn, M. P.; Fairlie, D. P. *Chem. Rev.* **2004**, *104*, 6085.
7. Yin, H.; Le, G. I.; Hamilton, A. D. Alpha-helix Mimetics in Drug Discovery in Drug Discovery Research: new Frontiers in the Post-Genomic Era In Huang, Z., Ed.; Wiley: Hoboken, 2007; pp 281–299.
8. Suat, K.; Jois, S. D. S. *Curr. Pharm. Des.* **2003**, *9*, 1209.
9. Tyndal, J. D. A.; Pfeiffer, B.; Abbenante, G.; Fairlie, D. P. *Chem. Rev.* **2005**, *105*, 793.
10. Holder, J. R.; Haskell-Luevano, C. *Med. Res. Rev.* **2004**, *24*, 325.
11. Adan, R. A. H.; Tiesjema, B.; Hillebrand, J. J. G.; la Fleur, S. E.; Kas, M. J. H.; de Krom, M. *Br. J. Pharmacol.* **2006**, *149*, 815.
12. Xi, N. *Drugs Future* **2006**, *31*, 163.
13. Dyck, B.; Parker, J.; Phillips, T.; Carter, L.; Murphy, B.; Summers, R.; Hermann, J.; Baker, T.; Cismowski, M.; Saunders, J.; Goodfellow, V. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3793.
14. Chen, C.; Tran, J. A.; Arellano, M.; Fleck, B. A.; Pontillo, J.; Marinkovic, D.; Tucci, F. C.; Wen, J.; Saunders, J. *Med. Chem.* **2008**, *4*, 67.
15. Jiang, W.; Tucci, F. C.; Tran, J. A.; Fleck, B. A.; Wen, J.; Markison, S.; Marinkovic, D.; Chen, C. W.; Arellano, M.; Hoare, S. R.; Johns, M.; Foster, A. C.; Saunders, J.; Chen, C. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5610.
16. Bednarek, M. A.; MacNeil, T.; Tang, R.; Kalyani, R. N.; Van der Ploeg, L. H.; Weinberg, D. H. *Biochem. Biophys. Res. Commun.* **2001**, *286*, 641.
17. Hruby, V. J.; Wilkes, B. C.; Hadley, M. E.; Al-Obeidi, F.; Sawyer, T. K.; Staples, D. J.; DeVaux, A. E.; Dym, O.; Castrucci, A. M. L. *J. Med. Chem.* **1987**, *30*, 2126.
18. Bednarek, M. A. *Chem. Today* **2009**, *27*, 28.
19. Todorovic, A.; Holder, J. R.; Scott, J. W.; Haskell-Luevano, C. *J. Peptide Res.* **2004**, *63*, 270.
20. Lee, J. H.; Lim, S. K.; Huh, S. H.; Lee, D.; Lee, W. *Eur. J. Biochem.* **1998**, *257*, 31.
21. Marschall, G. R. *Tetrahedron* **1993**, *49*, 3547.
22. Morimoto, H.; Fukushima, C.; Yamauchi, R.; Hosino, T.; Kikkawa, K.; Yasuda, K.; Yamada, K. *Bioorg. Med. Chem.* **2001**, *9*, 255.
23. Doyle, P. M.; Harris, J. C.; Moody, C. M.; Sadler, P. J.; Sims, M.; Thornton, J. M.; Uppenbrink, J.; Viles, J. H. *Int. J. Pept. Protein Res.* **1996**, *47*, 427.
24. MOE (Molecular Operating Environment) version 2005.06; Chemical Computing Group Inc.: 1010 Sherbrooke St. West, Suite 910, Montreal, Quebec, H3A 2R7, Canada.
25. EA Inventor, 4.1, Optive Research, Austin, TX, 2004; currently marketed by Tripos Inc., St. Louis, MO, http://www.tripos.com.
26. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.
27. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
28. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005.
29. Labute, P.; Williams, C.; Feher, M.; Sourial, E.; Schmidt, J. M. *J. Med. Chem.* **2001**, *44*, 1483.
30. Xue, C.-B., Decicco, C.P. and He, X. PCT Patent WO02055491A, 2006.
31. Feher, M.; Gao, Y.; Baber, J. C.; Shirley, W. A.; Saunders, J. *Bioorg. Med. Chem.* **2008**, *16*, 422.
32. ACD, version 2005. MDL Inc.; currently marketed by Accelrys Inc., San Diego, CA, http://www.accelrys.com.
33. Nickolls, Sarah A.; Fleck, Berth; Hoare, Sam R. J.; Maki, Rich A. *J. Pharm. Exp. Ther.* **2005**, *313*, 1281.