Paper 2

MR. BALMUKUND SURESHKUMAR THAKKAR (Orcid ID : 0000-0002-7831-0709)

PROF. RICHARD ALAN ENGH (Orcid ID : 0000-0002-6207-0560)

Received Date : 24-Oct-2016

Revised Date : 19-Dec-2016

Accepted Date : 21-Dec-2016

Article type : Research Article

Biofocussed chemoprospecting: An efficient approach for drug discovery

Balmukund S. Thakkar[a], Marte Albrigtsen[b], John-Sigurd M. Svendsen[a], Jeanette H. Andersen[b], Richard A. Engh[a]*

[a]Department of Chemistry, UiT The Arctic University of Norway, N-9037 Tromsø, Norway

[b]Marbio, UiT The Arctic University of Norway, Breivika, N-9037, Tromsø, Norway.

*Corresponding author:
Richard A. Engh
Department of Chemistry, UiT The Arctic University of Norway, N-9037 Tromsø, Norway
+47 77644073
Richard.Engh@uit.no

Running title: Biofocussed chemoprospecting

Key words: Drug discovery, Peptidomimetic, Cheminformatics, Chemoprospecting, Small molecule diversity

Abstract

Drug discovery strategies include from broad random screening to focussed target-based approaches. Structure and substrate information greatly enable target-based design, but this is limited to relatively few targets; cell-based screening can identify new targets but often suffers from low hit rates and difficult hit optimization. Thus, newer approaches are needed that can improve the efficiency of screening and hit optimization. Here we describe an efficient approach for hit-generation, which may be called "biofocussed chemoprospecting". With bio-likeness and ease of synthesis as priority criteria, libraries may be constructed with good optimization potential, physicochemical diversity, drug-likeness, and low cost. Following this approach, two libraries based on linear and cyclic dipeptide scaffolds were designed, first as virtual libraries comprising of more than 30000 compounds, and after subsequent filtering, as a small library of a total of 51 compounds. These provided good diversity at low cost, and were tested for bioactivities. The discovery of six active compounds demonstrates a hit-rate greater than 10%. This is comparable to target based approaches, but “chemoprospecting” described here has the additional potential to identify new targets and mechanisms.

Introduction

New drug discovery and development is typically long and costly process, involving enormous resources (1, 2) of time (8 to 12 years), money (more than a billion dollars per new approved compound), labour and intellect. The recent rise in the cost of drug discovery has also been accompanied by an increased attrition rate of drug candidates in clinical trials, with concomitant declines in pharma R & D productivity (3–5). However, it has also been observed (6, 7) that improved and more focussed preclinical research may decrease the chances of drug discovery failure.

Hit-finding approaches and library design

Conventional methods for hit-discovery and lead optimization may be distinguished into three basic approaches, each with its own basic focus (**Figure 1**):

Target-based approaches focus on specific disease mechanisms via validated drug-targets. They are considered relatively low-risk, and have dominated projects (8) in both, academia and industry over last decade as “structure based” (8–10) or “fragment based” (11–13) approaches with or without high-throughput screening. The lower risk comes at the cost of a lack of novelty in targets, including overexploitation of targets, with concomitant lowering of market potential and crowded patent space. It also provides fewer options for resistance-prone therapy such as cancer and antimicrobial treatments.

Scaffold-based approaches focus on pharmacophoric features of known bioactive molecules and/or efficient syntheses, hence tend to act via the same (known or unknown) target/mechanism as the known active molecules do (14–17).

Bioprospecting approaches aim to discover natural bioactivities that may also intervene in disease mechanisms (18–20). However, complexities arise at early stages, such as the isolation and identification of the bioactive molecule(s) and target determination, followed by difficulties of structure elucidation, synthetic route development and scale-up to commercial dimensions, which combine to undermine the bioprospecting approach (21). However, despite the limitations, bioprospecting is an important approach especially as it may identify new drug targets, mechanisms and chemical scaffolds.

In general, library design requires careful balance of synthetic feasibility, structural diversity and diversity of physicochemical properties for efficient hit finding. It has been observed that, while size is important, diversity in the library is also necessary, depending on its type (22–24). Further, replacing random diversity with a biogenic biased selection of compounds also greatly enhances the efficiency (25).

Biofocussed chemoprospecting

The fact that the total number of drug targets via which all FDA approved drugs act is only 324 (26, 27) indicates the huge potential to discover and utilize newer drug-targets and mechanisms. While a bioprospecting approach may be useful in this respect, its synthetic chemistry bottlenecks are typically prohibitive. This can be rectified by adapting a scaffold-based approach, choosing to synthesize compounds with bio-like scaffolds. This leads to a hybrid approach, which can be termed "**biofocussed chemoprospecting**". In other words, it can be described as diversity oriented synthesis (28) of bio-relevant scaffold based libraries from a small set of starting materials, and their use in bioactivity-screening to identify novel types of bioactivity. From a synthetic chemistry point of view, this approach not only uses diversity-oriented synthesis to create diversity from same starting materials, but also ensures that the diversity remains within the scope of bio-likeness.

Efficiency parameters

In order to increase the efficiency of finding good optimizable hits, the strategy was to establish specific efficiency parameters (**Figure 2**) for the libraries. The properties of both the scaffold and the substitutions are important for selection of compounds. It is desirable to have scaffolds with significant similarity to biomolecules. It is further important that substitutions also offer surface properties statistically compatible with pharmacophoric properties of biological targets.

Overall strategy

We aimed to develop the libraries of compounds that would preferably combine similarity to biomolecules with physicochemical criteria for "drug-likeness". Each library would possess a scaffold with more than one variable position, each allowing a good range of substitutions to diversify the pharmacophoric properties. Preferably, the synthesis of libraries would be easy with simple and small number of reaction steps, using cheap and safe reagents that are commercially available in wide variety, and preferably avoiding harmful/dangerous reagents.

Peptide structure as the starting point

With bio-likeness as a primary criterion for our library design, the versatility of peptides in nature prioritizes molecules with structural similarities to peptides to have good potential for activity in biological systems.

Among peptides, the simplest peptide structure is a dipeptide, which may be either linear or cyclic. A cyclic dipeptide would be piperazine-2,5-dione (**Figure 3**). One of the two nitrogen atoms may be substituted with an alkyl/aryl substitution to introduce fragments with diverse pharmacophoric features. Such substituted piperazine-2,5-diones can

be synthesized from corresponding *N*-substituted linear dipeptide esters, which in turn could be synthesized from starting materials such as aldehyde, amino acid and amino acid esters. Thus, congruent with the concept of diversity oriented synthesis, use of the same starting materials can provide two different libraries based on peptides: *N*-substituted dipeptide esters and piperazine-2,5-diones, as shown in **Figure 4**.

A few examples (34–36) of related *N*-substituted dipeptide esters were found but they were structurally different and there were no reports of any biological activity. They were synthesized via reductive amination of amino acid esters followed by peptide coupling with *N*-protected amino acid, and therefore had substitution on peptide bond nitrogen (**Figure 5**).

Piperazine-2,5-dione as a scaffold has been well established (37–44) for its bioactivity potential. A few compounds similar to our library have also been reported. Szardenings et al. used the zinc binding property of thiol group to synthesize thiol containing piperazine-2,5-dione derivatives (45–47) as matrix metalloproteinase inhibitors. Similarly, piperazine-2,5-dione derivatives as dual inhibitors of farnesyltransferase and geranylgeranyltransferase-1 were reported by Qiao et al. (48). They were similar to our designed library but they had substitutions only on one carbon of piperazine-2,5-dione scaffold, while our library was designed to include diverse substitutions from both carbons without any specific target considerations.

Materials and methods

Chemistry

General methods

All reagents and solvents were purchased from commercial vendors such as Aldrich and were used as received without further purification. The reactions were monitored by thin-layer chromatography (TLC) analysis on precoated Merck® silica gel 60 F_{254} TLC aluminum sheets and/or mass spectrometry. The spots on TLC were visualized by exposure to ultraviolet (UV) light (254 nm), or by staining with iodinated silica gel powder, or 5% ninhydrin solution in ethanol followed by heating. Column chromatography was carried out using silica gel 60 (230-400 mesh). Preparative thin layer chromatography (PTLC) was performed on precoated Merck® silica gel 60 F_{254} TLC aluminum sheets. For TLC as well as column chromatography, 0-100 % of ethylacetate in pentane or 0-15 % of methanol in dicholoromethane (DCM) were used as common mobile phase to elute compounds in increasing order of polarity. Microwave assisted reactions were carried out using Biotage® Initiator 300 W instrument. Spectroscopic methods such as ^{1}H NMR, ^{13}C NMR, MS and IR were used to confirm the products. NMR spectra were recorded at 400 MHz (^{1}H NMR) and at 100 MHz (^{13}C NMR) using Varian spectrometer. The high resolution mass spectrometric analysis (HRMS) were carried out on an LTQ Orbitrap XL (Thermo Scientific, Bremen, Germany) using positive mode ESI. The IR spectra were taken using Varian 7000e FT-IR spectrometer.

Synthesis of *N*-substituted amino acid

As per a reported protocol (49), an aldehyde (1.4 equiv.) was added to a stirred solution of an amino acid (1 equiv.) neutralized with finely powdered NaOH (1.05 equiv.) in methanol. After imine formation for 30 min to 2 h at room temperature, the solution was cooled on an ice-bath followed by slow addition of sodium borohydride or sodium cyanoborohydride (1.3 equiv.). The reaction mixture was stirred under inert nitrogen atmosphere for 2 h at room temperature followed by acidification with 37 % aq. HCl up to isoelectric pH of the amino acid. The solvent was then evaporated at reduced pressure. The solid mixture was triturated with acetone and filtered. The residue after filtration

was dried unless hygroscopic. The crude *N*-substituted amino acid product thus obtained was used in next step without further purification.

Synthesis of *N*-substituted dipeptide esters

A crude *N*-substituted amino acid (1 equiv.) was mixed in acetonitrile into which triethylamine (4.8 equiv.) and HBTU (1.2 equiv.) were slowly added sequentially. After 10 min stirring at room temperature, amino acid ester (1 equiv.) was added. The reaction mixture was stirred for about 20-30 min at room temperature and was monitored on TLC. After completion, the solvent was evaporated under reduced pressure. The crude reaction mixture was purified using column chromatography to yield *N*-substituted dipeptide ester product. When necessary, the product was further purified on preparative TLC.

Synthesis of piperazine-2,5-diones

An *N*-substituted dipeptide ester compound was mixed in water. The mixture was heated under microwave irradiation under microwave irradiation as per a reported protocol (35) with customized time (30 min to 5 h) and temperature to provide piperazine-2,5-diones. After completion, the solvent was evaporated to dryness under reduced pressure followed by isolation using column chromatography and/or preparative TLC.

Chemoinformatic analysis of libraries

Virtual libraries for both series were generated using the JChem software suite (ChemAxon). Chemoinformatic analysis of virtual libraries and synthesized compounds was carried out. Properties such as molecular weight, AlogP, hydrogen bond donor and acceptors, number of rotatable bonds, polar surface area, toxicity were calculated using the Canvas module of Schrodinger Suite and their distribution was analysed. Drug-likeness and clogP were estimated with Datawarrior software from Openmolecules.

Cost calculation

As a standard, the necessary amount of starting materials to carry out reaction at 1 mmol scale was calculated considering the mole equivalents. The estimated costs include only the cost of aldehyde, amino acid and amino acid esters. Based on the amount, the prices for each starting material were generated as per commercial catalogues of Aldrich and other suppliers (if the starting material was not available at Aldrich). From the cost of each starting material, the costs for all compounds from both virtual libraries were estimated. The costs of isolation, purification or characterization costs were not included because such calculations would be unpredictable and can show huge variation from one reaction to another.

Bioactivity assays

Kinase profiling studies were carried out at International centre for kinase profiling (50) and DiscoverX KINOMEscan (51). Cell viability (MTS) assays, anti-oxidant activity (CLPAA) assays, antibacterial activity assays and biofilm assays were conducted at MarBio, Tromsø.

Cellular lipid peroxidation antioxidant activity (CLPAA) assay

Approximately 90 000 HepG2 cells per well were seeded in black 96 well plates with clear bottoms (# 3603, Corning, NY, USA) and incubated overnight. The cells were labelled with 10 μM C11-BODIPY (#D3861, Invitrogen, Eugene, OR) for 30 min and incubated for 1 h with various concentrations of the test compounds. 50 μM Cumene hydroperoxide (cumOOH, #247502, Sigma-Aldrich, St.Louis, MO) was added to initiate lipid peroxidation and the plate was immediately placed in a Victor3 Plate Reader (Perkin Elmer, MA, USA). Both red (590/7 nm (exitation), 632/45 nm

(emission)) and green (485/14 nm, 520/10 nm) fluorescence was recorded every 3rd minute during ~1 h. Cells were washed with PBS between additions of new reagents. The total reaction volume was 100 µl. All incubations were carried out at 37 ºC with 5 % CO_2. Percent inhibition was calculated relative to the positive control (cumOOH without test compound).

Cell viability assay (MTS).

Cell viability was determined by a colorimetric [3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium] (MTS) assay.

The compounds were tested against three cancer cell lines: human melanoma (A2058 ATCC CRL-1147), human breast carcinoma (MCF7 ATCC HTB-22), and human colon carcinoma (HT29 ATCC HTB-38). In addition, non-malignant lung fibroblasts (MRC5 ATCC CCL-171) was used as toxicity control.

The cell lines were seeded in 96-well microtiter plates at 2000 (cancer cell lines) or 4000 (MRC5) cells/well. After 24 h incubation at 37 ºC in 5 % CO_2, the compounds were added in triplicates to each cell line to give 50 µM as the test concentration. The plates were incubated for 72 h. At the end of the exposure time, 10 µl Cell Titer 96® Aqueous One Solution Reagent (Promega, USA) was added to each well, and the plates were incubated for 1 h before absorbance was measured using DTX multimode detector (Beckman Coulter, INC CA92821 USA) at 485 nm. Cells treated with Roswell Park Memorial Institute-1640 medium were used as negative control, and compound effect was quantified as the percentage of control absorbance of reduced dye.

Antibacterial assay

The antibacterial activity was tested on five different strains; *E. faecalis* (ATCC 29212), *E. coli* (ATCC 25922), *P. aeruginosa* (ATCC 27853), *S. aureus* (ATCC 25923) and *Streptococcus agalactiae* group B (ATCC 12386). Growth medium with sterile MilliQ H_2O was used as a negative control while sterile MilliQ H_2O and bacteria suspension was used as a positive control. Bacteria were transferred from a blood plate to growth medium (MH-bullion #275730 Difco Becton Dickinson) for *E. coli*, *P. aeruginosa* and *S. aureus* and BHI-bullion (#53286 Sigma Aldrich) for *E. faecalis* and *S. agalactiae* gr. B) and incubated at 37°C overnight. The following day a part of the bacteria suspension was transferred to fresh medium and cultivated in a shaker incubator at 37°C for 1.5 h (*E. coli, E. faecalis* and *Streptococcus* gr. B) or 2.5 h (*S. aureus* and *P. aeruginosa*). The bacteria suspension was then diluted 1:100 in medium and added all wells in a 96-well microtiter plate (Nunc 167008), followed by test compounds in duplicates. The plates were incubated at 37 ºC overnight before growth was assessed visually and photometrical at 600 nm. The total reaction volume was 100 µL. Compounds were tested at 50 µM

Biofilm inhibition assay

S. epidermidis (RP62A 42-77, ATCC 35984) was used to assess the effect of the test compounds on biofilm formation. Growth media: tryptic soy broth (TS; #1.05459 Merck, Darmstadt, Germany). An overnight culture of *S. epidermidis* grown in TS was diluted with fresh TS containing 1 % glucose (1:100). Aliquots of 50 µL were transferred to a 96-well microtiter plate, and 50 µL of test compounds, dissolved in water at ranging concentrations, was added. After overnight incubation at 37 °C, the bacterial suspension was carefully discarded and the wells washed with water. The plate was dried and the biofilm fixed by incubation for 1 h at 55 °C before the surface attached cells were stained with 100 µL of 0.1 % crystal violet for 5 min. The crystal violet solution was removed and the plate once more washed with water and dried at 55 °C for 1 h. After adding 70 µL of 70 % ethanol, the plate was incubated at room temperature for 10 min.

Biofilm formation was observed by visual inspection of the plates. The MIC was defined as the lowest concentration where no biofilm formation was visible. A *S. epidermidis* suspension, diluted with 50 µL of water, was used as a positive control, and 50 µL *Staphylococcus haemolyticus* (clinical isolate 8-7A) suspension with 50 µL of water was employed as a negative control. A mixture of 50 µL water and 50 µL TS was used as assay control.

Results and discussion

Chemistry

Libraries of *N*-substituted dipeptide esters and piperazine-2,5-diones were created in just two and three steps respectively using aldehyde, amino acid and amino acid as starting materials as described in **Scheme 1**.

For couplings using uronium coupling agents such as HBTU, the reactivity difference (52) between 2°-amine (of *N*-alkylated amino acid) and 1°-amine (of amino acid esters) provided an opportunity to skip protection-deprotection steps. Therefore, opting for solid phase synthesis (as in previous examples) was no longer required, decreasing the number of steps and making the synthesis easier, cheaper and more compliant with our approach.

A total of 29 compounds from *N*-substituted dipeptide esters series were generated in two steps. The products were subjected to a step of cyclization with a microwave assisted protocol (35) after customization for highly substituted dipeptide esters with secondary amines as a nucleophile. Thus, a total of 22 compounds from piperazine-2,5-diones series were generated in three steps.

Analysis of efficiency parameters

Biofocus

Both libraries, *N*-substituted dipeptide esters and piperazine-2,5-diones, were based on linear and cyclized peptide scaffolds respectively. As discussed earlier, the basis of the selection of scaffolds itself was their strong bio-likeness.

Structural diversification potential

The starting materials for both series are aldehydes, amino acids and amino acid esters. Among them, aldehydes and amino acids are commercially available in wide varieties, including many in enantiomeric pure forms. Amino acid esters are also commercially available or easily accessible from corresponding amino acids (53, 54). A recent Aldrich catalogue mentions 1244 aldehydes and 551 amino acids. Considering 1200 aldehydes, 500 amino acids and 500 corresponding amino acid esters (only methyl esters), the number of total theoretically accessible compounds becomes 1200 * 500 * 500 = 300 million compounds for each series!

Skeletal (scaffold) diversification

While both libraries were synthesized from the same starting materials, *N*-substituted dipeptide esters and piperazine-2,5-diones are scaffolds with quite different topological features, surface area, rotatable bonds and adaptable conformations.

Appendage and functional group diversification

For both series, R^1 comes from an aldehyde; R^2 comes from an amino acid, while R^3 (and R^4) comes from an amino acid ester. All three starting materials are commercially available in wide varieties, corresponding to different appendages including acyclic, carbocyclic or heterocyclic chains and functional groups including carboxylic acid, halides, amide,

hydroxyl, ether, ester, cyano, phenol, thiol, amino group etc. Thus, overall appendage and functional group diversification potential is high.

Stereochemical diversification

Both libraries possess two chiral centres as a part of scaffolds, viz. contributed by the alpha carbons of amino acid and amino acid ester. Each chiral centre can be utilized to introduce desired chirality, either as an enantiomeric pure form or as racemic mixture that would include both enantiomers for the chiral centre, paving way for up to four stereoisomers for each compounds.

Diversity of properties and drug-likeness

Generation of virtual libraries

With directly accessible chemical space of more than 300 million compounds for each library, it would be computationally exhaustive to process chemoinformatic analysis. Hence, its small subset as a virtual library was generated and such created virtual library was analysed with chemoinformatic tools.

As Mao et al. have noted (55), an overwhelming majority of approved drug molecules contain at least one aromatic ring. However, not all amino acids and amino acid esters are aromatic. Hence, out of more than 1200 varieties of aldehydes, 22 aromatic aldehydes were chosen with simple carbocycles or heterocycles such as phenol, thiophene, furan, pyridine, indole, quinoline, imidazole. Out of more than 500 varieties of amino acids and amino acid esters, only natural stereoisomers (L-isomer) and their unnatural enantiomers (D-isomer) – except for achiral glycine were chosen. Being an imino acid, both isomers of proline were excluded from amino acids and amino acid esters. Thus, both virtual libraries of 22*37*37 = 30118 compounds each were generated and subjected to chemoinformatic analysis and distributions of various properties were studied.

Analysis of common properties of both libraries

The diversities of both virtual libraries were analysed with respect to various physicochemical parameters such as molecular weight, AlogP, hydrogen bond donor and acceptors, number of rotatable bonds, polar surface area etc. It can be observed from **Figure 6** to **Figure 11** that the virtual library shows promising and diverse distribution patterns for different parameters. While the absolute values for a given property differ significantly for an *N*-substituted dipeptide ester and its corresponding piperazine-2,5-dione compound, the relative changes in the property value are observed to be similar for such pairs, as expected given the conservation of the structural transformation of the paired compounds.

Figure 12 shows the correlation between molecular weight and clogP of compound libraries. Strongly negative clogP values represent extreme polar character, which is often counter-productive for drug development. Hence, the synthesized compounds were carefully chosen to be moderate-polar in order to ensure good balance between solubility and absorption in biological systems. Further, significant diversity of physicochemical properties of synthesized compounds can be observed as scattered red points around the chart, indicating an efficient selection. Similarly, **Figure 13** presents drug-likeness with respect to molecular weight and clogP for both libraries, calculated using Datawarrior software, which uses 5300 distinct substructure fragments with associated drug-likeness scores (56).

The overall drug-likeness of the virtual library compounds was found to be significantly high, and the distribution of synthesized compounds from both libraries was also found to be very efficient from diversity aspects.

Ease of synthesis

As described above, *N*-substituted dipeptide esters and piperazine-2,5-diones can be synthesized from commercially available materials in just two and three steps respectively. All steps are very safe. The first step does not require purification. The last step of piperazine-2,5-diones synthesis (cyclization step) does not require any reagent except solvent i.e. water.

With multistep synthesis being like a "norm" in library synthesis for drug discovery, it can be safely concluded that the overall ease of synthesis is significantly high for both libraries.

Cost and availability of starting materials

The starting materials for both libraries are aldehydes, amino acids and amino acid esters. They are cheap and commercially available in wide varieties.

As both series have share same starting material and only water is used in the last step, the cost for corresponding compounds from both libraries were same. The plot in **Figure 14** shows distribution of *N*-substituted dipeptide ester compounds on three parameters: molecular weight, clogP and cost of starting materials. It is evident from the plot that the distribution of synthesized *N*-substituted dipeptide library is reasonably diverse with respect to molecular weight and clogP while being skewed towards low cost, implying that reasonable diversity could be achieved at low cost.

Optimum substitution variations for synthesis

For optimum synthesis, it is important to synthesize libraries in such a way that despite being small, the library would provide much more information about structure activity relationships with respect to pharmacophoric features associated with substitutions.

Six aromatic aldehydes containing monocyclic or bicyclic carbocycles/heterocycles were chosen. Four amino acids and seven amino acid esters with different size, chirality and characteristics of side chains were selected. As shown in **Table 1**, the substitutions with specific pharmacophoric features were varied to generate libraries of synthesized compounds so that each position would provide a primary structure activity relationship.

When R^2 is C and R^3 is d, 4 variation for R^1 as 1, 2, 3 and 5 respectively are present. When R^1 is 1 and R^3 is a, 3 variations for R^2 as A, B and C are present. Similarly, when R^1 is 5 and R^2 is C, 7 variations for R^3 as a to g are present. Further, activity trends for such variations can also be evaluated with respect to more than one pairs. For example, variations of R^3 can be evaluated for different R^1-R^2 pairs such as 1-C, 3-C, 4-C, 5-C or 2-A, or variations of R^1 can be evaluated for different R^2-R^3 pairs such as C-a, C-b, C-d or C-e.

Bioactivity studies

Both libraries provided interesting hits, as hoped from their design as bio-like compounds. The studies included both cell-based screening and enzyme family profiling. An outline of the results is presented in **Figure 15**. The detailed results will be published at a later stage with specific focus on individual activities of both series.

Overall, one compound from piperazine-2,5-dione series was found to inhibit brain specific kinase 1 (BRSK1), while one more compound from piperazine-2,5-dione series and four compounds from *N*-substituted dipeptide ester series were found to cause viability loss on one to three cancer cell lines. Follow up cell viability studies of active anti-cancer compounds identified one *N*-substituted dipeptide ester that elicited cell morphology changes at a concentration of 5 μM **(Figure 16)**. The single-digit micromolar activity is a level promising for subsequent optimization (57) for which target identification will be a high priority. The type of morphology changes may also assist in target identification (58). None of the active compounds were predicted to show any toxicity as per the computational models from Datawarrior (59), nor did they violate Lipinski's rule of five. Considering that both libraries were not designed with any specific target in focus, the successful hit-finding for both series, even with very small libraries, serves as an affirmation of the efficiency by biofocussed diversity.

Hit-rates comparison

Hit-rates for diverse screening libraries are typically quite low (in range of < 0.1 %), while target-based approaches supported by X-ray crystallography, and computational methods such as virtual screening, docking, pharmacophore modelling etc. have been shown to provide much higher hit-rates (60–63). In our case, we found 6 biologically active compounds from two libraries comprising of total 51 compounds, which makes the hit-rate greater than 10 %, which is comparable to the hit-rates of target-based approaches, and shows at least 100 fold betterment over typical diverse screening libraries. The betterment can be attributed essentially to the efficient library design as an outcome of focus on bio-likeness and optimized diversity of physicochemical properties – the core ideas of the biofocussed chemoprospecting approach.

Conclusions

In the present research, we developed a hybrid approach to design and synthesize hit-finding libraries. The essence of the approach was to use diverse, yet bio-like compounds for efficient hit-finding. Two libraries based on dipeptide scaffolds – linear and cyclized – were designed. For both scaffolds, virtual libraries comprising of more than 30k compounds each were generated, and their diversity, physicochemical properties and drug-likeness were analysed computationally. From two libraries, total 51 compounds spanning a range of physicochemical and pharmacophoric properties were synthesized. Bioactivity studies of the synthesized compounds were carried out on different platforms, which showed that 4 compounds from *N*-substituted dipeptide esters series and 1 compound from piperazine-2,5-diones series caused loss of viability of cancer cell lines, while 1 compound from piperazine-2,5-diones series inhibited brain-specific kinase 1 (BRSK1).

With this, we have successfully used "biofocussed chemoprospecting" as an efficient hit-finding approach. While conventional approaches use libraries comprising thousands of compounds synthesized at very high cost, we have been able to demonstrate that a biofocussed diversity approach for hit-finding libraries can provide hits even from small library at cheap cost with hit-rates comparable to target-based approaches.

Acknowledgement

The authors are grateful for funding from University of Tromsø and from the MABIT program of NorInnova, Tromsø. The authors declare that they have no conflict of interest.

References

1. Hughes J., Rees S., Kalindjian S., Philpott K. (2011). Principles of early drug discovery. Br. J. Pharmacol.; 162:1239–1249.
2. Mullin R. (2014). Cost to develop new pharmaceutical drug now exceeds $2.5B. Scientific American. (cited 19 Oct 2016): http://www.scientificamerican.com/article/cost-to-develop-new-pharmaceutical-drug-now-exceeds-2-5b/.
3. DiMasi J.A., Grabowski H.G., Hansen R.W. (2015). The Cost of Drug Development. N. Engl. J. Med.; 372:1972–1972.
4. Nicolaou K.C. (2014). Advancing the Drug Discovery and Development Process. Angew. Chem. Int. Ed.; 53:9128–9140.
5. Pammolli F., Magazzini L., Riccaboni M. (2011). The productivity crisis in pharmaceutical R&D. Nat. Rev. Drug Discov.; 10:428–438.
6. Hartung T. (2013). Food for Thought Look Back in Anger – What Clinical Studies Tell Us About Preclinical Work. ALTEX; 30:275–291.
7. Begley C.G., Ellis L.M. (2012). Drug development: Raise standards for preclinical cancer research. Nature; 483:531–533.
8. Lounnas V., Ritschel T., Kelder J., McGuire R., Bywater R.P., Foloppe N. (2013). Current Progress In Structure-Based Rational Drug Design Marks A New Mindset In Drug Discovery. Comput. Struct. Biotechnol. J.; 5:1–14.
9. Verma S., Prabhakar Y.S. (2015). Target based drug design - a reality in virtual sphere. Curr. Med. Chem.; 22:1603–1630.
10. Anderson A.C. (2003). The Process of Structure-Based Drug Design. Chem. Biol.; 10:787–797.
11. Wang T., Wu M.-B., Chen Z.-J., Chen H., Lin J.-P., Yang L.-R. (2015). Fragment-based drug discovery and molecular docking in drug design. Curr. Pharm. Biotechnol.; 16:11–25.
12. Scott D.E., Coyne A.G., Hudson S.A., Abell C. (2012). Fragment-Based Approaches in Drug Discovery and Chemical Biology. Biochemistry (Mosc.); 51:4990–5003.
13. Joseph-McCarthy D., Campbell A.J., Kern G., Moustakas D. (2014). Fragment-Based Lead Discovery and Design. J. Chem. Inf. Model.; 54:693–704.
14. Zhang K.Y.J., Milburn M.V., Artis D.R. (2007). Scaffold-Based Drug Discovery. In Structure-Based Drug Discovery, (Springer Netherlands), pp. 129–153.
15. Kirkpatrick P. (2005). Small is beautiful. Nat. Rev. Drug Discov.; 4:190–190.
16. Welsch M.E., Snyder S.A., Stockwell B.R. (2010). Privileged Scaffolds for Library Design and Drug Discovery. Curr. Opin. Chem. Biol.; 14:347–361.
17. Card G.L., Blasdel L., England B.P., Zhang C., Suzuki Y., Gillette S., Fong D., Ibrahim P.N., Artis D.R., Bollag G., Milburn M.V., Kim S.-H., Schlessinger J., Zhang K.Y.J. (2005). A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold-based drug design. Nat. Biotechnol.; 23:201–207.
18. Saslis-Lagoudakis C.H., Savolainen V., Williamson E.M., Forest F., Wagstaff S.J., Baral S.R., Watson M.F., Pendry C.A., Hawkins J.A. (2012). Phylogenies reveal predictive power of traditional medicine in bioprospecting. Proc. Natl. Acad. Sci.; 109:15835–15840.
19. L. A., Gericke N. (2011). Bioprospecting: Creating a Value for Biodiversity. In Research in Biodiversity - Models and Applications, I. Pavlinov, ed. (InTech), p.
20. Bailey F., Dundas I. (2001). Bioprospecting: Discoveries changing the future. Aust. TPOTCO Ed Canberra House Represent. Standing Comm. Prim. Ind. Reg. Serv.;
21. Firn R.D. Bioprospecting – why is it so unrewarding? Biodivers. Conserv.; 12:207–216.
22. Galloway W.R.J.D., Isidro-Llobet A., Spring D.R. (2010). Diversity-oriented synthesis as a tool for the discovery of novel biologically active small molecules. Nat. Commun.; 1:80.
23. Wassermann A.M., Camargo L.M., Auld D.S. (2014). Composition and applications of focus libraries to phenotypic assays. Front. Pharmacol.; 5.
24. Petrone P.M., Wassermann A.M., Lounkine E., Kutchukian P., Simms B., Jenkins J., Selzer P., Glick M. (2013). Biodiversity of small molecules – a new perspective in screening set selection. Drug Discov. Today; 18:674–680.
25. Hert J., Irwin J.J., Laggner C., Keiser M.J., Shoichet B.K. (2009). Quantifying Biogenic Bias in Screening Libraries. Nat. Chem. Biol.; 5:479–483.
26. Overington J.P., Al-Lazikani B., Hopkins A.L. (2006). How many drug targets are there? Nat. Rev. Drug Discov.; 5:993–996.

27. Frearson J.A., Collie I.T. (2009). HTS and hit finding in academia – from chemical genomics to drug discovery. Drug Discov. Today; 14:1150–1158.
28. Tan D.S. (2005). Diversity-oriented synthesis: exploring the intersections between chemistry and biology. Nat. Chem. Biol.; 1:74–84.
29. Brüstle M., Beck B., Schindler T., King W., Mitchell T., Clark T. (2002). Descriptors, Physical Properties, and Drug-Likeness. J. Med. Chem.; 45:3345–3355.
30. Walters W.P., Murcko M.A. (2002). Prediction of "drug-likeness." Adv. Drug Deliv. Rev.; 54:255–271.
31. Ursu O., Rayan A., Goldblum A., Oprea T.I. (2011). Understanding drug-likeness. Wiley Interdiscip. Rev. Comput. Mol. Sci.; 1:760–781.
32. Lipinski C.A., Lombardo F., Dominy B.W., Feeney P.J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev.; 23:3–25.
33. Congreve M., Carr R., Murray C., Jhoti H. (2003). A "Rule of Three" for fragment-based lead discovery? Drug Discov. Today; 8:876–877.
34. Alberto López-Cobeñas P.C. (2005). Microwave-Assisted Synthesis of 2,5-Piperazinediones under Solvent-Free Conditions. Synthesis; 2005:3412.
35. Pérez-Picaso L., Escalante J., Olivo H.F., Rios M.Y. (2009). Efficient Microwave Assisted Syntheses of 2,5-Diketopiperazines in Aqueous Media. Molecules; 14:2836–2849.
36. Pachaly P., Pelzer H.-J. (1983). 6-Alkyl-2,5-bisethoxy-3-ethoxycarbonyl-3,6-dihydropyrazine aus Piperazin-2,5-dionen 6-Alkyl-2,5-bisethoxy-3-ethoxycarbonyl-3,6-dihydropyrazines from Piperazine-2,5-diones. Arch. Pharm. (Weinheim); 316:653–655.
37. Borthwick A.D. (2012). 2,5-Diketopiperazines: Synthesis, Reactions, Medicinal Chemistry, and Bioactive Natural Products. Chem. Rev.; 112:3641–3716.
38. Martins M.B., Carvalho I. (2007). Diketopiperazines: biological activity and synthesis. Tetrahedron; 63:9923–9932.
39. Daugan A., Grondin P., Ruault C., Le Monnier de Gouville A.-C., Coste H., Linget J.M., Kirilovsky J., Hyafil F., Labaudinière R. (2003). The Discovery of Tadalafil: A Novel and Highly Selective PDE5 Inhibitor. 2: 2,3,6,7,12,12a-hexahydropyrazino[1',2':1,6]pyrido-[3,4-b]indole-1,4-dione Analogues. J. Med. Chem.; 46:4533–4542.
40. Fintan K., Richard T.L., Angus M.M., Kevin J.M. (1994). Piperazine derivatives. GB2271774 (A).
41. Ankersen M., Arndt K., Conde-Frieboes K.W., Krist B., Lustenberger P., Mueller S., Rudolf K., Schindler M., Sensfuss U., Stenkamp D., Thoegersen H., Wieland H., Wulff B.S. (2004). 2,5-diketopiperazines for the treatment of obesity. WO2004048345 A3
42. Borthwick A.D., Liddle J., Davies D.E., Exall A.M., Hamlett C., Hickey D.M., Mason A.M., Smith I.E.D., Nerozzi F., Peace S., Pollard D., Sollis S.L., Allen M.J., Woollard P.M., Pullen M.A., Westfall T.D., Stanislaus D.J. (2012). Pyridyl-2,5-Diketopiperazines as Potent, Selective, and Orally Bioavailable Oxytocin Antagonists: Synthesis, Pharmacokinetics, and In Vivo Potency. J. Med. Chem.; 55:783–796.
43. Gomez-Monterrey I., Campiglia P., Carotenuto A., Califano D., Pisano C., Vesci L., Lama T., Bertamino A., Sala M., di Bosco A.M., Grieco P., Novellino E. (2007). Design, Synthesis, and Cytotoxic Evaluation of a New Series of 3-Substituted Spiro[(dihydropyrazine-2,5-dione)-6,3'-(2',3'-dihydrothieno[2,3-b]naphtho-4',9'-dione)] Derivatives. J. Med. Chem.; 50:1787–1798.
44. Gomez-Monterrey I., Campiglia P., Carotenuto A., Stiuso P., Bertamino A., Sala M., Aquino C., Grieco P., Morello S., Pinto A., Ianelli P., Novellino E. (2008). Spiro[(dihydropyrazin-2,5-dione)-6,3'-(2',3'-dihydrothieno[2,3-b]naphtho-4',9'-dione)]-Based Cytotoxic Agents: Structure–Activity Relationship Studies on the Substituent at N4-Position of the Diketopiperazine Domain. J. Med. Chem.; 51:2924–2932.
45. Szardenings A.K., Harris D., Lam S., Shi L., Tien D., Wang Y., Patel D.V., Navre M., Campbell D.A. (1998). Rational Design and Combinatorial Evaluation of Enzyme Inhibitor Scaffolds: Identification of Novel Inhibitors of Matrix Metalloproteinases. J. Med. Chem.; 41:2194–2200.

46. Szardenings A.K., Antonenko V., Campbell D.A., DeFrancisco N., Ida S., Shi L., Sharkov N., Tien D., Wang Y., Navre M. (1999). Identification of highly selective inhibitors of collagenase-1 from combinatorial libraries of diketopiperazines. J. Med. Chem.; 42:1348–1357.
47. Campbell D., Look G.C., Szardenings A.K., Patel D.V. (1999). Collagenase-1 and stromelysin-1 inhibitors, pharmaceutical compositions comprising same and methods of their use.
48. Qiao Y., Gao J., Qiu Y., Wu L., Guo F., Lo K.K.-W., Li D. (2011). Design, synthesis, and characterization of piperazinedione-based dual protein inhibitors for both farnesyltransferase and geranylgeranyltransferase-I. Eur. J. Med. Chem.; 46:2264–2273.
49. Verardo G., Geatti P., Pol E., Giumanini A.G. (2002). Sodium borohydride: A versatile reagent in the reductive N-monoalkylation of α-amino acids and α-amino methyl esters. Can. J. Chem.; 80:779–788.
50. Services | Premier Screen | International Centre for Kinase Profiling. (Cited 19 Oct 2016): http://www.kinase-screen.mrc.ac.uk/services/premier-screen
51. Explore KINOMEscan - DiscoverX. (Cited 19 Oct 2016): https://www.discoverx.com/services/drug-discovery-development-services/explore-our-panels/explore-kinomescan
52. Montalbetti C.A.G.N., Falque V. (2005). Amide bond formation and peptide coupling. Tetrahedron; 61:10827–10852.
53. Li J., Sha Y. (2008). A convenient synthesis of amino acid methyl esters. Mol. Basel Switz.; 13:1111–1119.
54. McGhie S. (2002). Preparation of Methyl Ester Derivatives of Amino Acids Bearing Hydrolysable N-Protection. Synth. Commun.; 32:1275–1278.
55. Mao F., Ni W., Xu X., Wang H., Wang J., Ji M., Li J. (2016). Chemical Structure-Related Drug-Like Criteria of Global Approved Drugs. Mol. Basel Switz.; 21:75.
56. Chemical Property Overview. (cited 19 Oct 2016): http://www.openmolecules.org/properties/properties.html.
57. Lain S., Hollick J.J., Campbell J., Staples O.D., Higgins M., Aoubala M., McCarthy A., Appleyard V., Murray K.E., Baker L., Thompson A., Mathers J., Holland S.J., Stark M.J.R., Pass G., Woods J., Lane D.P., Westwood N.J. (2008). Discovery, In Vivo Activity, and Mechanism of Action of a Small-Molecule p53 Activator. Cancer Cell; 13:454–463.
58. Futamura Y., Kawatani M., Kazami S., Tanaka K., Muroi M., Shimizu T., Tomita K., Watanabe N., Osada H. (2012). Morphobase, an encyclopedic cell morphology database, and its use for drug target identification. Chem. Biol.; 19:1620–1630.
59. Toxicity Assessment. (cited 19 Oct 2016): http://www.openmolecules.org/propertyexplorer/toxicity-assessment.html
60. Ekins S., Berbaum J., Harrison R.K., Zecher M., Yuan J., Ishchenko A.V., Berezin K., Chubukov V., Lawson J.D., Hupcey M.A. (2005). Applying Computational and InVitro Approaches to Lead Selection. In Pharmaceutical Profiling in Drug Discovery for Lead Selection, p. 361.
61. Valler M.J., Green D. (2000). Diversity screening versus focussed screening in drug discovery. Drug Discov. Today; 5:286–293.
62. Orry A.J.W., Abagyan R.A., Cavasotto C.N. (2006). Structure-based development of target-specific compound libraries. Drug Discov. Today; 11:261–266.
63. Hoelder S., Clarke P.A., Workman P. (2012). Discovery of small molecule cancer drugs: Successes, challenges and opportunities. Mol. Oncol.; 6:155–176.

Figures and Scheme

Figure 1: Different approaches in drug discovery research with their advantages and disadvantages

Figure 2: Efficiency parameters for library design

Figure 3: Biofocused chemoprospecting – The idea of using biomolecules as a starting point led to *N*-substituted dipeptide esters and piperazine-2,5-diones libraries

Figure 4: Scaffolds for *N*-substituted dipeptide esters (left) and piperazine-2,5-diones (right)

Figure 5: Difference between *N*-substituted dipeptides from previous reports and our scheme

Scheme 1: Synthesis scheme for *N*-substituted dipeptide esters and piperazine-2,5-dione libraries

Figure 6: Molecular weight distribution for *N*-substituted dipeptide esters (left) and piperazine-2,5-diones (right)

Figure 7: AlogP distribution for *N*-substituted dipeptide esters (left) and piperazine-2,5-diones (right)

Figure 8: Hydrogen bond acceptors distribution for *N*-substituted dipeptide esters (left) and piperazine-2,5-diones (right)

Figure 9: Hydrogen bond donors distribution for *N*-substituted dipeptide esters (left) and piperazine-2,5-diones (right)

Figure 10: Number of rotatable bonds distribution for *N*-substituted dipeptide esters (left) and piperazine-2,5-diones (right)

Figure 11: Polar surface area distribution for *N*-substituted dipeptide esters (left) and piperazine-2,5-diones (right)

Figure 12: Distribution of AlogP relative to molecular weight for the virtual libraries of *N*-substuituted dipeptide esters (left) and piperazine-2,5-diones (right). Red dots represent the synthesized compounds.

Figure 13: Distribution of drug-likeness, molecular weight and clogP for the virtual libraries of *N*-substuituted dipeptide esters (left) and piperazine-2,5-diones (right). Red dots represent the synthesized compounds.

Figure 14: Distribution of cost, molecular weight and clogP for the virtual library of *N*-substuituted dipeptide esters. The red dots signify the synthesized compounds.

Figure 15: Bioactivity studies on both libraries

Figure 16: Change in morphology of A2058 melanoma cells caused in presence of 5 μM concentration of an *N*-substituted dipeptide ester compound (right) in 4 hours compared to the negative control (left).

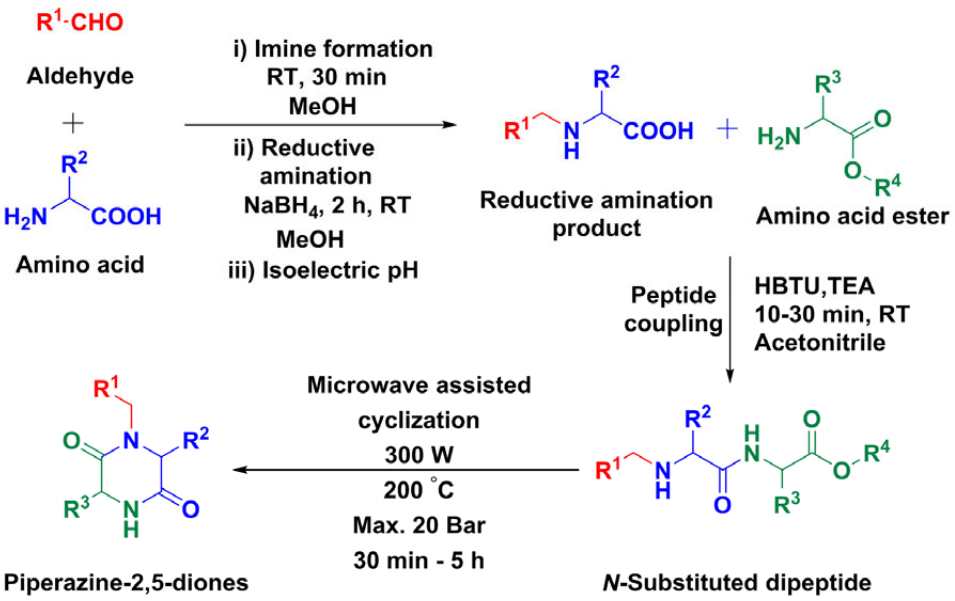


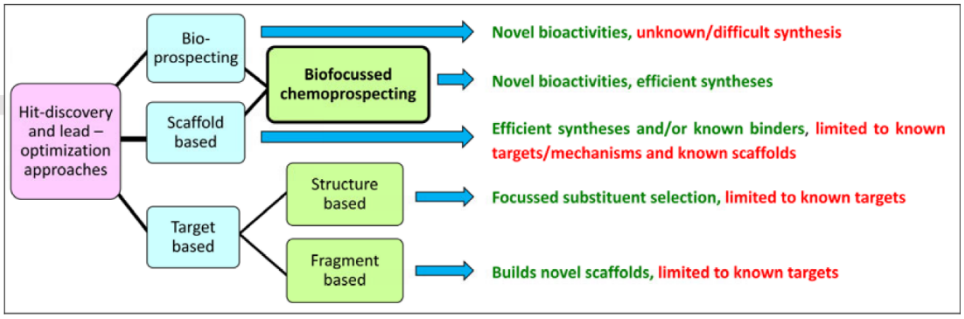
Table 1: Substitution variations for library synthesis

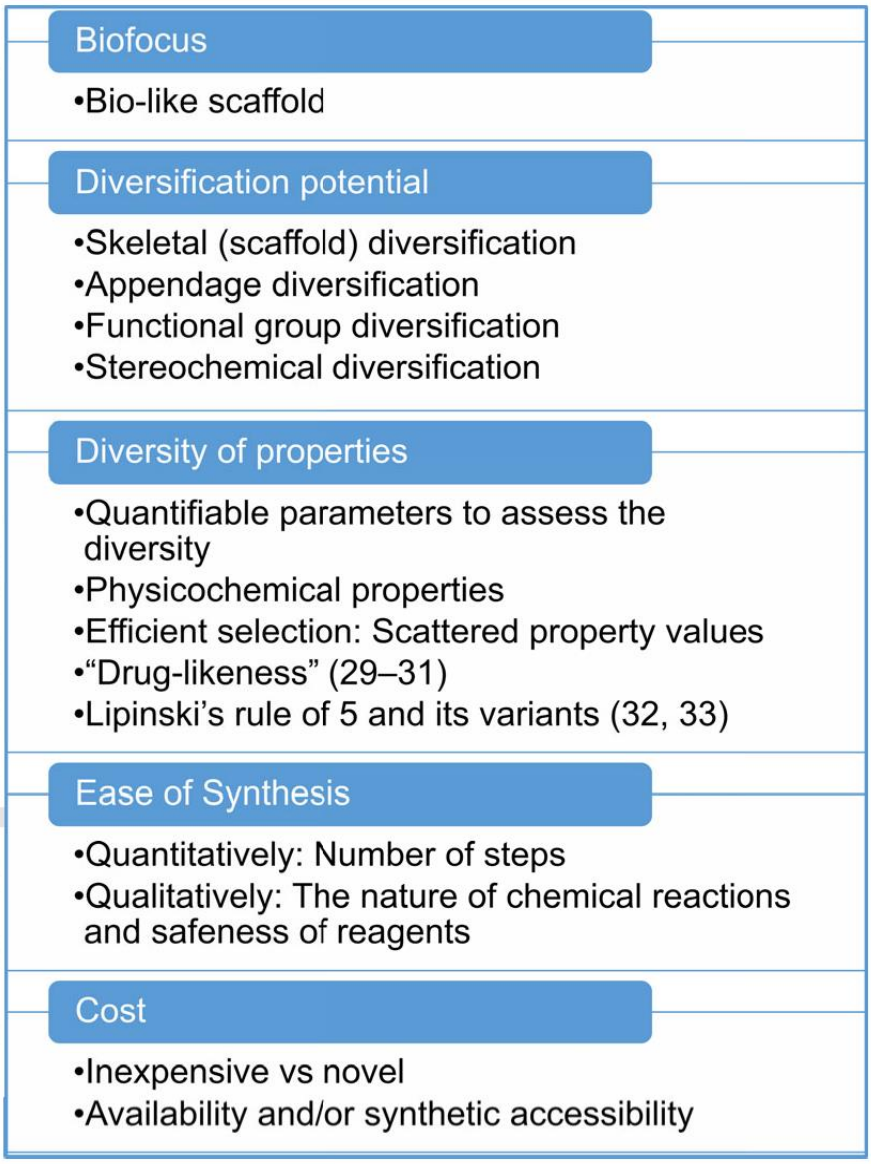
No.	R^1	R^2	R^3
1	1	A	a
2	1	B	a
3	1	B	c
4	1	B	e
5	1	C	a
6	1	C	b
7	1	C	d
8	2	A	a
9	2	A	c
10	2	A	d
11	2	A	e
12	2	C	a
13	2	C	d
14	3	C	b
15	3	C	d

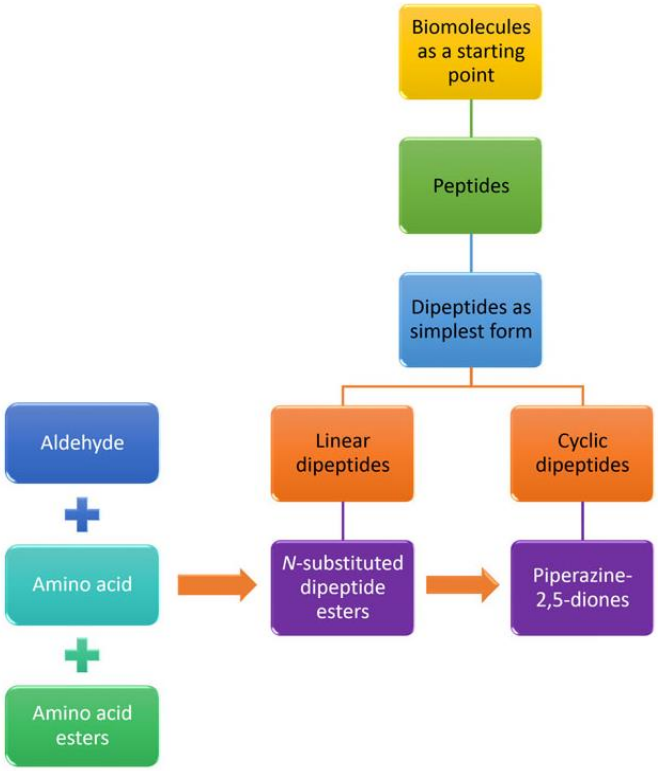
No.	R^1	R^2	R^3
16	3	C	e
17	4	C	a
18	4	C	b
19	4	C	c
20	4	C	e
21	5	C	a
22	5	C	b
23	5	C	c
24	5	C	d
25	5	C	e
26	5	C	f
27	5	C	g
28	6	B	d
29	6	D	d

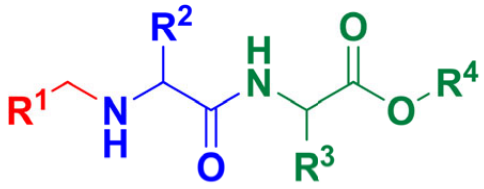
Small aliphatic	Aromatic non-polar	Aromatic polar
No HBD or HBA	Only HBA	HBD + HBA

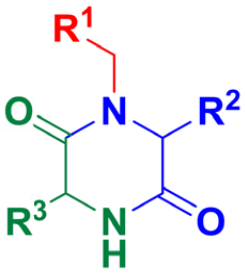
HBD = Hydrogen bond donor; HBA = Hydrogen bond acceptor

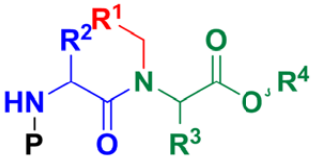


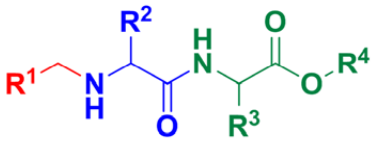












***N*-Substituted dipeptide esters from previous examples**

***N*-Substituted dipeptide esters from our protocol**