Paper 2
Biofocussed chemoprospecting: An efficient approach for drug discovery

Balmukund S. Thakkar\textsuperscript{a}, Marte Albrigtsen\textsuperscript{b}, John-Sigurd M. Svendsen\textsuperscript{a}, Jeanette H. Andersen\textsuperscript{b}, Richard A. Engh\textsuperscript{a*}

\textsuperscript{a}Department of Chemistry, UiT The Arctic University of Norway, N-9037 Tromsø, Norway
\textsuperscript{b}Marbio, UiT The Arctic University of Norway, Breivika, N-9037, Tromsø, Norway.

\textsuperscript{*}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

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/oby.12934
This article is protected by copyright. All rights reserved.
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 “biococussed 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).

This article is protected by copyright. All rights reserved.
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 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 alkylaryl 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: Nsubstituted 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 metallopeptidase inhibitors. Similarly, piperazine-2,5-dione derivatives as dual inhibitors of farnesyltransferase and geranylgeranyltransferase-I 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 F254 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 F254 TLC aluminum sheets. For TLC as well as column chromatography, 0-100 % of ethyl acetate in pentane or 0-15 % of methanol in dichloromethane (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 1H NMR, 13C NMR, MS and IR were used to confirm the products. NMR spectra were recorded at 400 MHz (1H NMR) and at 100 MHz (13C 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 pI 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 (iK) and DiscoverX KINOMEscan (S1). 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 9000 HepG2 cells per well were seeded in black 96 well plates with clear bottoms (Im 3603, Corning, NY, USA) and incubated overnight. The cells were labelled with 10 µM C11-BODIPY (H#3861, Invitrogen, Eugene, OR) for 30 min and incubated for 1 h with various concentrations of the test compounds. 50 µM Cumene hydroperoxide (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

This article is protected by copyright. All rights reserved.
(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 °C with 5% CO₂. Percent inhibition was calculated relative to the positive control (cumODH 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 °C in 5% CO₂, 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 H₂O was used as a negative control while sterile MilliQ H₂O and bacteria suspension was used as a positive control. Bacteria were transferred from a blood plate to growth medium (MH-bouillon #275730 Difco Becton Dickinson) for E. coli, P. aeruginosa and S. aureus and Tl-h-bouillon (#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 °C overnight before growth was assessed visually and photometrally 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; #1.05439 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.

This article is protected by copyright. All rights reserved.
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' comes from an aldehyde; R comes from an amino acid, while R' 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
erstereoisomers 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 piperazino-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 3300 distinct substructure fragments with associated drug likeness scores [56].

This article is protected by copyright. All rights reserved.
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, logP 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 logP 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 R1 is C and R2 is d, 4 variation for R1 as 1, 2, 3 and 5 respectively are present. When R1 is 1 and R2 is a, 3 variations for R1 as A, B and C are present. Similarly, when R1 is 5 and R2 is C, 7 variations for R1 as g to a are present. Further, activity trends for such variations can also be evaluated with respect to more than one pairs. For example, variations of R2 can be evaluated for different R1-R2 pairs such as 1-C, 3-C, 4-C, 5-C or 2-A, or variations of R2 can be evaluated for different R1-R2 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.

This article is protected by copyright. All rights reserved.
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 10). 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 synthesised. 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.

This article is protected by copyright. All rights reserved.
References


This article is protected by copyright. All rights reserved.

This article is protected by copyright. All rights reserved.

Figures and Scheme

This article is protected by copyright. All rights reserved.
Figure 1: Different approaches in drug discovery research with their advantages and disadvantages.

Figure 2: Efficiency parameters for library design.

Figure 3: Biofocussed chemo-prospecting – 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-substituted 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-substituted 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-substituted dipeptide esters. The red dots signify the synthesized compounds.

Figure 15: Bioactivity studies on both libraries.

Figure 16: Change in morphology of A2038 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).

This article is protected by copyright. All rights reserved.
$R^1\text{CHO}$

Aldehyde

$+$

$H_2N \overset{\text{COOH}}{\text{R}}$

Amino acid

i) Imine formation

RT, 30 min

MeOH

ii) Reductive amination

NaBH$_4$, 2 h, RT

MeOH

iii) Isoelectric pH

Reductive amination product

Amino acid ester

Peptide coupling

HBTU, TEA

10-30 min, RT

Acetonitrile

Microwave assisted cyclization

300 W

200 $^\circ$C

Max. 20 Bar

30 min - 5 h

Piperazine-2,5-diones

$N$-Substituted dipeptide

This article is protected by copyright. All rights reserved.
Table 1: Substitution variations for library synthesis

<table>
<thead>
<tr>
<th>No.</th>
<th>R^1</th>
<th>R^2</th>
<th>R^3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>c</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>e</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>A</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>A</td>
<td>c</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>A</td>
<td>e</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>C</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>C</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>C</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>C</td>
<td>d</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No.</th>
<th>R^1</th>
<th>R^2</th>
<th>R^3</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>C</td>
<td>c</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>C</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>C</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>C</td>
<td>c</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>C</td>
<td>e</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>C</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>C</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>C</td>
<td>c</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>C</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>C</td>
<td>e</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>C</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>C</td>
<td>g</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>B</td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>D</td>
<td>d</td>
<td></td>
</tr>
</tbody>
</table>

Small aliphatic | Aromatic non-polar | Aromatic polar
No HBD or HBA | Only HBA | HBD = HBA

HBD = Hydrogen bond donor; HBA = Hydrogen bond acceptor

---

This article is protected by copyright. All rights reserved.
<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biofocus</td>
<td>• Bio-like scaffold</td>
</tr>
<tr>
<td>Diversification potential</td>
<td>• Skeletal (scaffold) diversification</td>
</tr>
<tr>
<td></td>
<td>• Appendage diversification</td>
</tr>
<tr>
<td></td>
<td>• Functional group diversification</td>
</tr>
<tr>
<td></td>
<td>• Stereochemical diversification</td>
</tr>
<tr>
<td>Diversity of properties</td>
<td>• Quantifiable parameters to assess the diversity</td>
</tr>
<tr>
<td></td>
<td>• Physicochemical properties</td>
</tr>
<tr>
<td></td>
<td>• Efficient selection: Scattered property values</td>
</tr>
<tr>
<td></td>
<td>• “Drug-likeness” (29–31)</td>
</tr>
<tr>
<td></td>
<td>• Lipinski’s rule of 5 and its variants (32, 33)</td>
</tr>
<tr>
<td>Ease of Synthesis</td>
<td>• Quantitatively: Number of steps</td>
</tr>
<tr>
<td></td>
<td>• Qualitatively: The nature of chemical reactions and safeness of reagents</td>
</tr>
<tr>
<td>Cost</td>
<td>• Inexpensive vs novel</td>
</tr>
<tr>
<td></td>
<td>• Availability and/or synthetic accessibility</td>
</tr>
</tbody>
</table>
N-Substituted dipeptide esters from previous examples

N-Substituted dipeptide esters from our protocol

This article is protected by copyright. All rights reserved.
This article is protected by copyright. All rights reserved.
This article is protected by copyright. All rights reserved.
This article is protected by copyright. All rights reserved.