UiT
THE ARCTIC
UNIVERSITY
OF NORWAY

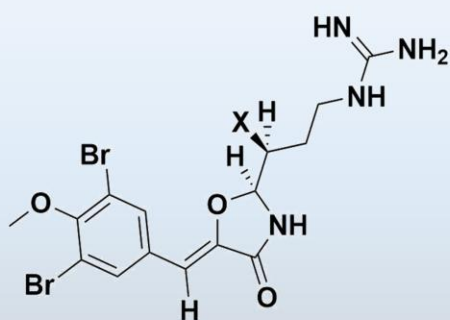Faculty of Science and Technology
Department of Chemistry

A Biofocussed Chemoprospecting Approach to Drug Discovery

Design, Synthesis and Bioactivity Screening of Diverse Biofocussed Chemical Libraries

—

Balmukund Sureshkumar Thakkar
A dissertation for the degree of Philosophiae Doctor – February 2017

A dissertation for the degree of Philosophiae Doctor

A Biofocussed Chemoprospecting Approach to Drug Discovery:

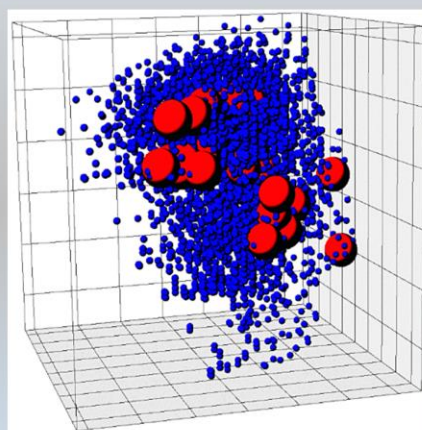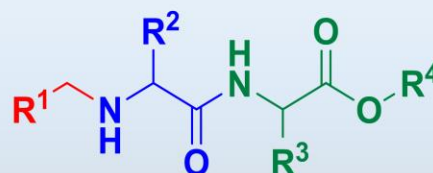# Design, Synthesis and Bioactivity Screening of Diverse Biofocussed Chemical Libraries

Balmukund Sureshkumar Thakkar

Tromsø 2017

Department of Chemistry

Faculty of Science and Technology

University of Tromsø

Norway

योगः कर्मसु कौशलम् ।

Yogah Karmasu Kaushalam.

"To acquire skills in doing the right things is the essence of Yoga"

Table of Contents

Acknowledgements

Abbreviations and acronyms

ACN – Acetonitrile

ATP – Adenosine triphosphate

BRSK1 – Brain specific kinase 1

^{cal}E – Calculated energy value for a structure obtained based on additive principle

DCC - *N,N′*-Dicyclohexylcarbodiimide

DCM – Dichloromethane

DKP – Piperazine-2,5-dione

DMF – Dimethylformamide

DMSO – Dimethylsulfoxide

EBG – Energy barrier geometry – obtained with relaxed coordinate scan at a fixed ω value close to the energy barrier

EDC - 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide

FDA – Food and drug administration, USA

GGMe – Glycylglycine methyl ester

GP – Gas phase

GPE – Gas phase energy

HATU - 1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate

HBA – Hydrogen bond acceptor

HBD – Hydrogen bond donor

HBTU - *N,N,N′,N′*-Tetramethyl-*O*-(1*H*-benzotriazol-1-yl)uronium hexafluorophosphate

HOMO – Highest occupied molecular orbital

HPLC – High performance liquid chromatography

IRC – Intrinsic reaction coordinates

Me – Methyl

MS – Mass spectrometry

NMA – *N*-methylacetamide

NMR – Nuclear magnetic resonance

NSD – *N*-substituted dipeptide ester

OMe – Methyl ester

^{opt}E – Energy of optimized minimum energy geometry

PDB – Protein data bank

PKA – Protein kinase A

POP – Prolyl oligopeptidase

QSAR – Quantitative structure activity relationship

RCS – Relaxed coordinate scan

ROCK – Rho-associated protein kinase

RT – Room temperature

SAR – Structure activity relationship

SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SILAC – Stable Isotope Labeling with Amino acid in Cell culture

SPE – Solution phase energy

^{st}E – Steric effect estimated with additive principle

TAB – Tartaric acid bisamide

TEA – Triethylamine

THF – Tetrahydrofuran

TS – Transition state

TS_{anti} 120 – Transition state geometry of *anti* type, with ω values near 120º

TSG – Transition state geometry – obtained with transition state search method, with only one imaginary frequency

TS_{syn} 60 – Transition state geometry of *syn* type with ω values near 60º

UPLC – Ultra performance liquid chromatography

WP – Water phase

Introduction

The introduction will begin with a brief discussion of the current status of the drug discovery process, followed by a discussion on why newer approaches are necessary, and how a "biofocussed chemoprospecting" approach can lead to more efficient hit-generation in drug discovery. This will be followed by a detailed discussion on specific efficiency parameters for library design and effective compound selection as the core of this approach. The principle and approach of biofocussed chemoprospecting thereby emerges; it developed further by describing specific examples of compound libraries. Finally, the detailed goals of the thesis research will be shown to follow from these analyses. Thus, this introduction should guide the reader from current status of drug discovery to the specific goals of this research and prepare for the results and discussion sections.

1. Biofocussed chemoprospecting approach

This chapter serves as an introduction to the current status of drug discovery processes and timelines, the need to find new approaches of basic research, and how biofocussed chemoprospecting can provide a way forward for efficient drug discovery. This includes discussion of a typical drug discovery process, followed by discussion of different types of biochemical screening, their role in drug discovery and the importance of linking the screening approaches with synthetic chemistry accessibility. This leads to "chemoprospecting", the approach introduced and followed here. These discussions are expected to bridge the reader to the more specific introduction on chemical libraries for chemoprospecting, to be discussed in chapter 2.

1.1. Need for faster and cheaper drug discovery approaches

New drug discovery and development is typically a long process involving enormous resources in terms of time, money, labour and intellect. If we consider the drug discovery process as a sequence from the original idea to the market launch, it can take from 8 to 12 years of time, and more than 1 billion USD, to develop a single new chemical entity (NCE) into a marketable drug.[1,2] In fact, according to an estimate[3] from the Tufts Center for the Study of Drug Development (CSDD), the average drug discovery investment has risen to over 2.6 billion USD per approved drug. Moreover, in recent years, the rise in the cost of drug discovery has been accompanied by the high attrition rates and an overall decline in pharma R & D productivity, raising the risks of investment.[4–6] Further, most pharmaceutical research is carried out with target based approach, which (despite being a low-risk approach) necessitates significantly improved therapeutic activity from existing molecules – often requiring longer and larger clinical trials, in turn increasing the cost and time before market-launch, resulting high drug prices of new drugs.

A drug discovery research project must pass through several stages before a drug can be launched to market.[7] The period before the preclinical testing is considered basic research in which only a small number of compounds out of thousands of molecules are selected for further research. This basic research phase can take up to 5 years. After this phase, the possibilities to minimize cost and time are limited, as each stage must be carried out in compliance with extensive regulatory guidelines. This shows the need to adopt newer approaches in basic research phase (especially for hit-finding and lead-optimization) that can make the drug discovery process faster, cheaper and more efficient. In addition to streamlining the process, new approaches also promise improvements in the quality of the results, as it been observed[8,9] that improved and more focused preclinical research may

lessen the possibility of drug discovery failure. The following chapters will throw more light on this and how "Chemoprospecting" approach offers a good alternative to current approaches.

1.2.   Bioactivity screening

The primary goal of any drug discovery process is to elicit a specific biological response that can translate into corresponding therapeutic activity. This can be analyzed by measuring the effect of compounds in question on the biological responses in bioactivity screening. Thus, bioactivity screening has a central role in any drug discovery model. The screenings may be phenotypic screening or target specific screening.

1.2.1. Phenotypic screening

Phenotypic screening is useful to identify the effect of compounds on specific cellular activity. Hence, such screening is useful when the objective of the research is to identify whether compounds exert any therapy relevant activity at all, independent of whether the drug-target at the core of the activity mechanism is known. High throughput phenotypic screening[10–13] has generally involved *in vitro* (biochemical or cellular) assays since the 1980s, in contrast to earlier reliance on *in vivo* (i.e. animal models) phenotypic assays. They identify effects of compounds on specific cellular activity, depending on the assay readout. Usually, the assays are performed on well characterized cell lines and a specific read-out parameter, such as cell-growth or particular protein production. For more complex analysis with parallel measurement of more than one parameters, high content screening[14,15] is also used.

1.2.2. Target specific screening

Target specific screening is the logical approach when the objective of the research is to identify the effect of compounds on a specific cellular or extracellular target, usually an enzyme or receptor. Such assays are always *in vitro* assays, and are carried out on isolated targets, generally as activity or binding assays.[16] As indicated by the names, activity assays measure the activity of a target as a function of compound concentration, while binding assays measure the binding of a compound to the often immobilized target, independent of its activity effects.

Table 1 provides an overview of phenotypic and target specific screening approaches.

Table 1. Comparison of phenotypic and target specific screenings

Criteria	Phenotypic screening	Target specific screening
In vitro / *In vivo*	Can be both, but mostly *in vitro* in initial stage of drug discovery	*In vitro*
Subject	Cells or Cell lines	Specific target such as enzyme / receptor
Read-out	Growth / growth inhibition / cell count / protein expression / effect on cellular organnels etc.	Enzyme activity or binding
Techniques	Cellular imaging, fluorescence-luminescence-absorbance, flow-cytometry, etc.	Flurorescence-absorbance-luminescence, isothermal calorimetry, SPR (surface-plasmon resonance),
Identification and validation of target	Yes, target identification and validation is required, often it proves to be a bottleneck	No, assays are done on the validated target itself
Correlation for cellular level activity	Assays are done on cells, so results are indicative of the activity at cellular level	Critically depends factors such as on absorption, pH and stability inside the cell and cellular organelles
Correlation with in vivo activity	Generally good correlation, except where distribution and metabolism affect severely.	Very unpredictable

1.3. Hit-finding approaches and library design

As described above, bioactivity screening is the core of any drug discovery project. The overall goal is to achieve valuable therapeutic activity by eliciting a suitable phenotypical response due to target interactions. For medicinal chemists, the means to achieve the specific biological response is usually to select—somehow—a promising set of small chemical molecules, from the vast "chemical space" of possible molecules, and optimize them. The typical hit-finding and lead optimization process involves screening a large number of molecules with an assay simple enough to enable high throughput testing, choosing the most promising hits, and optimizing them based on hypotheses of the best properties for clinical trials. The fundamental requirement for this process is the supply of the molecules for screening, which may be either natural or synthetic. Based on the source and choice of molecules, and corresponding bioactivity screening approaches, there are several conventional methods for hit-discovery and lead optimization: (1) target-based approaches (structure- or fragment based), (2) a scaffold-based approach, or (3) bioprospecting (Figure 1). The bioprospecting approach relies on natural product extracts

and purified biomolecules to provide the molecules for screening, while all other approaches rely largely on synthetic chemistry.
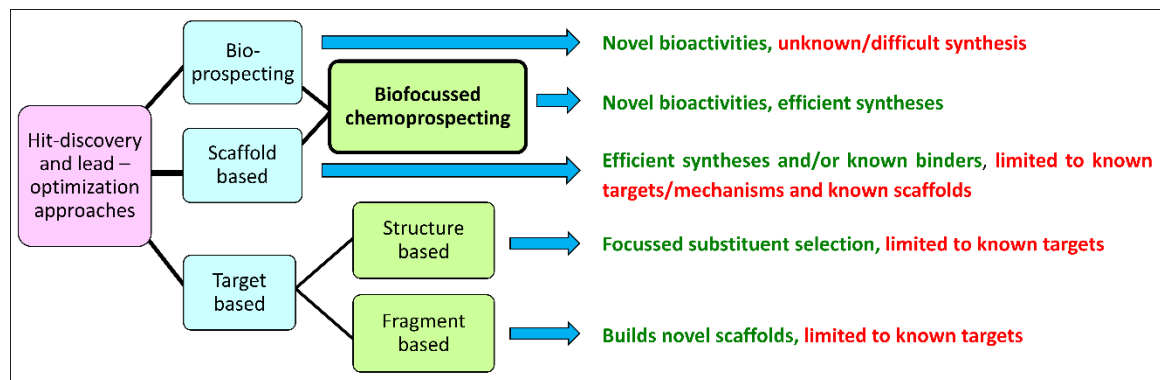
Figure 1. Different approaches in drug discovery research

1.3.1. Target based approaches

Target based approaches use validated and "druggable" targets. Druggable targets are ones that have chemical characteristics that are compatible with high affinity binding at sites where binding should cause the desirable therapeutic effect. Usually, high throughput screening is used to identify "hit" molecules, which is followed by computational studies and combinatorial/parallel synthesis approaches to modify the molecules for further testing. Target based approaches are relatively low-risk, as only the hits that act on known and validated drug targets are considered for further development. Hence, target based approaches have been followed most[17] in both academia and industries in the past decade.

A significant limitation of target-based approaches is over-exploitation of a target. It often reduces the target value, both from a biological and from a market perspective. Biologically, this is especially true for chronic use for resistance-prone diseases such as cancer or antimicrobial therapy. From a market perspective, the use of the same target and hence the same mechanism of action requires that new drug candidates must show significant improvement over the existing molecules that act on the same target, which in turn requires longer and larger clinical trials, making the drug development slower and costlier.

A. Structure based

A structure based approach is possible when the interactions of known active compounds/leads are known in detail, enabling the design of specific modifications to the target binding interactions.[17–19] It is a relatively recent approach, starting in the 1980s as

protein crystallography was increasingly efficient due to improvements of recombinant protein production methods, X-ray sources, and crystallography techniques. Now, with the support of high-speed computation, high-throughput X-ray crystallography techniques, and extensive databases of known structures, opportunities exist to analyze and/or predict drug-target interactions for many types of targets. Despite the availability of target structures, however, *de novo* structure-based drug design is relatively less frequent, due to the still limited ability to predict binding strengths.

B. Fragment based

The fragment based approach is based on the idea that different fragments of a molecule interact at different sites inside the binding pocket of a protein. It is a type of structure based approach. In this approach, different fragments binding to the residues inside the active site of the target are identified, and then a lead molecule is built with the fragments in such a way that would enable spatial arrangement of the fragments in proximity of the respective residues.[20–22] Generally, the lead molecules are built by anchoring the fragments as substitutions around suitable chemical scaffolds. This provides an opportunity for introducing newer, simpler and easily synthesized scaffolds. However, the interactions of the fragments anchored around a scaffold in the biological system may deviate significantly from that observed from the fragment crystal structure, or as predicted computationally. Small changes in the number of rotational bonds, atom-specific conformational changes such as amino-nitrogen inversion, and conformational changes in saturated rings may also cause significant differences in interactions at the active site. Therefore, the fragment based approach is robust when the key fragment properties that are important for binding allow for such changes. Because fragment binding geometries are only approximately reproduced in the larger molecules, diverse methods relevant to pocket identification and verification of fragment binding, such as NMR spectroscopy, may be utilized in the fragment based approach.[23]

1.3.2. Scaffold based

The scaffold based approach is a conventional approach of drug discovery and development. It is based on analyzing the structural similarities of active molecules and developing their variants.[24–27] Initially, compounds from diverse scaffolds are screened. Based on the structures of active compounds, common scaffolds are recognized and kept constant, while a library of compounds is created around the scaffold by variations in its

appendages (substitutions), enabling specific design to reproduce desired pharmacophoric features. Exploratory diversification at this stage may be called "hit explosion". A significant advantage of this approach is that it can be used independent of knowledge of the target structure. Without a target structure, this approach can be used to create SAR and QSAR hypotheses, though usually ambiguous, that predict the suitability of diverse substituents (and corresponding pharmacophoric features) at variable sites of the scaffold. If the target structure is known, that knowledge greatly reduces the ambiguity of the SAR hypotheses with concomitant improvement in the choice of substituents that may facilitate better interactions.

A specific scaffold based approach is also known as Plexxikon approach, which combines use of a conventional scaffold based approach iteratively with X-ray crystallography in order to evaluate the effectiveness of substitutents and thereby to find a hit.[27]

1.3.3. Bioprospecting

For thousands of years, humankind has used natural biodiversity for therapeutic purposes. Bioprospecting is simply a modern version of this, with a more systematic approach to investigate potentially large numbers natural compounds, for a potentially large diversity of therapy applications; natural biodiversity is huge.[28–30] A common practice in bioprospecting is to use high-throughput screens to identify bioactivity first on crude extracts of natural products. The extracts showing bioactivity are then purified further, aiming to identify individually the active natural compound or compounds, which is followed by the derivatization and further testing. Usually the high throughput screening in bioprospecting take place on both phenotypic screens as well as target-based screens.

A general limitation of bioprospecting is lack of direct identification of target and mechanism of action in phenotypic screening. When biological activity is observed via phenotypic screening, it often turns out to be difficult to identify the drug target and mechanism of action, especially when it is novel. However, this difficulty also leads to the greatest advantage of the methods: it allows the discovery of novel targets, novel mechanisms of actions, and new scaffolds. It can be a challenge to isolate the active constituent of an active extract in amounts sufficient to carry out structure elucidation. Moreover, the chemical synthesis of the natural compound for scale-up and derivatization is often a bottleneck for library synthesis for drug-development, as well as for commercial

production. Hence, while bioprospecting approach is beneficial for novelty, it is also disadvantageous for practical commercial application.

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.[31–33] Further, replacing random diversity with a biogenic biased selection of compounds also greatly enhances the efficiency.[34]

1.4. Biofocussed chemoprospecting: A hybrid approach

From above approaches, it would be apparent that each approach has its own spectrum of advantages and disadvantages. Hence, it was important to us to develop a hybrid approach that fits to the constraints of a PhD research project, maximizing the potential value, while minimizing the disadvantages and uncertainties that can be tolerated in larger scale enterprises.

With some 30000 encoding genes in the human genome, and a much larger diversity among pathogens, there would seem to be a huge number of potential drug targets. Thus, it is remarkable—even *stunning*—that there are only 324 drug targets, in total, for all FDA approved drugs[35,36]. There is a huge potential for the discovery of new drug-targets and therapeutic mechanisms. Bioprospecting approaches are certainly useful in this respect. Nevertheless, to alleviate its synthetic chemistry bottlenecks, a bioprospecting inspired scaffold-based approach may be chosen, synthesizing libraries of compounds with bio-like scaffolds for screening. This hybrid approach can be called "**biofocussed chemoprospecting**". In other words, it can be described as a type of diversity oriented synthesis[37] of bio-relevant scaffold based libraries, from a small set of starting materials, and using them to screen for 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.

2. Efficiency parameters for chemoprospecting approach

With a background about biofocussed chemoprospecting as a hybrid approach for faster, cheaper and efficient drug discovery in chapter 1, we will now discuss specific aspects of the approach, such as scopes of diversification and tailoring of library properties. This chapter intends to explain to the readers how such libraries can be designed while carefully considering and balancing the efficiency-parameters. This reading is expected to prepare the ground for the specific examples of chemoprospecting libraries chosen for this research project (to be discussed in chapter 3) and to facilitate the explanation of the specific statement of purpose / goals of the project in chapter 4.

2.1. Biofocus

The core idea of the approach is to adapt chemical synthesis to bioprospecting, but at the same time to keep advantages of synthetic efficiency to avoid the supply bottleneck of natural products. In other words, chemoprospecting libraries were designed for ease of synthesis but also with a focus on "bio-like" properties, i.e. maintaining close similarity with biomolecules. Hence, the actual relevance to biomolecules, or "biofocus", was considered the most important criterion, among many, in the selection of the scaffolds.

2.2. Structural diversification potential

While biofocus is the primary aim for good activity potential, the diversification potential of the scaffold is essential for synthesis of a library with good potential for the discovery of novel activities.[31,37] There are four different levels where diversification needs consideration.

2.2.1. Skeletal diversification

Diversity oriented synthesis aims for syntheses of more than one scaffold from same starting materials. It can be achieved with the use of one type of compound as the precursor for another type of compound (not just simple derivative, but structurally different class), or by using different reaction conditions of same precursors; both create different scaffolds. Skeletal diversification provides an opportunity to target more than one class of biomolecules, and thereby a much broader range of targets/cellular pathways. This level of diversity is further expanded by other levels.

2.2.2. Appendage diversification

For each scaffold, appendages (substitutions) at different positions provide another level of diversification. The potential of diversification increases exponentially with number of the appendage-sites and available variations for any given appendage site. For a scaffold

with only 3 variable substitution positions, and each position with just 5 variations, the total number of possible compounds using a small set of 5 substituents would be as many as 5 * 5 * 5 =125.

As with the concept behind the fragment-based approach to drug discovery, the substitutions anchored on a scaffold are the most important part of molecule, as they are the ones that interact with the biological target; the choice of such fragments determines not only diversity but also the likelihood of good target interactions.

2.2.3. Functional group diversification

The functional groups, either as parts of a scaffold or as appendages, represent characteristic pharmacophoric features, and determine the total ligand-receptor interaction. Diversity of functional groups in a library provides an opportunity for different kinds of interactions, and therefore better chances of finding hits. Hence, the substitutions should be carefully chosen from fragments important for interactions with drug-targets, corresponding to the fragments that can form hydrogen bonding, π-π stacking, hydrophobic interactions, metal-π interactions, etc.

2.2.4. Stereochemical diversification

The binding of ligand to its drug target critically depends on three-dimensional spatial arrangement of atoms. Use of different stereoisomers can provide access to different residues/coordinates at the active site. Because the three-dimensional structure changes drastically with a change from e.g. an "*S*" configuration to an "*R*" configuration at a single chiral center, the bioactivity profile can change completely, as in case of quinine (an antimalarial drug) vs quinidine (an anti-arrhythmic drug). The presence of chiral centers in a scaffold thus provides additional scope for stereochemical diversity.

2.3. Diversity of properties and drug likeness

The discussion of structural diversity in the previous section (2.2), alluded to the diversity of physicochemical and pharmacophoric properties important for the structural interactions with target molecules. The total set of properties relevant for chemical library design include others, such as molecular weight, partition coefficient, number of rotatable bonds, in addition to ones that may involve specific interactions, such hydrogen bond donor/acceptor capacities. In fact, such properties can be used as quantifiable parameters to assess the library diversity. The "appending" substitutions can serve as the means to introduce such properties on an anchoring scaffold. An efficient selection of compounds

would show such parameter values scattered across a broad range of combinations. A library designed this way, although including only an extremely small subset of the theoretical possibilities in numerical terms, would still provide diversity reasonable enough to represent the scaffold. This is analogous to "sparse matrix" types of searches to identify key parameters from a large set of variables.

It has been a recent tradition to evaluate "drug likeness" based on a set of physicochemical properties, such as molecular weight, lipophilicity, functional groups, no. of hydrogen bond donor, no. of hydrogen bond acceptor, polar surface area, etc.[38–40] Various sets of "rules" have been derived from empirical data in order to focus research efforts on "drug like" compounds. Lipinski's rule of five[41] (also known as Pfizer's rule of five) is an example. The Lipinski's rule has been largely followed "religiously" by medicinal chemists across the world. Its variant such as "rule of 3" has also been employed for fragment compounds.[42] Recently, a new measure of drug likeness based on a concept of desirability called Quantitative Estimate of Drug likeness (QED) has also been proposed.[43]

However, depending on the source of empirical data, such rules may focus on properties of relevance for a specific target area or therapeutic type, such as bioavailability, toxicity and other pharmacokinetic factors. For example, the criteria may be valid only for orally administered drugs. Further, the criteria are defined for human cells as target only; hence, these criteria may be irrelevant[44] or nearly so when the target cells are not human cells, as with as antibiotics. In the wake of recently developed drugs, such as kinase inhibitors, questions have been raised on the validity of such criteria.[45,46] For this work, the traditional drug likeness criteria were considered secondary to bio-likeness, and attempts were made to design the libraries in a way that most drug likeness criteria would be satisfied.

2.4. Ease of synthesis

With the aim to ease synthetic chemistry bottleneck typical of bioprospecting approaches, ease of synthesis is an important criterion for library design. From synthetic chemistry viewpoint, ease of synthesis can roughly be measured as the number of steps to synthesize a final compound of the series. Hence, it is important to choose those scaffolds as libraries that can be synthesized with a small number of steps, while providing good scope for diversification. It is of course also important to consider the nature of chemical reactions and the safety of reagents as parameters relevant to ease of synthesis.

2.5. Cost and availability of starting materials

The importance of cost as a criterion for library design depends on the available budget, but universally becomes a consideration for any practical application. As a model of chemoprospecting research, it is interesting to assess the cost of library synthesis. For simple estimation, the costs of the starting compounds may be used, ignoring the costs of solvents, isolation or structure analysis (which may vary widely).

The cost of a reagent is inevitably determined by its availability (natural or synthetic), and demand. Thus, for hit-finding purposes, it will generally be preferable to choose starting materials that are available commercially as non-specialized compounds.

2.6. Summary of principle

In summary, we aimed to develop the libraries of compounds that would combine similarity to biomolecules with physicochemical criteria for "drug likeness" and give good diversity at low cost. 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.

3. Libraries for biofocussed chemoprospecting

After discussions on how chemoprospecting provides a hybrid approach for drug discovery in chapter 1, and on the efficiency parameters of biofocussed chemoprospecting library design in chapter 2, this chapter will now provide the specific examples of the chemoprospecting libraries designed, synthesized and tested as a part of the PhD research.

3.1. Biomolecules as the starting points

As discussed in previous chapters, bio-relevance was the primary criteria for our chemoprospecting library design. Following this, we considered two very important biomolecules–peptides and tartaric acid–as the starting points (Figure 2).

Peptides are versatile in nature. Hence, the molecules having structural similarity with peptides may be expected to show activity on biological systems. In order to create our libraries, we decided to synthesize simple peptide derivatives with minor modification from natural peptides. The simplest peptide structure is a dipeptide, which can be either linear or cyclic. A cyclic dipeptide would be piperazine-2,5-dione. 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.

Tartaric acid is a naturally occurring compound, which can be viewed as an open chain form of ribose. The similarity with ribose can be used to mimic ATP structure, and hence may be expected to show some activity on the ATP-binding targets. Thus, this provided us with another library: Tartaric acid bisamides.
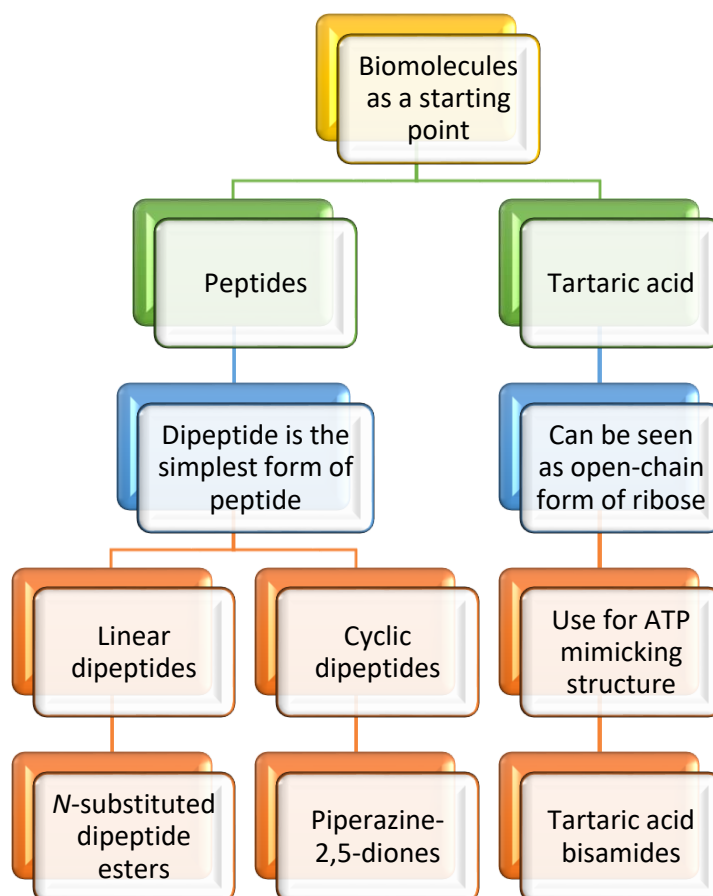
Figure 2. Biofocussed chemoprospecting libraries from biomolecules as starting points

3.2.   *N*-substituted dipeptide esters (NSDs)

3.2.1. **Peptides and peptidomimetics**

Peptides are chains of amino acid monomers linked together by covalent peptide (amide) bonds formed by coupling of carboxylic acid group of one amino acids and amino group of another amino acid. Proteins are made of one or more chains of peptides. Based on the number of monomers present, the peptides are called dipeptides (2 monomers), tripeptides (3 monomers), tetrapeptides (4 monomers) and so on. Peptide molecules having monomers between 2 to approximately 20 are collectively known as oligopeptides.

Peptidomimetics are molecules that mimic peptide structure[47,48]; they are often derivatives of natural peptides or are synthesized using scaffolds similar to natural peptide components.[49] Owing to the similarity with peptides, peptidomimetics are expected to affect peptide binding targets, such as proteases and peptide receptors. Such targets maybe found relevant to diverse pathophysiological conditions including cancer, arthritis, pancreatitis, Alzheimer's disease, and others.[50]

3.2.2. *N*-substituted dipeptide ester as scaffold

The potential of using peptide-like structures as a library for chemoprospecting is self-evident. Dipeptides are the simplest peptides, and linear dipeptide esters can also serve as precursors of cyclic dipeptides i.e. piperazine-2,5-diones. Consequently, *N*-substituted derivatives of dipeptide esters (Figure 3) were chosen for one of libraries of this research.
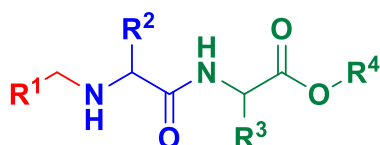
Figure 3. *N*-substituted dipeptide ester scaffold

The scientific literature describes some previous examples[51–53] of structurally related *N*-substituted dipeptide esters (henceforth NSDs). However, they differed from the compounds chosen here. Their synthetic route involved reductive amination of amino acid esters, followed by peptide coupling with *N*-protected amino acids. This strategy resulted in dipeptide esters with substitutions on the peptide bond nitrogen. However, our scaffold allows *N*-alkyl substitution on the terminal nitrogen, at the end of the chain (Figure 4), keeping the peptide bond nitrogen unsubstituted.
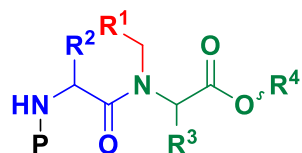
N-substituted dipeptide esters from previous examples　　**N-substituted dipeptide esters from our protocol**
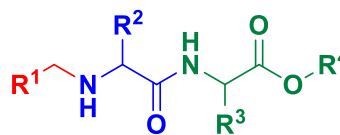
Figure 4. Difference between *N*-substituted dipeptides from previous reports and our scheme

Advantages of the scaffold

Selection of the NSD scaffold offered certain advantages. The peptidomimetic scaffold is "bio-like" and is suitable for chemoprospecting library synthesis. The scaffold offers of 4 sites for variation. Further, there are two chiral centers in the scaffold, which means that the scaffold offers effectively 6 stereospecific sites to introduce 4 chemical groups, providing high potential for diversification. The opportunity of introducing 4 desired fragments at 6 stereospecific positions can be important to tailor the physicochemical and pharmacophoric properties, which is a very important benefit for chemoprospecting, as well as for optimization after hit-finding. A significant benefit of this scaffold is also that it can be synthesized from simple, cheap and safe-to-use starting materials such as aldehydes, amino acids and amino acid esters, which are available commercially with good diversity.

3.3.　Piperazine-2,5-diones (2,5-diketopiperazines, DKPs)

3.3.1.　Marine bioprospecting background

Marine biodiscovery and bioprospecting is one of the most important areas of Norwegian research. In the course of the research on marine bioprospecting and biodiscovery, many marine species have been studied by UiTø and MabCent. As a part of such studies, the antimicrobial and antifouling activity of extracts of *Synoicum pulmonaria,* and isolation and characterization of the active molecules, such as synoxazolidinones A, B, C., pulmonarins A and B etc. have been described.[54,55] Synoxazolidinone A and B have been observed to have antimicrobial activity while synoxazolidinone C also possesses anticancer cytotoxic activity in addition.[56,55] Similarly, ianthelline from an arctic Sponge *Stryphnus fortis* has been established to possess anticancer activity via kinase inhibition activity.[57] Further, anti-inflammatory and anti-oxidant activity of barettin from a marine sponge *Geodia barrette* has also been reported.[58] Many hybrid variants of synoxazolidinones,

ianthelline and barettin have been synthesized and tested for antimicrobial and anticancer activity. Among those, the compounds containing a simplified 6-member piperazine-2,5-dione (2,5-diketopiperazine, referred to as "DKP" henceforth) ring structure (as in barettin) were synthesized[55], replacing the 5-member 4-oxazolidinone ring (e.g. synoxazolidinones) did not affect the antifouling activity.

Synoxazolidinones A and B
X = Cl : Synoxazolidinone A,
X = H : Synoxazolidinone B

Synoxazolidinone C

Barettin

Ianthelline

Debromobarettin

Where, R^1 =

R^2 =

Where, R^1 = Aromatic carbocycles and heterocycles such as phenol, indole, pyridine, thiophene, furan etc.

R^2 and R^3 = Side chains of amino acid such as tryptophan, phenyl alanine, tyrosine, alanine etc.

Piperazine-2,5-diones
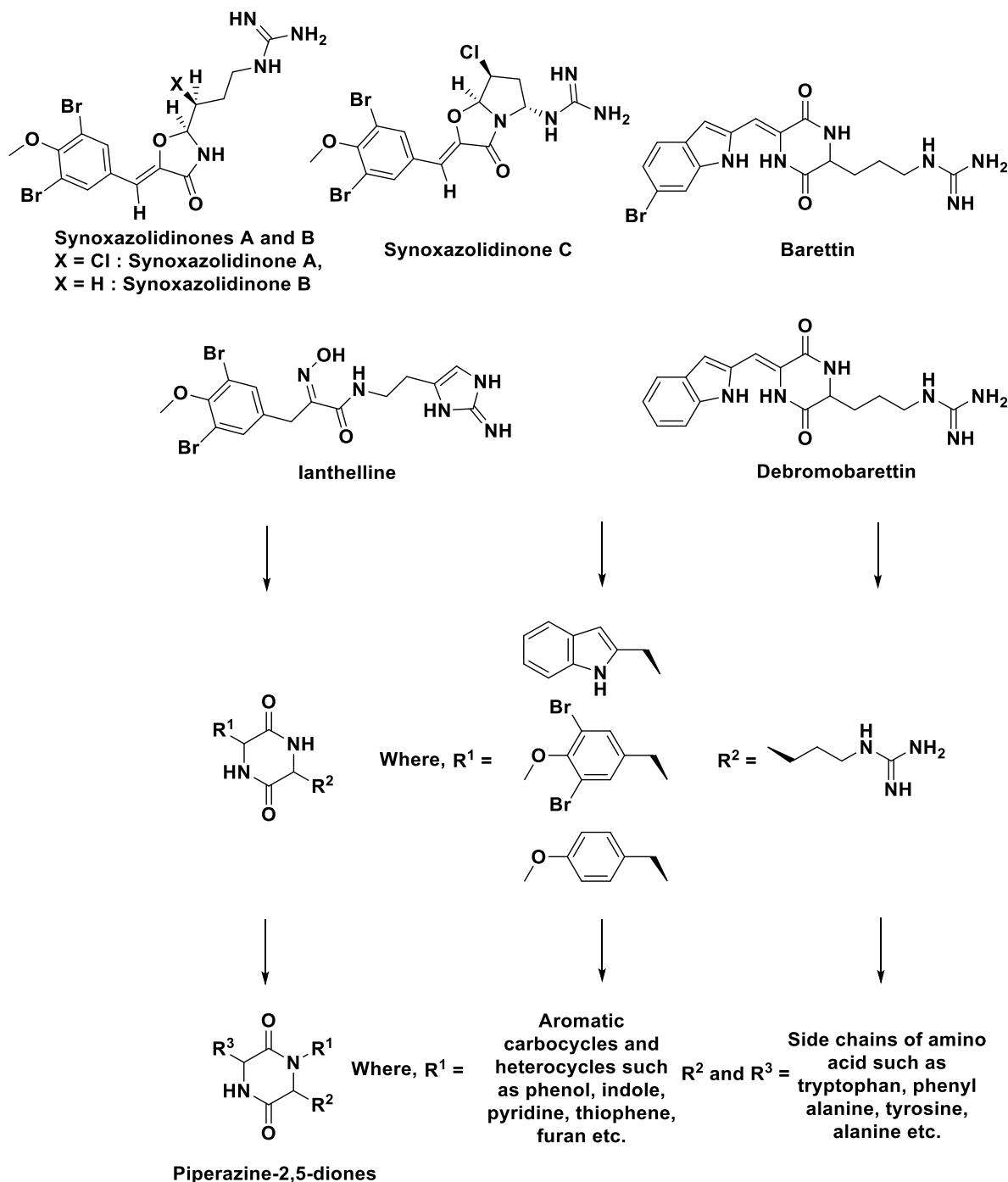
Figure 5. From marine bioprospecting research to chemoprospecting of DKPs

16

3.3.2. Use of the DKP scaffold as an anchor of important fragments

The above examples demonstrated that DKP could prove a good anchoring scaffold for introduction of suitable substitutions, such as various amino acid side chains or aromatic substitutions as in synoxazolidinones, barettin, ianthelline and other analogues. In line with our chemoprospecting approach, the DKP derivatives would then be available for various bioactivity assays. Thus, as shown in Figure 5, the evolution of a marine biodiscovery process eventually came to include chemoprospecting using the DKP library.

3.3.3. DKP as a bioactive scaffold

DKP as a scaffold has been well established[59,60] for its therapeutic potential. An interesting case[61] is the development of tadalafil, a DKP containing PDE5 inhibitor, from a hydantoin lead. Merck patented[62] piperazine and DKP derivatives as tachykinin receptor antagonists, while Novo Nordisk and Boehringer patented[63] DKP derivatives as MC4 receptor agonists for anti-obesity treatment. A series of pyridyl-2,5-diketopiperazines have been shown[64] to be orally bioavailable potent oxytocin antagonists.

Apart from the therapeutic areas mentioned above, the DKP scaffold has also been used as an effective core for anticancer drug development. In the pursuit to develop DNA binding agents, DKP derivatives similar to anthracyclin derivatives such as daunorubicin and doxorubicin were also developed[65,66] and their anti-cancer activities were established. Exploiting the zinc binding property of thiol group, various thiol containing DKP derivatives have been developed[67–69] as matrix metalloproteinase inhibitors. Qiao et al.[70] reported DKP derivatives as dual inhibitors of farnesyltransferase and geranylgeranyltransferase-1. While these derivatives were similar to our designed library, the library compounds had substitutions only at one carbon. In contrast, our library was not designed against any specific target and included diverse substitutions on both carbons.

3.3.4. Benefits of the DKP scaffold
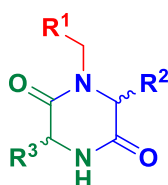
Figure 6. General DKP scaffold for our compounds

In addition to the similarity with previously reported compounds, the DKP scaffold (Figure 6) offered a few benefits also.

1) It is possible to introduce chemically 3 substitutions (R^1, R^2, R^3) at 5 possible stereospecific spaces, allowing to address greater possibilities of interactions inside active site. It is also possible to exploit the 2nd nitrogen to introduce a 4th substitution if required.

2) DKP occurs in two conformational states[59,71–73], flat or slightly puckered boat conformations. There is only a minor energy difference between both conformations, which allows DKP scaffold to interchange its conformation between both conformational states. From medicinal chemistry point of view, it is interesting as this enables DKP scaffold to address an even greater diversity of pharmacophores with minor changes in structure.

3) The polarity of the overall structure heavily depends on the selection of 3 substitutions. Hence, the scaffold allows necessary flexibility for pharmacokinetic considerations for drug-development.

4) The most common starting materials for synthesis of the scaffold are aldehydes, amino acids and amino acid esters. These starting materials are cheap, and readily available in a wide variety. They are also easily accessible from other class of compounds for introduction of specific substitutions when required.

3.4. Tartaric acid bisamides (TABs)

3.4.1. ATP binding proteins as possible drug targets

ATP is one of the most important biomolecule in nature, often considered as a universal energy-currency inside a cell. ATP serves multiple roles in cellular functions such as metabolism, active-transport, cell-signaling and synthesis of other biomolecules including DNA and RNA. With such a wide-range of functions, it is apparent that ATP binds to multiple cellular targets, such as kinases and ATPases, and these targets can be exploited to elicit significant therapeutic activities.[74,75]

3.4.2. The rationale behind the tartaric acid bisamide (TABs) scaffold

As shown in Figure 7, tartaric acid can be viewed as an open chain analogue of ribose, with the vicinal diol of tartaric acids as similar to 2'- and 3'- hydroxyl groups of ribose. The stereospecific considerations can be explored with use of different stereoisomers of tartaric acid. Further, the conversion of both carboxyl groups of tartaric acid to amides can be a

versatile approach to introduce favourable substitutions at both ends. From the varieties of commercially available amines and accessibility of amines from other classes, the scaffold may be developed into a new class of compounds for ATP-binding targets.
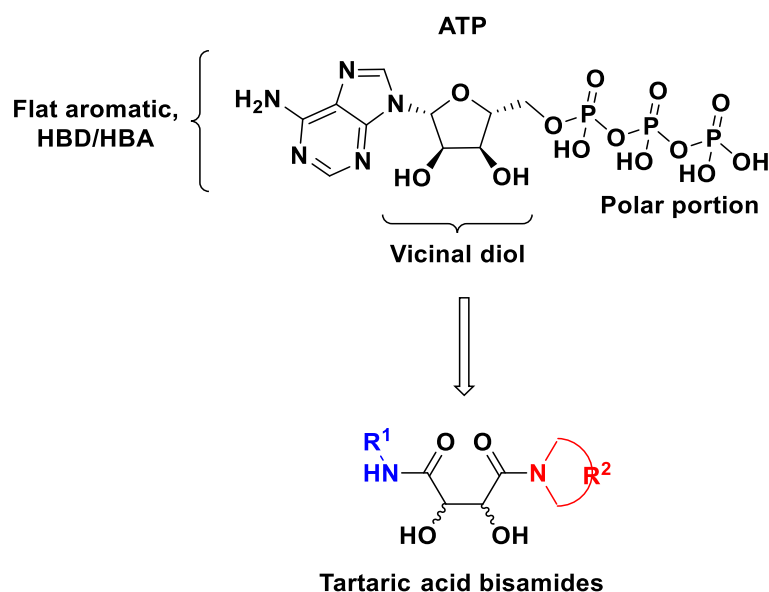
Figure 7. Tartaric acid bisamides as open chain analogues of ATP

3.4.3. Benefits of TAB scaffold

Compared to NSD or DKP scaffolds, TAB is a smaller scaffold, which makes TABs more like fragments that can be further extended via substituted fragments on both variable positions. Moreover, stereospecific vicinal diol functionality provides more potential to tailor physicochemical and pharmacophoric properties. This is compatible with importance of generating highly specific analogues of ATP, which itself is a highly nonspecific biomolecule.

19

4. Statement of purpose / Goals

Based on the principle of the "biofocussed chemoprospecting" approach described in previous chapters, the overall goals of the research work were defined as follows:

Goal: Design and synthesis of diverse chemical libraries based on bio-relevant scaffolds with "appropriate" properties (foci) and their bioactivity studies.

<u>Subgoals</u>:

- To employ a hybrid approach – biofocussed chemoprospecting - to minimize limitations of conventional approaches of hit-finding and lead determination, such as bioprospecting and scaffold based approaches, while maximizing their efficiency/opportunities.

- Determine scaffolds and substitutions for compound-libraries fulfilling desired criteria.

- Design easy and short synthetic routes with conventional, customized, or novel methods, followed by the synthesis of libraries and assessment of the protocols.

- Test the synthesized libraries at different platforms for bioactivities including phenotypic assays and cell-based assays. Create target hypotheses for the compounds found active.

Results and Discussion

In the introduction part, a brief outline of biofocussed chemoprospecting libraries was given as a theoretical treatise with examples of three libraries inspired from biomolecules. In this part, the practicalities of this approach will be demonstrated using the same libraries, with details on library design, synthesis, cheminformatics analysis and bioactivity studies (including computational approaches).

5. Library I: *N*-substituted dipeptide esters (NSDs)

This chapter first describes synthesis of N-substituted dipeptide ester library, followed by the cheminformatics analysis of library properties regarding the efficiency parameters described in the introduction part. The discussion will then focus on bioactivity studies and initial target identification approaches. These comprise especially assays on cancer cell lines and proteomics analysis with the SILAC method, protease profiling, target-prediction efforts using cheminformatics fingerprinting based models and BioPrint® profiling. Overall, this chapter describes the N-substituted dipeptide esters library as a practical example of a biofocussed chemoprospecting library.

5.1.   Synthesis

5.1.1. Approach for synthesis

As described in chapter 3, the *N*-substituted dipeptide esters can be synthesized from common commercially available starting materials, such as aldehydes, amino acids and amino acid esters. As a dipeptide ester scaffold, a peptide coupling protocol would also be the most convenient approach. In order to introduce substitutions on the terminal nitrogen, the amino acid may be alkylated before the peptide coupling. The retrosynthetic analysis can be explained as below in Scheme 1.
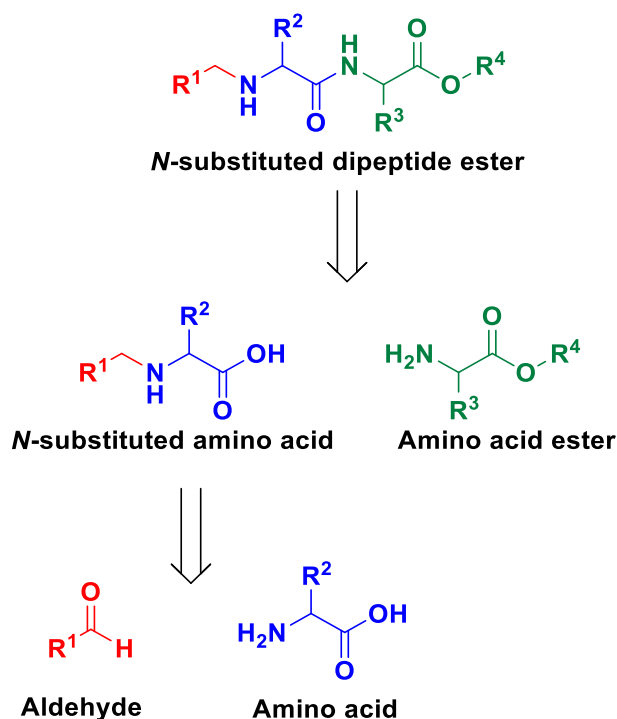
Scheme 1. Retrosynthesis for *N*-substituted dipeptide esters

Following the retrosynthetic analysis, the two-step synthetic scheme (Scheme 2) was designed. Unlike the previously reported examples discussed in chapter 3, we chose to carry out reductive amination of amino acids to generate *N*-substituted amino acid, followed by its peptide coupling with an amino acid ester.
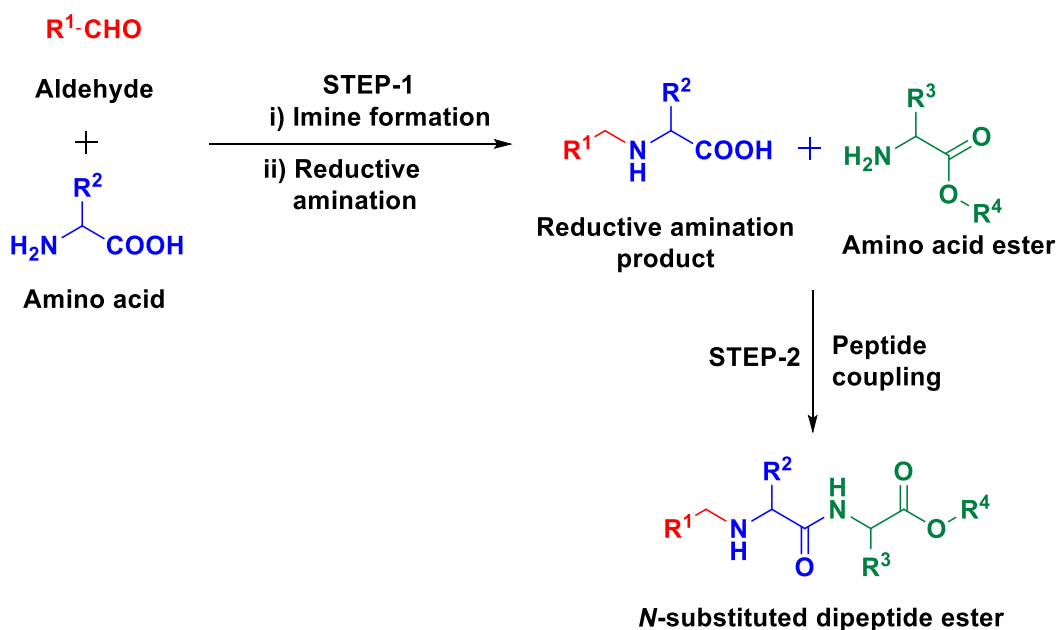
Scheme 2. Synthesis scheme for *N*-substituted dipeptide esters

As typical peptide coupling requires many steps of protection and deprotection, it is common to carry out peptide couplings using solid phase synthesis. In our case, for couplings using uronium coupling agents such as HBTU, the reactivity difference[76] between the 2°-amine (of *N*-alkylated amino acid) and the 1°-amine (of amino acid ester) provided an opportunity to skip protection-deprotection steps. Therefore, the solid phase synthesis, as in previous examples, was no longer required. Thus, using a slightly different strategy for synthesis, we avoided the need of protection and deprotection steps, decreasing the number of steps and making the synthesis easier, cheaper and more compliant with our approach.

5.1.2. Selection of R¹, R² and R³ substitutions

As described in the introduction part, appending substitutions play a major role in determining physicochemical properties of final compounds. For efficient library synthesis, it was important to select the substitutions that can provide sufficient diversity with respect to such properties at low cost.

As per the synthetic strategy, R^1 comes from an aldehyde, while R^2 and R^3 come from amino acid and amino acid esters. Aldehydes were selected among diverse aromatic carbocycles or heterocycles such as phenol, pyridine, thiophene, furan, indole etc. Amino acids and amino acid esters were chosen to be either natural ones (L-isomers) or their enantiomers (D-isomers), except for proline which is an imino acid. Further, acidic or basic amino acids were excluded in order to make synthesis simpler. Hence, out of 19 natural amino acids (excepting the imino acid proline) we selected neutral amino acids and their esters such as:

- Glycine (polar amino acid with no substitution)
- Alanine (relatively non-polar amino acid with small methyl substitution)
- Phenyl alanine (hydrophobic aromatic substitution without any HBD or HBA)
- Tyrosine (partially hydrophobic aromatic substitution with hydrogen bond donor and acceptor groups) and
- Tryptophan (large hydrophobic aromatic heterocyclic substitution)

Except for L-phenylalanine (ethyl ester), all amino acid esters used were methyl esters.

5.1.3. *N*-monoalkylation via reductive amination

N-monoalkylation of amino group is a common method to create a C-N bond. There are many ways to achieve *N*-alkylation, including simple nucleophilic substitution of halides, using alcohol with palladium catalyst in presence[77] or absence[78] of hydrogen, reductive amination, the Buchwald-Hartwig reaction etc. Though the nucleophilic substitution is simple and straightforward, multiple alkylation and quaternary salt formation are its main limitations.

Reductive amination is a common method for *N*-monoalkylation of amino acids. The reductive amination of amino acids has been reported with reagents such as sodium borohydride[79–81], sodium cyanoborohydride[82–84], 1,2,3-triazole-boranes[85]. Typically, the reaction takes place in two steps, whereby the first step requires imine formation between aldehyde and amino acid. As the imine formation is reversible, one reagent is generally used in excess to drive the equilibrium towards the imine side, which is then converted to amine by reducing agents such as borohydrides and boranes.
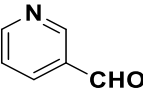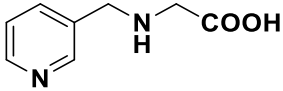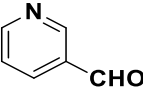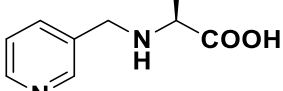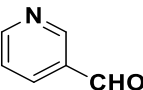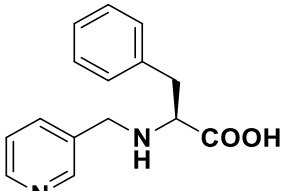
Using a protocol[81] reported by Verardo et al. (protocol A in Scheme 3), different aldehydes were reacted with amino acids (Table 2) in methanol. After allowing imine formation, they were reduced with sodium borohydride. To drive the equilibrium in the forward direction,
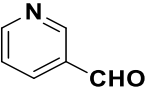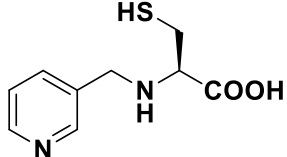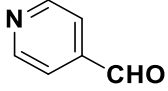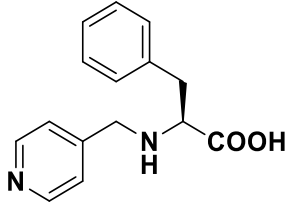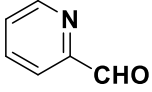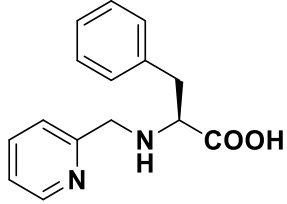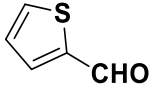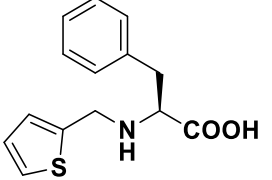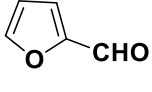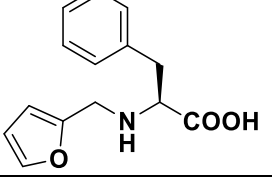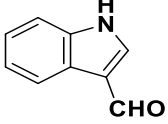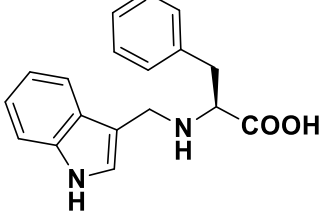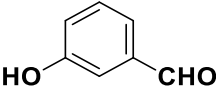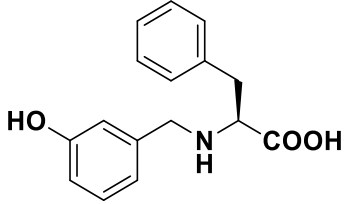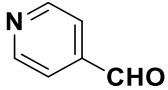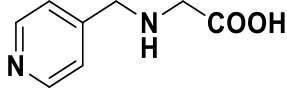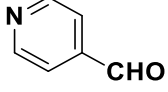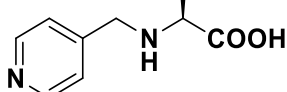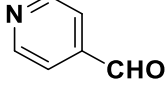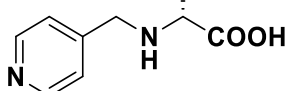
a 1.4 equivalent amount of aldehyde was taken. The reactions were mostly clean, with only a few exceptions. As the products were highly polar, it was not possible to isolate the impure products. As all products were also more or less water-soluble, the water-soluble salts were not separated and the products were used in crude form for the next step. Hence, it was not possible to calculate yields.
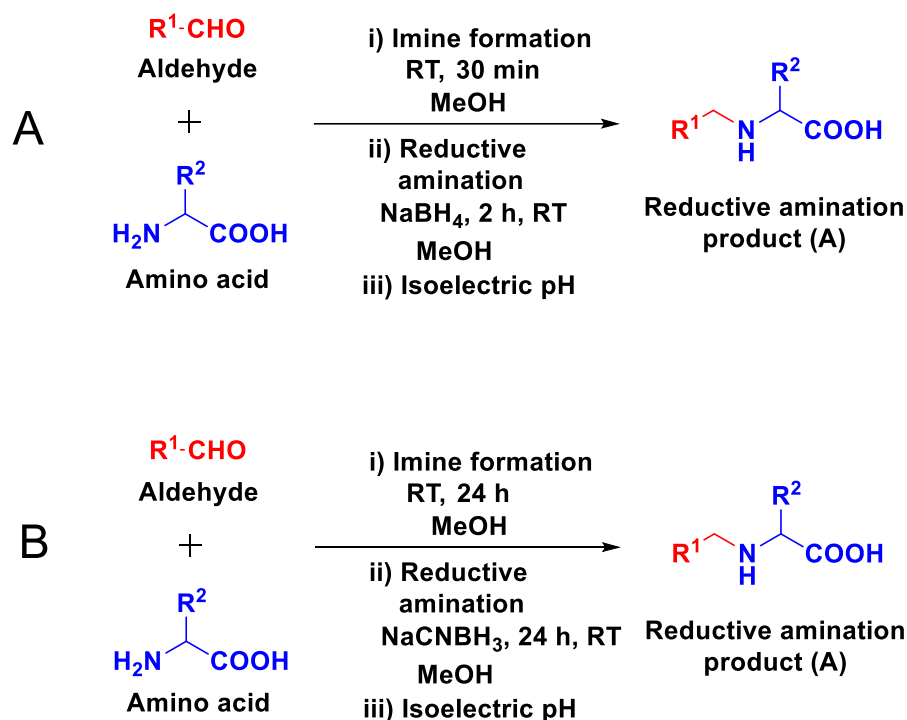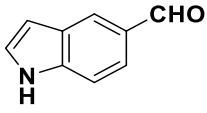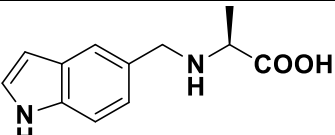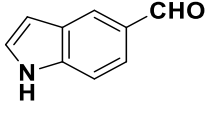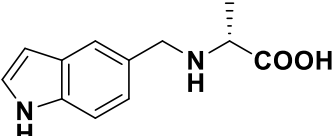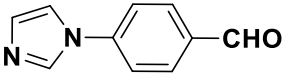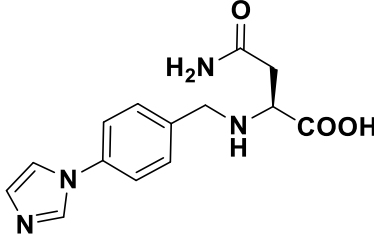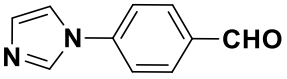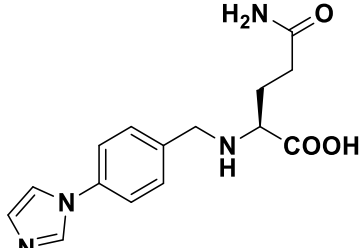
Scheme 3. Synthesis of *N*-substituted amino acids via reductive amination

Table 2. Table for reductive amination reaction (Step-1)

Aldehyde	Amino acid	Protocol	Product	Product structure
	Glycine	A[a]	A1	
	L-Alanine	A	A2	
	L-Phenylalanine	A	A3	

	L-Cysteine	A[b]	A4	
	L-Phenylalanine	A	A5	
	L-Phenylalanine	A	A6	
	L-Phenylalanine	A	A7	
	L-Phenylalanine	A	A8	
	L-Phenylalanine	A[b]	A9	
	L-Phenylalanine	A	A10	
	Glycine	A	A11	
	L-Alanine	A	A12	
	D-Alanine	A	A13	

	L-Alanine	B	A14	
	D-Alanine	B	A15	
	L-Asparagine	B[b]	A16	
	L-Glutamine	B[b]	A17	

[a]Extremely hygroscopic product, could not be dried for long to remove acetone
[b]Impure products – not used further

The reaction of indole-3-carboxaldehyde with phenylalanine was not clean. Hence, the reactions of indole-5-carboxaldehyde and 4-(1*H*-imidazol-1-yl)benzaldehyde were carried out using a different protocol (protocol B in Scheme 3), using sodium cyanoborohydride to allow a longer reaction time for imine formation and selective reduction of imines driving the equilibrium to forward direction.

5.1.4. Peptide coupling

A typical peptide coupling protocol involves activation of carboxylic acid using singly or in combination, coupling reagents such as HBTU, PyBOP, 3°-amine like diisopropylethylamine (DIPEA), in aprotic solvents like DMF and then adding amine to form the amide product. After completion, the reaction mixture is usually washed with citric acid solution to remove excess amines and water soluble impurities, and is extracted with ethylacetate or DCM. The crude reaction mixture thus obtained is subjected to purification.
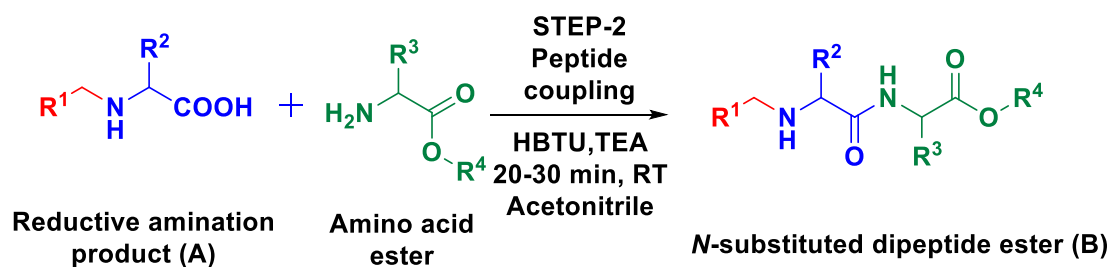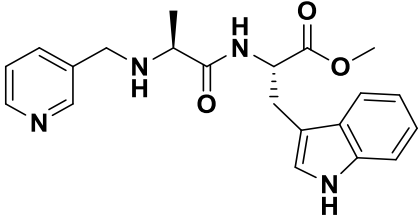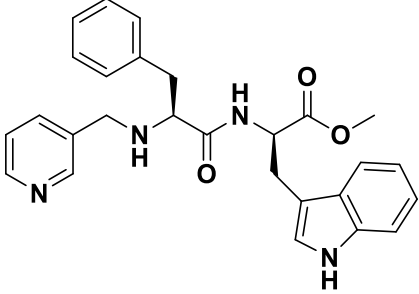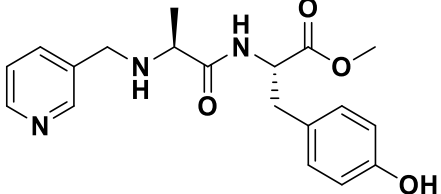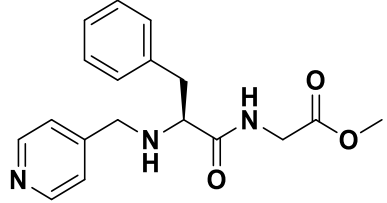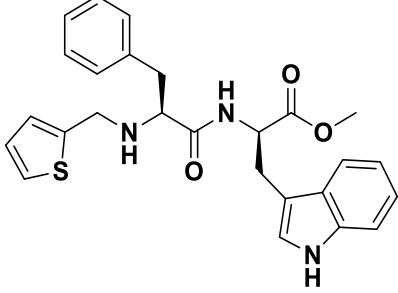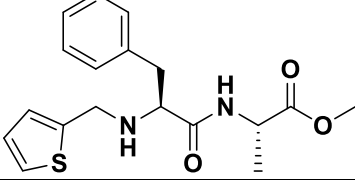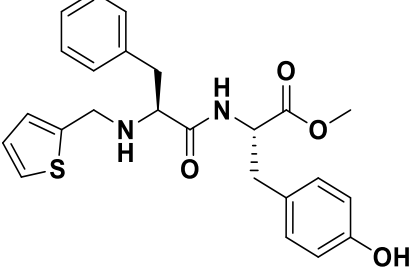
Scheme 4. Peptide coupling reaction for synthesis for NSDs
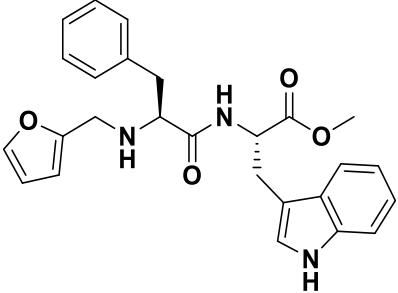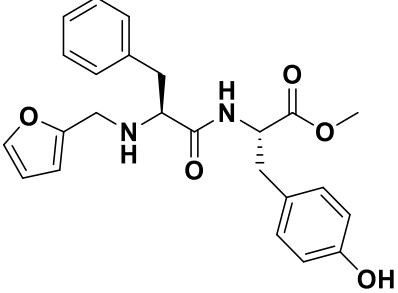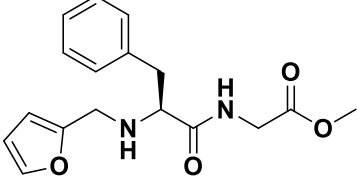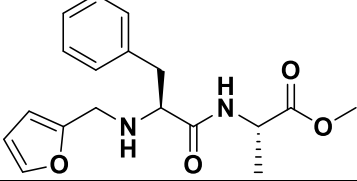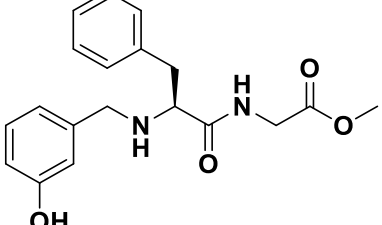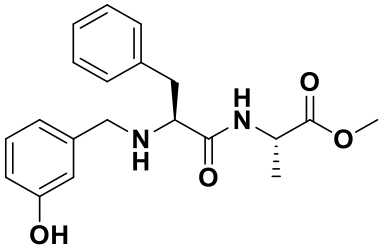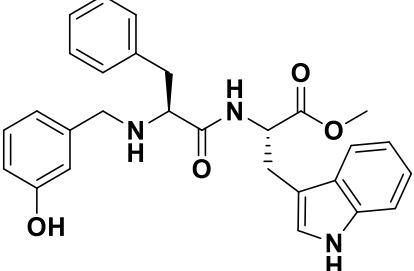
In our case (Scheme 4), the NSDs were synthesized using a modified peptide coupling protocol (Table 3). Most of the products were highly or partially soluble in water. Moreover, the products also contained an active 2°-amine group that would react with acids. Hence, the reaction mixtures could not be washed with aqueous citric acid solutions in order to remove reagents like excess DIPEA. Thus, it was necessary to substitute DIPEA with a 3°-amine, which can be evaporated when in excess. So, triethylamine (TEA) was used. Instead of a high boiling point solvent like DMF, a low boiling point aprotic solvent acetonitrile was used. R^4 was kept as a methyl group for all compounds except B28, which contained ethyl group as R^4.
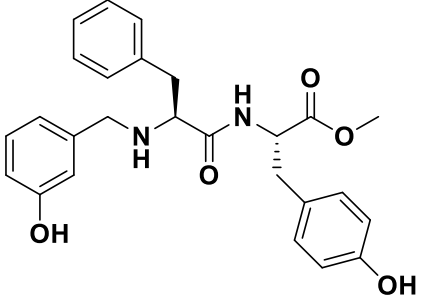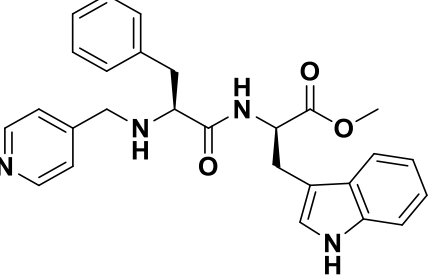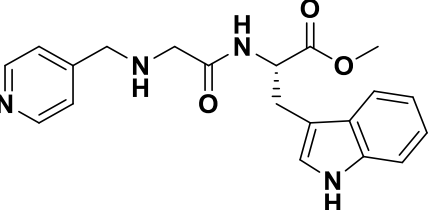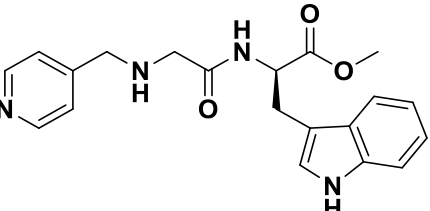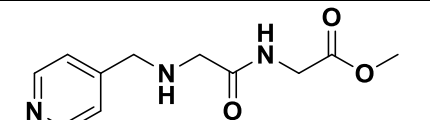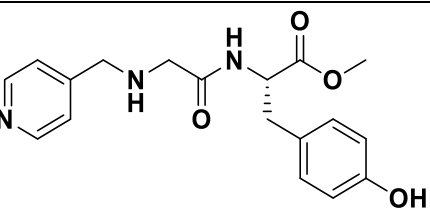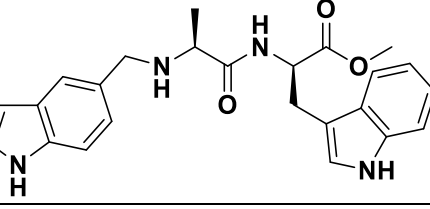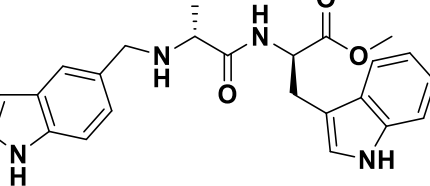
Table 3. Table for peptide coupling reaction (Step-2)
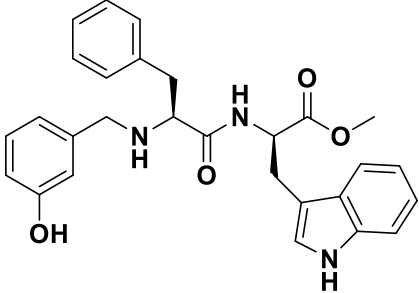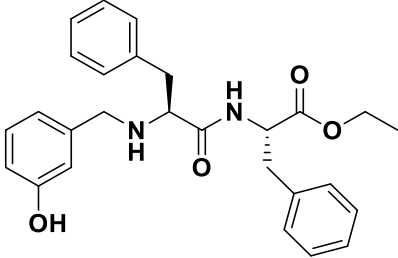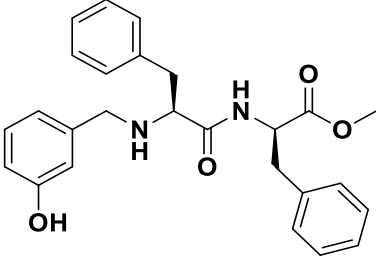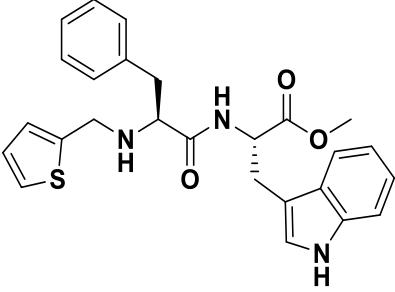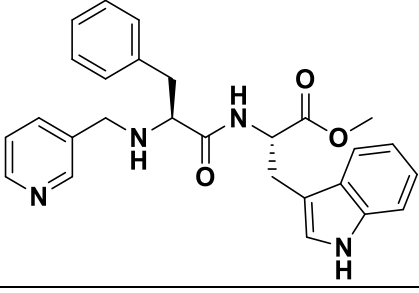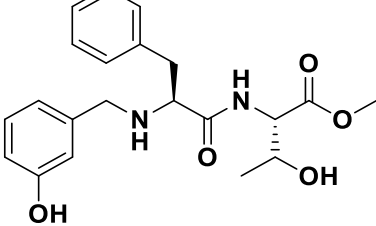
Step-1 products	Amino acid ester	Product	Product structure	Isolated Yield %
A1	GlyOMe	B1		73[a]
A2	GlyOMe	B2		49[a]
A3	GlyOMe	B3		67[a]
A3	L-AlaOMe	B4		52[a]

28

A2	L-TrpOMe	B5		53[a]
A3	D-TrpOMe	B6		53
A2	L-TyrOMe	B7		58
A5	GlyOMe	B8		55
A7	D-TrpOMe	B9		47[a]
A7	L-AlaOMe	B10		45
A7	L-TyrOMe	B11		47

A8	L-TrpOMe	B12		43
A8	L-TyrOMe	B13		48[b]
A8	GlyOMe	B14		68[b]
A8	L-AlaOMe	B15		63[b]
A10	GlyOMe	B16		61
A10	L- AlaOMe	B17		54
A10	L-TrpOMe	B18		45

A10	L-TyrOMe	B19		46
A5	D-TrpOMe	B20		50
A11	L-TrpOMe	B21		43
A11	D-TrpOMe	B22		39
A11	GlyOMe	B23		59
A11	L-TyrOMe	B24		64[a]
A14	D-TrpOMe	B25		50
A15	D-TrpOMe	B26		43

A10	D-TrpOMe	B27		52
A10	L-PheOEt	B28		47
A10	D-PheOMe	B29		53
A7	L-TrpOMe	B30		54
A3	L-TrpOMe	B31		48
A10	L-ThrOMe	B32		44

[a]5-10 % cyclic dimer was formed
[b]Product has very low UV activity

5.1.5. Cyclic dimer formation as a self-condensation product

After the carboxylic group of amino acid is activated by a coupling agent like HBTU, it can react with both 1°- and 2°-amines. As Scheme 5 shows, if the activated carboxylic group reacts with 2°-amine (instead of the 1°-amine of amino acid esters, which would give an NSD product) of another molecule of the same starting material, it leads to a cyclic dimer as a self-condensation product. However, the rate of reaction for coupling with 1°-amine is much faster than that with the 2°-amine. While designing the reaction scheme, we attempted to utilize the difference in order to avoid protection and deprotection steps. It was observed that in some reactions, the self-condensation product was still formed in small amounts, even as the dipeptide product was the major one.
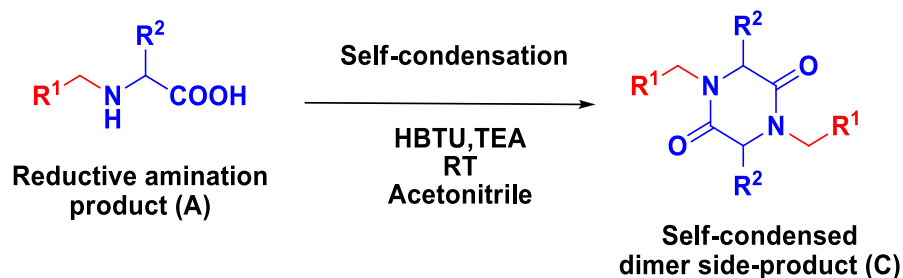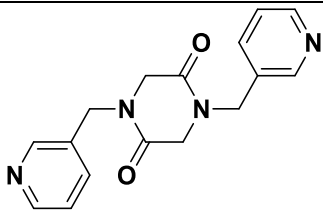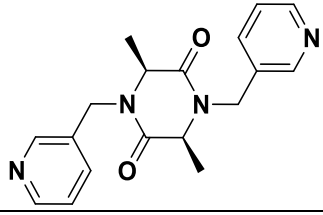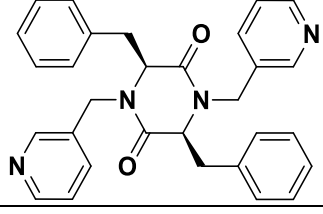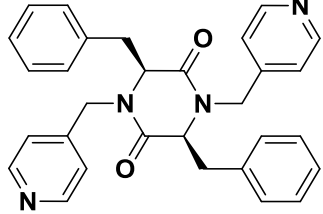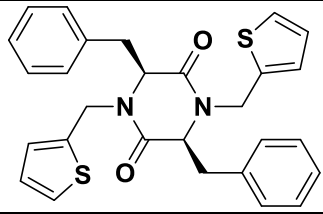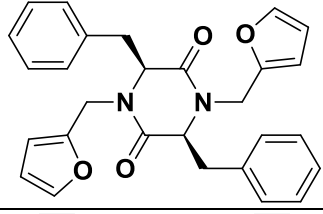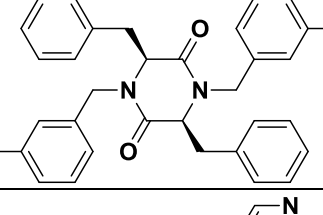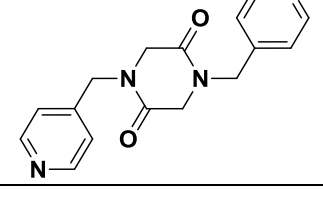
Scheme 5. Cyclized dimer formation as a result of self-condensation of *N*-substituted amino acids

As the cyclized dimer product is a symmetrical DKP derivative, this side-product formation was considered an opportunity to generate a new series. Hence, when such dimer product was formed as a side product, it was also isolated along with the desired NSD product. However, as we had expected, such self-condensation product formation was not common. When it did not form as a side product, the reactions were set up according to the same protocol, but without addition of amino acid ester (Table 4). The rationale was to facilitate the self-condensation in absence of 1°-amine.

From the perspective of synthesizing a new series, the strategy worked quite well, as it was possible to synthesize several compounds with this method. It is important to note that specific reactions were set up only when self-condensation did not occur as a side-reaction, implying that the self-condensation was not easy in such cases. Hence, the seemingly poor yields for such reactions also appear relatively good.

Table 4. Cyclic dimer (symmetrical DKP) product formation.

Step-1 products	Product	Product structure	Isolated Yield %
A1	C1		Isolated as a side-product
A2	C2		34
A3	C3		26
A5	C4		29
A7	C5		Isolated as a side-product
A8	C6		No product formation
A10	C7		No product formation
A11	C8		Isolated as a side-product

5.2. Library properties

5.2.1. Diversifiability

The starting materials for this series such as aldehydes and amino acids are commercially available in wide varieties. Amino acid esters are also commercially available or easily accessible from corresponding amino acids. This implies that, consistent with the idea of chemoprospecting, there is a huge diversification potential for this library. A latest Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO, USA) catalogue mentions 1244 aldehydes and 551 amino acids. Even if we consider 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!

The computation required to evaluate such a huge chemical space exhaustively with respect to different properties, is prohibitively expensive, so we decided to create its small subset as a virtual library and then to analyze this virtual library computationally to study the distributions of various properties for these compound sets.

As Mao et al. have noted[86], an overwhelming majority of approved drug molecules contain at least one aromatic ring. Additionally, the presence of at least one aromatic group in a compound can be helpful in chromatographic isolation. However, not all amino acids and amino acid esters are aromatic. Hence, out of more than 1200 varieties of aldehydes, 22 aromatic / heterocyclic aldehydes were chosen as R^1. All the aldehydes were chosen from simple aromatic carbocycles or heterocycles such as phenol, thiophene, furan, pyridine, indole, quinoline, imidazole etc. Similarly, out of more than 500 varieties of amino acids and amino acid esters, only natural isomers (L-isomer) and their unnatural enantiomers (D-isomer – except for achiral glycine) were chosen as R^2 and R^3 respectively (Table 5). Being an imino acid, both isomers of proline were excluded from amino acids and amino acid esters. Using the JChem software suite (ChemAxon, Budapest, Hungary), a virtual library of 22 * 37 * 37 = 30118 compounds was generated. The virtual library was subjected to computational analysis.

Table 5. Aldehydes, amino acids and amino acid esters chosen for virtual library generation.

No.	Aldehydes	Amino acids	Amino acid esters
1	4-pyridaldehyde	Glycine	Glycine methyl ester HCl
2	3-pyridaldehyde	L-Alanine	L-Alanine methyl ester HCl
3	2-pyridaldehyde	L-Serine	L-Serine methyl ester HCl

4	4-hydroxybenzaldehyde	L-valine	L-valine methyl ester HCl
5	3-hydroxybenzaldehyde	L-leucine	L-leucine methyl ester HCl
6	2-hydroxybenzaldehyde	L-isoleucine	L-isoleucine methyl ester HCl
7	Thiophene-2-carboxaldehyde	L-tyrosine	L-tyrosine ethyl ester HCl
8	Thiophene-3-carboxaldehyde	L-phenylalanine	L-phenylalanine ethyl ester HCl
9	Furan-2-carboxaldehyde	L-cysteine	L-cysteine ethyl ester HCl
10	Furan-3-carboxaldehyde	L-threonine	L-threonine methyl ester HCl
11	Imidazole-4-carboxaldehyde	L-methionine	L-methionine methyl ester HCl
12	4-(1-imidazolyl)-benzaldehyde	L-tryptophan	L-tryptophan methyl ester HCl
13	*N*-methyl-Imidazole-2-carboxaldehyde	L-histidine	L-histidine methyl ester diHCl
14	*N*-methyl-Imidazole-5-carboxaldehyde	L-asparagine	L-asparagine *tert*-butyl ester HCl
15	4-anisaldehyde	L-glutamine	L-glutamine methyl ester HCl
16	3-anisaldehyde	L-aspartic acid	L-aspartic acid dimethyl ester HCl
17	2-anisaldehyde	L-glutamic acid	L-glutamic acid dimethyl ester HCl
18	4-(*N,N*-dimethylamino)benzaldehyde	L-arginine	L-arginine methyl ester diHCl
19	Quinoline-3-carboxaldehyde	L-lysine	L-lysine methyl ester diHCl
20	Isoquinoline-5-carboxaldehyde	D-Alanine	D-Alanine methyl ester HCl
21	Indole-3-carboxaldehyde	D-Serine	D-Serine methyl ester HCl
22	Indole-5-carboxaldehyde	D-valine	D-valine methyl ester HCl
23		D-leucine	D-leucine methyl ester HCl
24		D-isoleucine	D-isoleucine methyl ester HCl
25		D-tyrosine	D-tyrosine methyl ester HCl
26		D-phenylalanine	D-phenylalanine methyl ester HCl
27		D-cysteine	D-cysteine methyl ester HCl
28		D-threonine	D-threonine methyl ester HCl
29		D-methionine	D-methionine methyl ester HCl
30		D-tryptophan	D-tryptophan methyl ester HCl
31		D-histidine	D-histidine methyl ester diHCl
32		D-asparagine monohydrate	D-asparagine methyl ester HCl
33		D-glutamine	D-glutamine methyl ester HCl
34		D-aspartic acid	D-aspartic acid dimethyl ester HCl
35		D-glutamic acid	D-glutamic acid dimethyl ester HCl
36		D-arginine	D-arginine methyl ester diHCl
37		D-lysine	D-lysine methyl ester diHCl

5.2.2. Diversity

The diversity of the virtual library of NSDs was analyzed with respect to various physicochemical parameters such as molecular weight, AlogP, number of rotatable bonds and polar surface area. It can be observed in Figure 8 that the virtual library shows promising distribution patterns for different parameters.
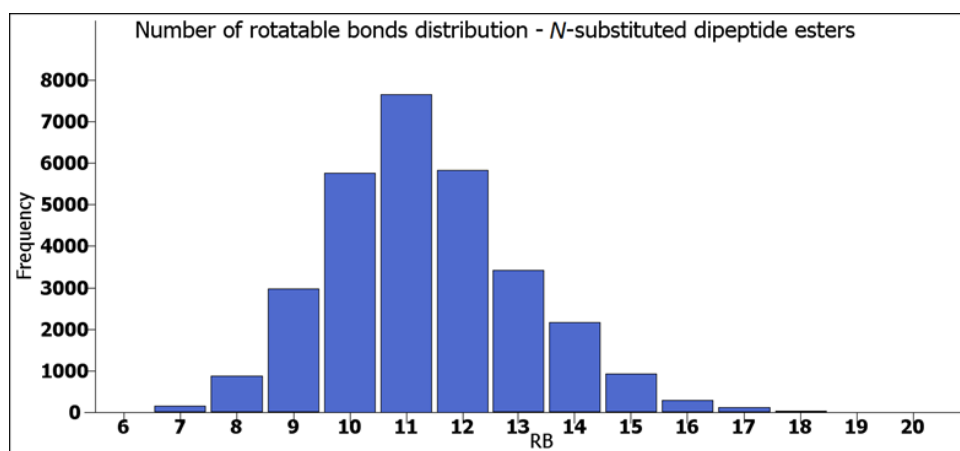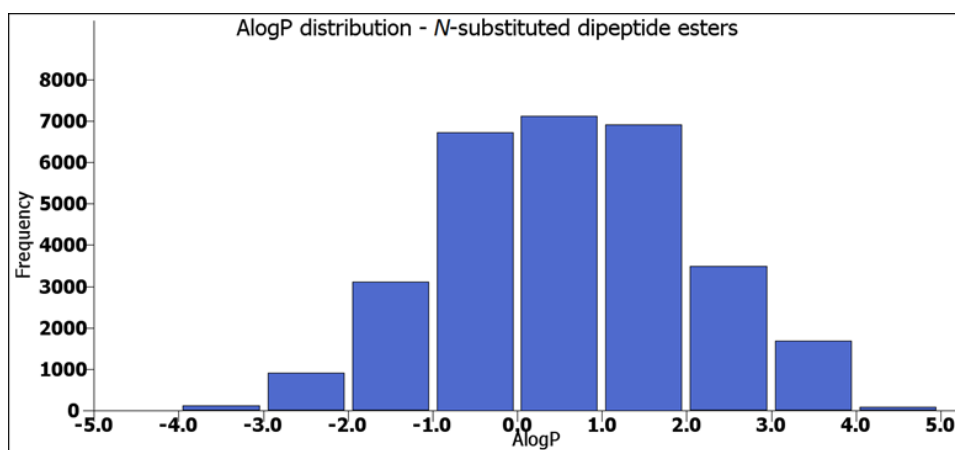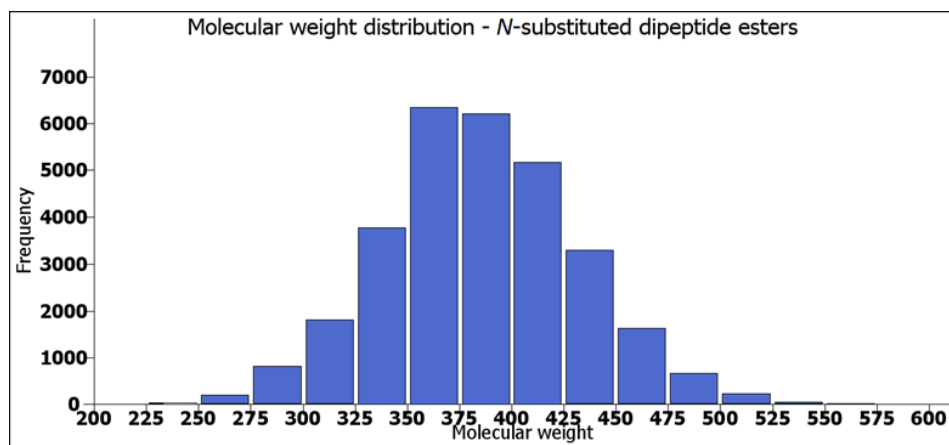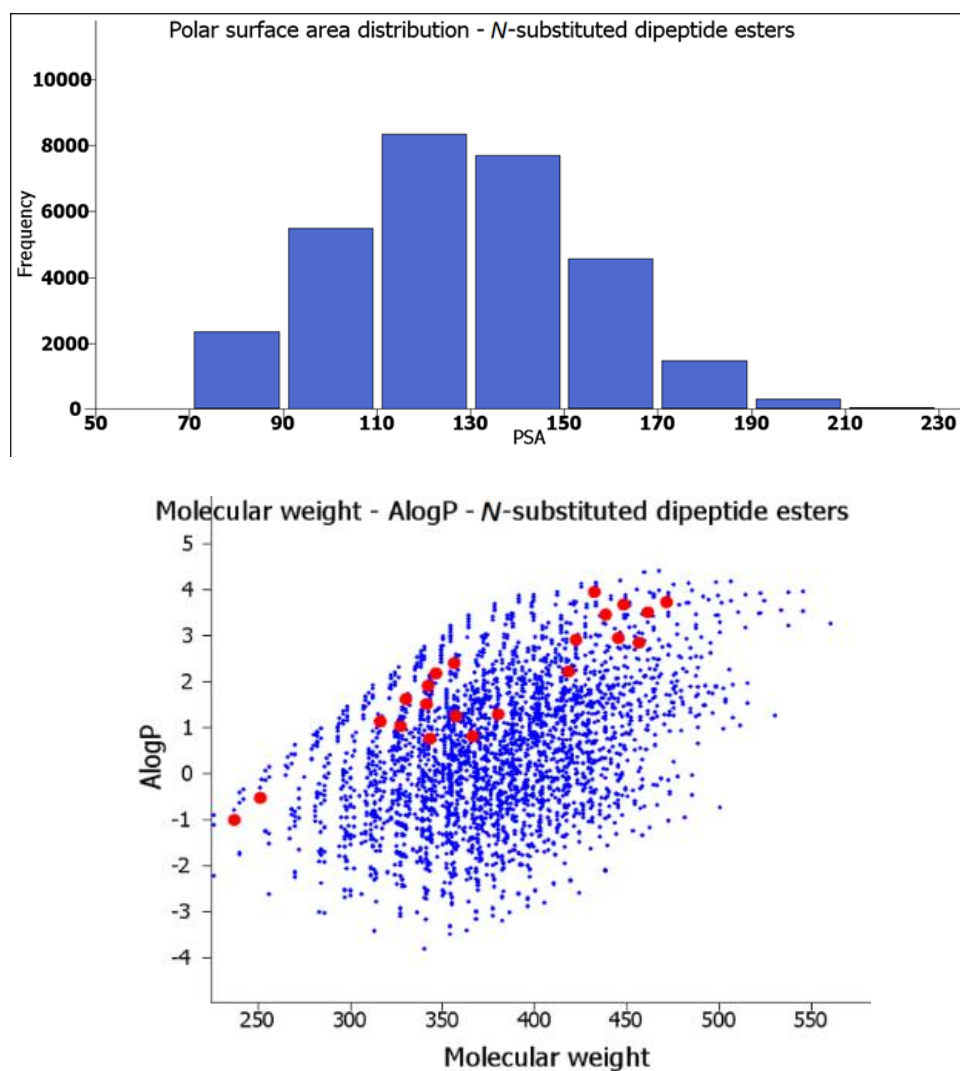
Figure 8. Distribution of a few properties for the virtual library of NSDs

5.2.3. Ease of synthesis

As explained in the introduction, the ease of synthesis can be approximately correlated with the number of steps (quantitatively) and safety parameters of the reagents and reaction conditions. The NSD library can be synthesized from commercially available materials in just two steps. Both steps are very safe, easy to scale-up and do not necessarily require work-up. The first step does not require purification as well, as the crude reaction mixture is sufficient for the second step. The second step is prone to side product formation and it needs isolation with column chromatography.

Considering multistep synthesis as a "norm" in library synthesis for drug discovery, it can be safely concluded that the overall ease of synthesis is significantly high for the NSD library.

5.2.4. Cost and availability of starting materials

The cost and availability of starting materials can make a significant impact on overall budget of research work. The starting materials for NSDs are aldehydes, amino acids and amino acid esters. They are cheap and commercially available in wide varieties.
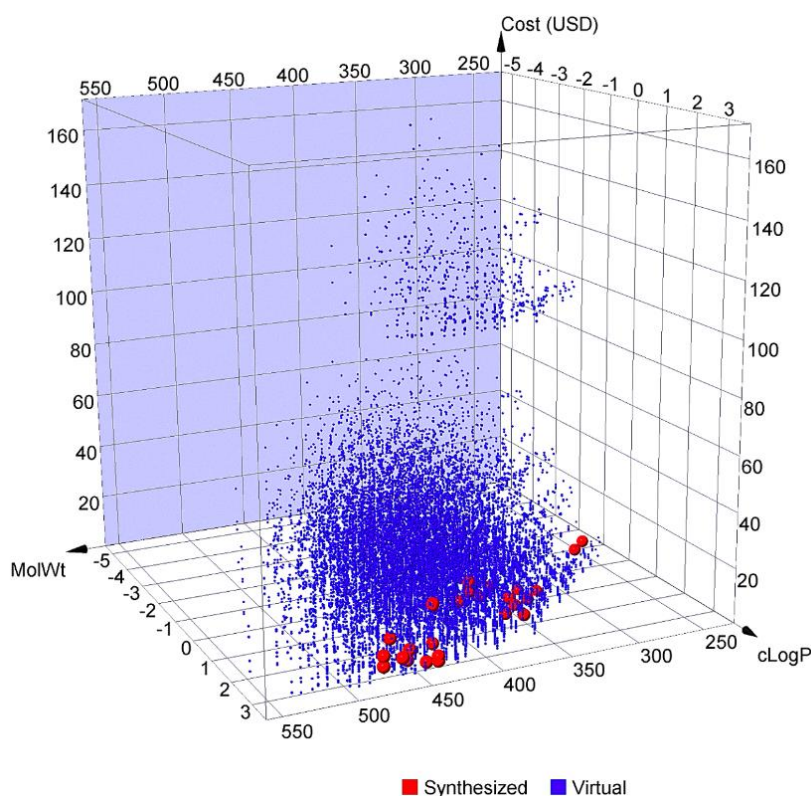
Figure 9. The plot shows 3D representation of virtual library of NSDs with respect to molecular weight, AlogP and cost. The red marks signify the synthesized library and represent the extent of diversity achieved by the chemoprospecting.

In order to estimate the cost, an estimation was made for all 22 aldehydes, 37 amino acids and 37 amino acid esters. As a standard, the necessary amount of compounds to carry out reaction at 1 mmol scale was calculated considering the mole equivalents. Based on the amount, the prices for each starting material was generated based on commercial catalogues of Aldrich (and other suppliers if the compound was not available at Aldrich). From the cost of each starting material, the costs for all 30118 compounds from virtual library were estimated. The estimated costs include only the cost of aldehyde, amino acid and amino acid esters. They do not include the isolation, purification or characterization costs because such calculations would be unpredictable and can show huge variation from one reaction to another.

The plot in Figure 9 shows distribution of compounds on three parameters: molecular weight, clogP and cost of starting materials. It is evident from the plot that the synthesized compounds are scattered along the two property axes while remaining strongly skewed towards cheaper side, implying that good diversity was achieved at low cost.

5.3. Bioactivity studies and target-search

The NSD library was tested for various bioactivities on multiple platforms such as viability assays on cancer cell lines, anti-microbial assays, anti-oxidant activity assay, and detailed profiling on targets such as kinases, proteases and CEREP BioPrint® profile. Significant and promising activities were observed on cancer cell lines, on prolyl oligopeptide (POP) and on multiple targets in BioPrint® profile. No other significant activities were observed. The detailed observations of bioactivity assays are included in the appendix part.

5.3.1. Viability loss on cancer cell lines

At MarBio, NSD compounds were tested on various cancer cell lines such as A2058 melanoma, HT29 colon cancer, MCF7 breast cancer and MRC lung cells. The results are presented in Figure 10.
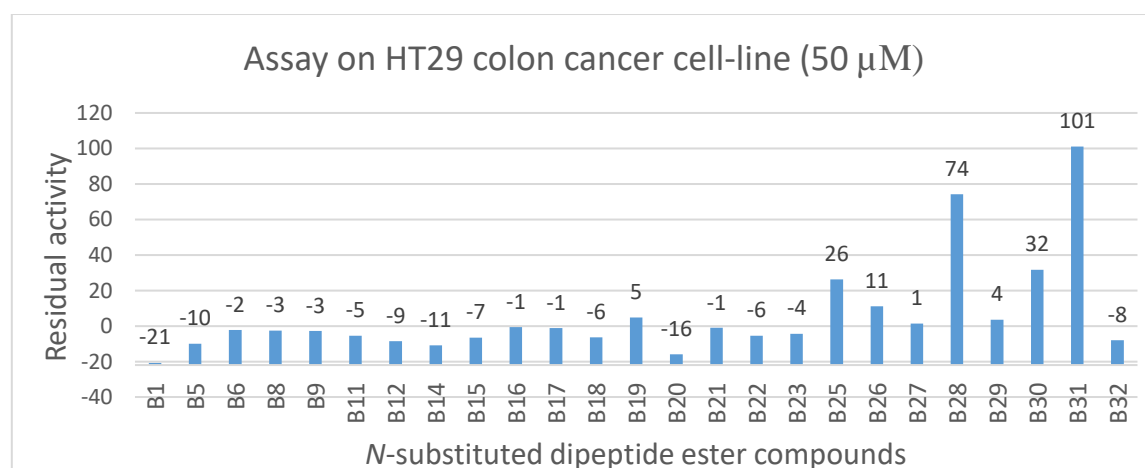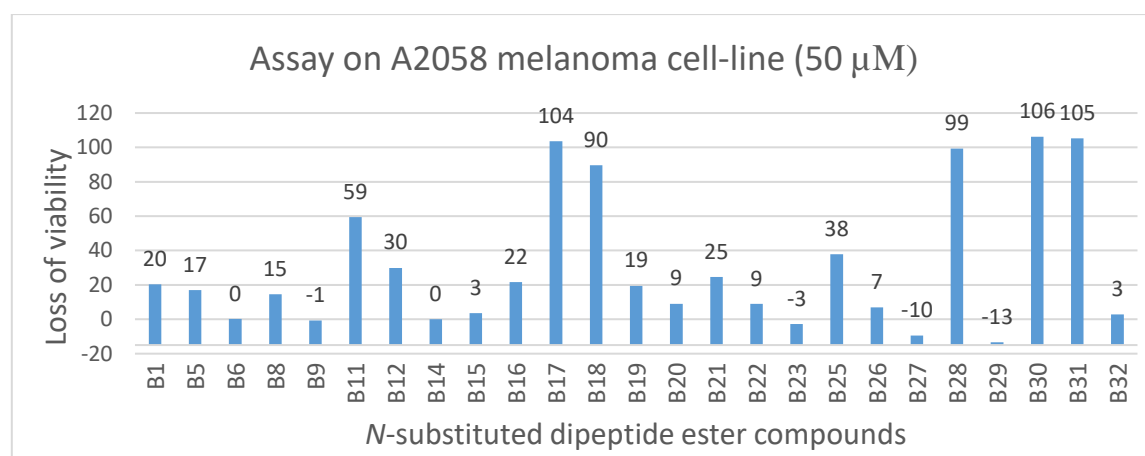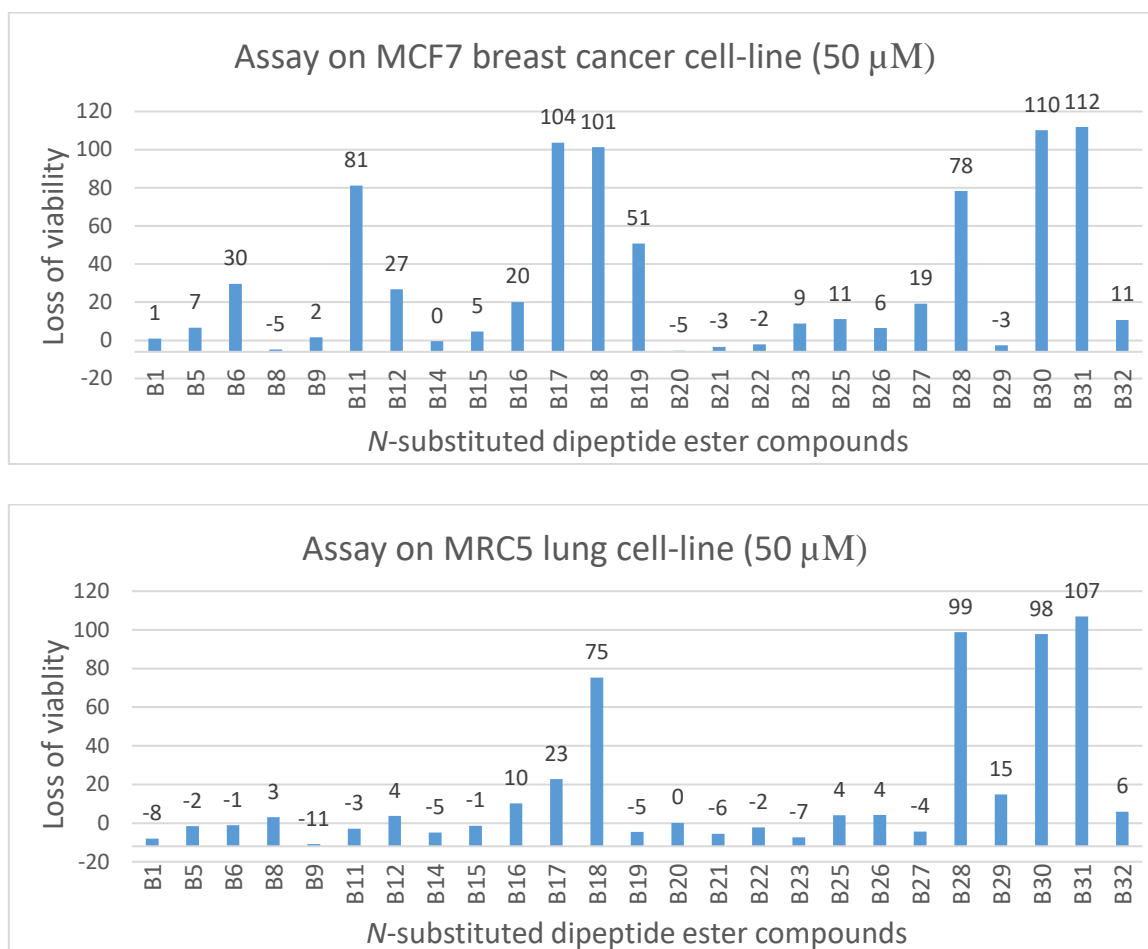
Figure 10. Cell viability assays on cancer cell lines for NSD series

Table 6. Viability loss of cell lines treated with seven active compounds (50 µM)

Compound	A2058	MCF7	HT29	MRC5
B31	105	112	101	107
B28	99	78	74	99
B30	106	110	32	98
B18	90	101	-6	75
B11	59	81	-5	-3
B17	104	104	-1	23
B19	19	51	5	-5

As shown in Table 6, seven NSD compounds (Figure 11) have been identified that cause significant loss of viability on at least one cancer cell line. Notably, compound B27 and compound B6 (stereoisomers of active compounds B18 and B31 respectively, both with *R*-configuration at C-terminal α-carbon i.e. attachment point of R^3) showed no activity. This

highlights the importance of chirality analogous to natural amino acid for the loss of viability (Figure 12).
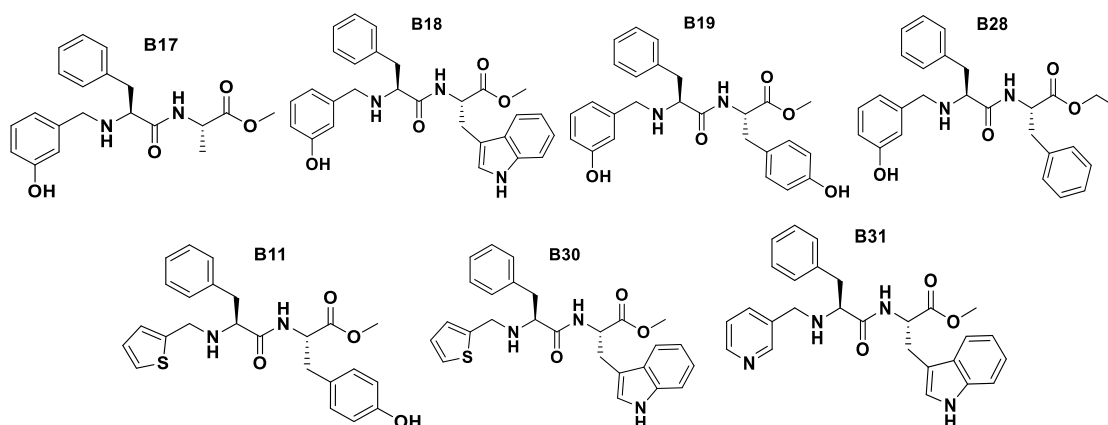
Figure 11. NSD compounds showing viability loss on cell lines

Upon morphological analysis of cells, compound B18 was observed to cause morphological change in the A2058 melanoma cells from concentrations as low as 5 µM in 4 hour exposure (Figure 13). The affected cells changed their shape from angular to round and did not remain immobilized at the bottom of the well. At higher concentrations such as 30 µM, the morphological change was immediate (within 5 minutes). Compound B27, an inactive stereoisomer of B18, did not cause any morphological change at 30 µM, confirming that the morphological change was associated with the loss of viability.
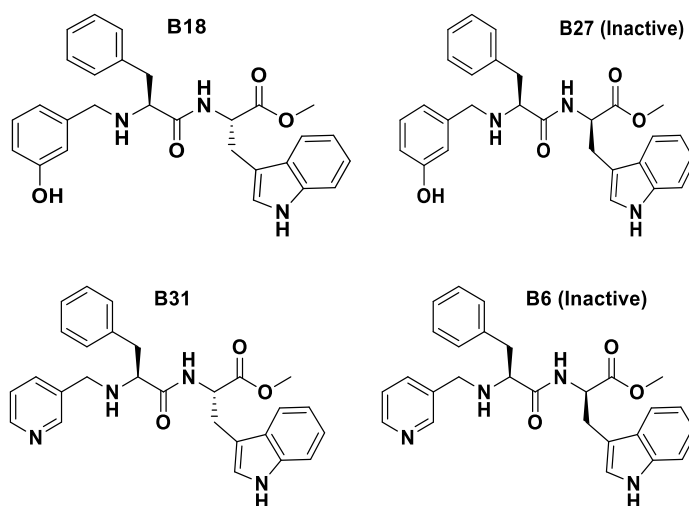
Figure 12. Compounds B18 and B31 (both *S,S*) causes viability loss, while their stereoisomers B27 and B6 (both *S,R*) are inactive, indicating that chirality analogous to natural amino acids is vital for loss of viability.
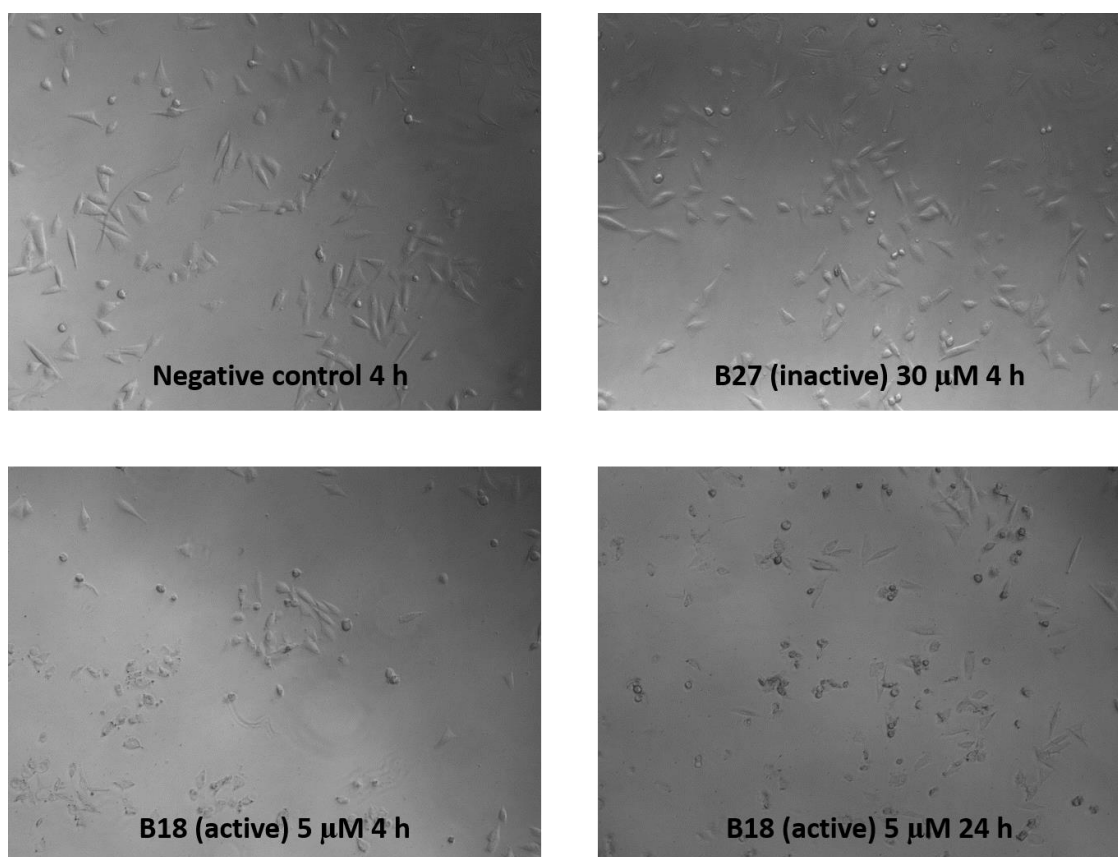
Figure 13. Effect of active compound B18 and its stereoisomer B27 on morphology of A2058 melanoma cells. B18 causes change in the shape of cells at concentrations as low as 5 µM in 4 hours, while B27 does not cause any viability loss.

In order to understand the underlying mechanism via which the *N*-substituted dipeptide esters cause loss of viability, proteomics analysis was has been carried out using SILAC method, which will be discussed in section 5.3.2.

Structure activity relationship (SAR) for loss of viability

Even though the drug-target was not identified, a rough SAR was generated for R^1, R^2 and R^3 substitutions and their effect on viability loss on cancer cell lines. These were based on available variations of substitutions of tested compounds and were as described below:

SAR for R^1: Aromatic carbocyclic or heterocyclic substitutions appear beneficial. The aromatic ring may contain hydrogen bond donor and acceptor groups.

SAR for R^2: A hydrophobic aromatic ring is beneficial. Absence of substitution or small aliphatic groups are counter-productive.

SAR for R³: Hydrophobic aromatic rings are beneficial. Polar hydrogen bond donor groups may decrease the activity. Absence of substitutions or substitutions with unnatural chirality result in complete loss of activity, which suggests that conformational state analogous to natural peptides is vital for the activity.

5.3.2. Quantitative proteomics analysis with SILAC studies

As described in section 5.3.1, seven compounds from NSD series showed significant loss of viability on cancer cell lines, starting from as low as 5 µM concentration in 4 hours. However, the drug target and underlying mechanism for the loss of viability is not known. A broad profiling of active compounds on kinases and proteases did not provide any conclusive target. Even though BioPrint® profiling did provide some interesting new targets, those targets do not appear relevant to the observed viability loss. Hence, a quantitative proteomics analysis was carried out using SILAC methods in order to understand change in protein expression driving the viability loss.

Principle: SILAC (Stable Isotope Labeling with Amino acid in Cell-culture) is a method for quantitative proteomics studies, which uses automated HPLC system coupled with tandem mass spectrometry to separate and identify specific cellular proteins.[87,88] Heavy amino acids (containing isotopes ^2H, ^{13}C, and/or ^{15}N) are present in media used to grow cells. As the cells grow, the heavy amino acid(s) are incorporated into its proteins. Thus, the cellular proteins from the cells grown in a heavy media have higher molecular weight than the proteins from the cells grown in a normal media, while the other properties, such as cellular functions or HPLC retention time, remain essentially the same. This difference in molecular weight separates the protein from heavy-media treated cells from the proteins from normal media treated cells when subjected to mass spectrometry after HPLC purification. The relative amount of proteins can thus be measured quantitatively to measure upregulation or downregulation of proteins..

Method: In a typical experiment, one group of cells (test or control) are grown in the heavy-media, while the other group (control or test) are grown in the normal media, . The test cells are subjected to the factors under study, while control cells are kept untreated. After treatment, the cells are lysed, and the protein mixtures are resolved on SDS-PAGE gel and digested to generate smaller "signature" peptides, which represent known peptide fragments for corresponding parent proteins. At this stage, peptides obtained from test and control cells are mixed, and the mixture is subjected to HPLC (or UPLC) purification as

per a pre-optimized automated protocol. The HPLC output is then subjected to tandem MS analysis, where signature peptide fragments are identified. For quantitative analysis, the log2 ratio between "Heavy" and "Light" labelled proteins in the sample are compared to identify significant upregulation and downregulation of proteins, which in turn would provide information regarding involved pathways.
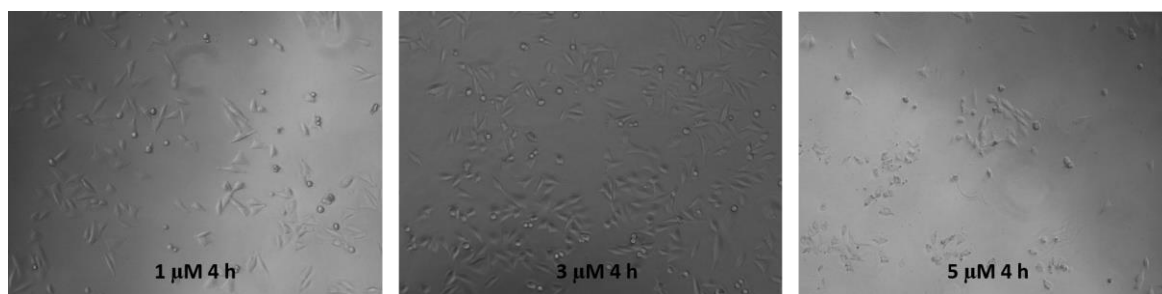
Figure 14. A2058 cells treated with B18 show morphological change at 5 µM concentration but not at lower concentrations such as 3 µM or 1 µM.

As the morphological change occurs as a result of numerous cellular events initiated by an affected cellular pathway caused by binding of compound to the specific drug target, it may be expected that the specific drug target and pathway are affected at much lower concentration than the concentration causing the morphological change. In our case, A2058 cells treated with B18 showed morphological change at a 5 µM concentration, while concentrations lower than 5 µM, such as 3 µM and 1 µM did not cause any observable change (Figure 14). Hence, for SILAC studies, A2058 melanoma cells were treated with B18 for 4 hours at 3 different concentrations: 0.5 µM, 2.5 µM and 5 µM. For each concentration, test and control samples were grown in heavy media and normal media respectively.

Results: The primary results of SILAC studies give an interesting insight into the cellular events driving the loss of viability. The histograms in Figure 15 shows that the overall log2 ratio (test/control) for the majority of proteins increases from less than zero to ~0.7 as per raw data with increasing concentration of B18. This suggests that increasing concentration of B18 results in stress conditions, which results in overall higher metabolic activity as cells react to the stress, e.g. prepare for apoptosis or other responses. Hence, the ratios were normalized for analysis of upregulation and downregulation of proteins, and the proteins showing significant change between test and control samples were analyzed for their involvement in cellular signaling pathways. The detailed data has been included in the appendix. Please see section 6.
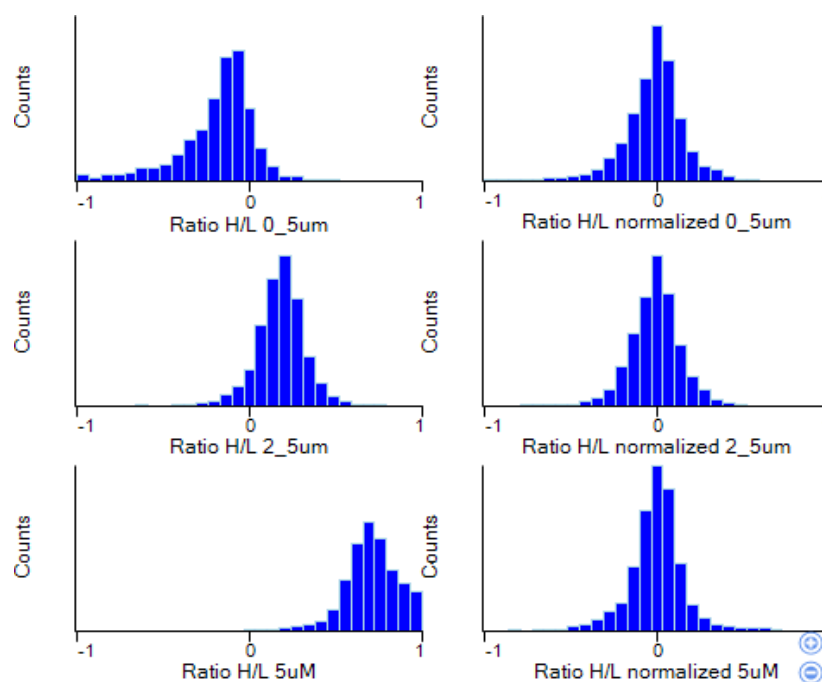
Figure 15. Histogram showing distribution of proteins in terms of amount ratio (test/control) for raw (left) and normalized (right) data. The ratio H/L are the log2 ratio between Heavy (test) and Light (control) labelled proteins.

It would be too early to conclusively establish the specific target or pathway from the primary results of SILAC studies. However, with the aim to understand the initial change in proteomics profile, the proteins showing significant upregulation (59 proteins) or downregulation (31 proteins) in both 0.5 µM and 2.5 µM samples were identified. The genes corresponding to such proteins were associated with specific cellular pathways with the help of Panther Database.[89,90] From this, the following four pathways were identified to be associated with one or more upregulated proteins as well as downregulated proteins:

1) FAS signaling pathway
2) Inflammation mediated by chemokine and cytokine signaling pathway
3) Huntington disease pathway
4) Cytoskeletal regulation by Rho GTPase

The Panther database was not able to associate some genes to any pathway, therefore, it cannot be ruled out that these four pathways may manifest an aftereffect of the activity of the compound on its target. Nevertheless, the identification of these pathways serves as an important milestone for further detailed and more focused studies to identify the target and mechanisms conclusively.

5.3.3. Protease profiling

The role of proteases has been well-established in cancer biology.[91,92] As the NSD compounds are peptidomimetic dipeptide esters, their potential to show activity on peptide-binding targets is evident. In order to explore this potential, compound B28-which showed significant activity on cancer cell lines-was subjected to protease profiling assays against 38 proteases at BPS bioscience, San Diego, USA. The compound B28 showed selective 80 % inhibition of prolyl oligopeptidase (POP) enzyme at 50 µM concentration. Its IC_{50} value was measured to be 6.0 µM (Figure 16). Following this, other related compounds, including B18, which also showed strong activity against cancer cell lines, were also subjected to assays on POP, but neither of them showed very significant inhibition at 100 µM (Table 7). This also means that the loss of viability observed on cancer cell lines is not due to POP inhibition. However, at the same time, the activity on POP also indicates that the NSD scaffold has potential to inhibit proteases, and an untested protease may potentially be the target.
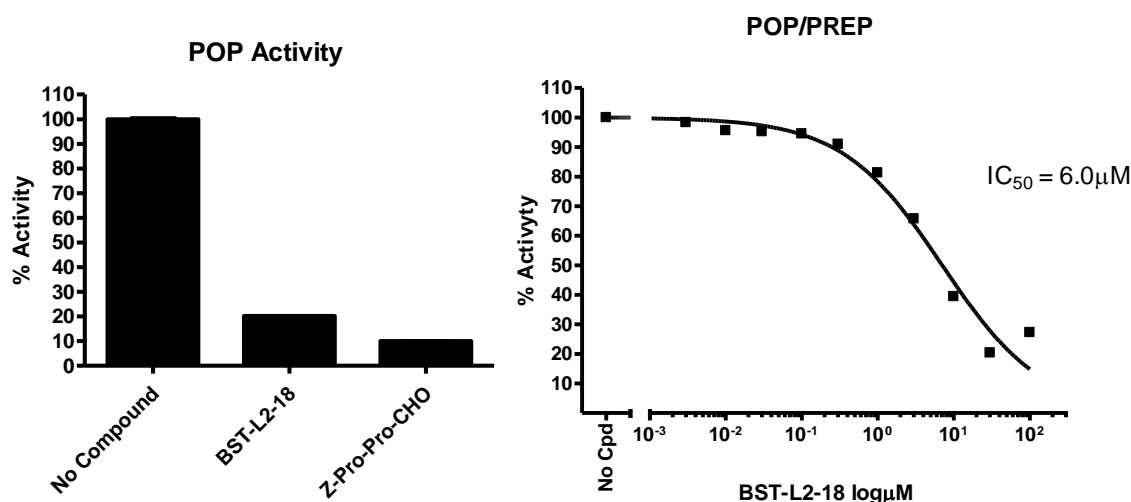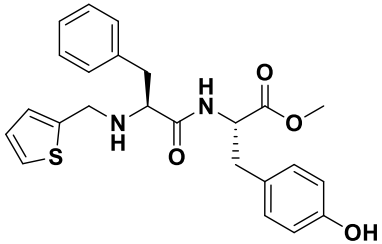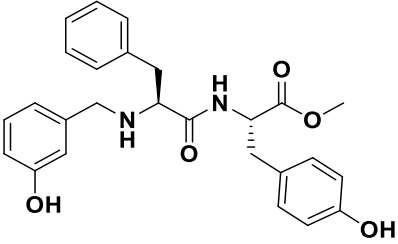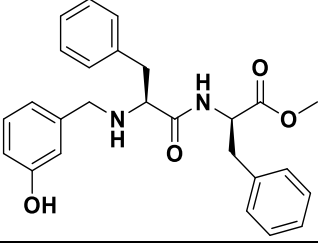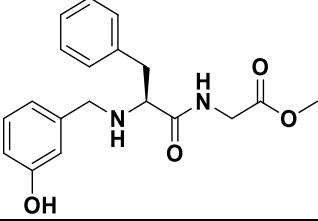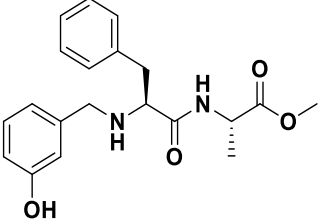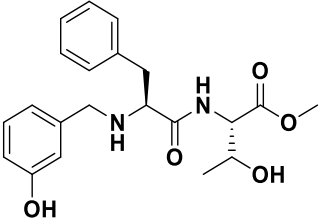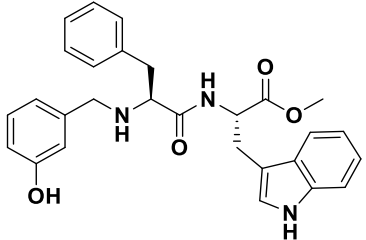
Figure 16. Inhibition of POP by compound B28 at 50 µM (left) and its IC_{50} studies (right).

Table 7. Assays on prolyl oligopeptidase (POP)

Compound	Structure	IC50, (POP inhibition)
B28		IC_{50} = 6 µM

B11		IC$_{50}$ > 100 μM, (23 % at 100 μM)
B19		IC$_{50}$ > 100 μM, (39 % at 100 μM)
B29		IC$_{50}$ > 100 μM, (9 % at 100 μM)
B16		IC$_{50}$ > 100 μM, (6 % at 100 μM)
B17		IC$_{50}$ > 100 μM, (10 % at 100 μM)
BST-L2-30		IC$_{50}$ > 100 μM, (27 % at 100 μM)
B18		IC$_{50}$ > 100 μM, (26 % at 100 μM)

5.3.4. Cheminformatics fingerprinting based target-prediction

Multiple methods[93–100] have been and continue to be developed for the *in silico* prediction of drug-target interactions using cheminformatics based fingerprinting analysis. Though the accuracy of such methods remains debatable, such methods help in narrowing down the list of probable targets. Among these models, we used a machine learning model[98] developed by Mervin et al. based on Bayesian methods using activity data (ChEMBL) and inactivity data (PubChem), and 2D fingerprinting of the compounds. Five representative compounds from *N*-substituted dipeptide esters library were subjected to the target prediction. The predictions were encoded as probability estimates that the compound binds to the target.

Table 8. Cheminformatics fingerprinting based target-prediction as probability estimates

Targets	Uni.No	B5	B11	B18	B19	B28
Cathepsin B	P07858	++	++	++	++	++
Cathepsin L1	P07711	++	++	++	++	++
Calpain-1 catalytic subunit	P07384	+	++	++	++	++
Neprilysin	P08473		++	++	++	++
Endothelin-converting enzyme 1	P42892	++	++	++	++	
Integrin alpha-4	P13612		++	++	++	++
Prolyl oligopeptidase (POP)	P48147					

Table 8 shows the target most likely to be inhibited, along with the predictions for prolyl oligopeptidase (POP). The method predicted (++ = near certainty, + = likely, blank = unlikely) binding of NSDs to cathepsin B and other related proteases. However, as described earlier, the protease profiling did not reveal any such significant activity on cathepsins or other related targets except µM level binding affinity of B28 to POP, which was predicted to be unlikely.

Overall, analogous to the results reported by Urich et al.[101] for protein kinases, this cheminformatics fingerprinting based prediction correlates with the measurements within the enzyme type, but not for the specific enzymes themselves. This may be attributed to the fact that cheminformatics based methods take into consideration mostly 2D structures, which neither includes information on 3D interactions with targets nor it identifies the most important structural features to give higher importance for probability calculation. These observations, however, may be useful for development of more a more inclusive and holistic cheminformatics fingerprinting based model.

5.3.5. Multiple receptor binding

A broader Bioprint® profiling of compound B18 at Eurofins CEREP revealed significant new bioactivities on multiple targets as shown in Table 9. Further detailed studies in a larger scale are underway to establish primary SAR.

Table 9. BioPrint® profile of B18 shows strong binding to multiple receptor and enzyme targets

Assay	% Inhibition of Control Specific Binding (50 µM)	Follow up study
NK2 (h)	102	Antagonist, IC_{50} = 3.1 µM, K_b = 400 nM
kappa (KOP)	98	Agonist, EC_{50} = 2.3 µM
COX2(h)	93	
COX1(h)	92	
motilin (h)	90	
delta (DOP) (h)	90	Antagonist, 37 % inhibition at 5 µM
mu (MOP) (h)	90	Weak agonist, EC_{50} = 47 µM
A3 (h)	84	
V1a (h)	84	
CCK1 (CCKA) (h)	83	
PPARγ (h)	80	
Ca^{2+} channel (L, diltiazem site) (benzothiazepines)	79	
MT1 (ML1A) (h)	77	
FP (h)	77	
NK1 (h)	76	

Considering that, most of these targets are peptide receptors; the activities are promising and optimizable. The strong activities on opioid receptors and cyclooxygenase are particularly interesting owing to their common therapeutic potential in treatment of pain. Moreover, strong activity on opioid receptors and neurokinin receptors also provides an opportunity in psychotherapeutics. In the follow-up studies, the compound B18 has been established to be agonist on κ (kappa) – opioid receptor with EC_{50} = 2.3 µM, a weak agonist on µ (mu) – opioid receptor with EC_{50} = 47 µM (20 times weaker agonist than on κ) and an antagonist on δ (delta) - opioid receptor with 37 % inhibition at 5 µM. This subtype selectivity is intriguing, especially from therapeutic point of view. The compound was also found to be antagonist on NK2 receptors with IC_{50} = 3.1 µM and K_b = 400 nM, further extending the potential of the series in psychotherapy.
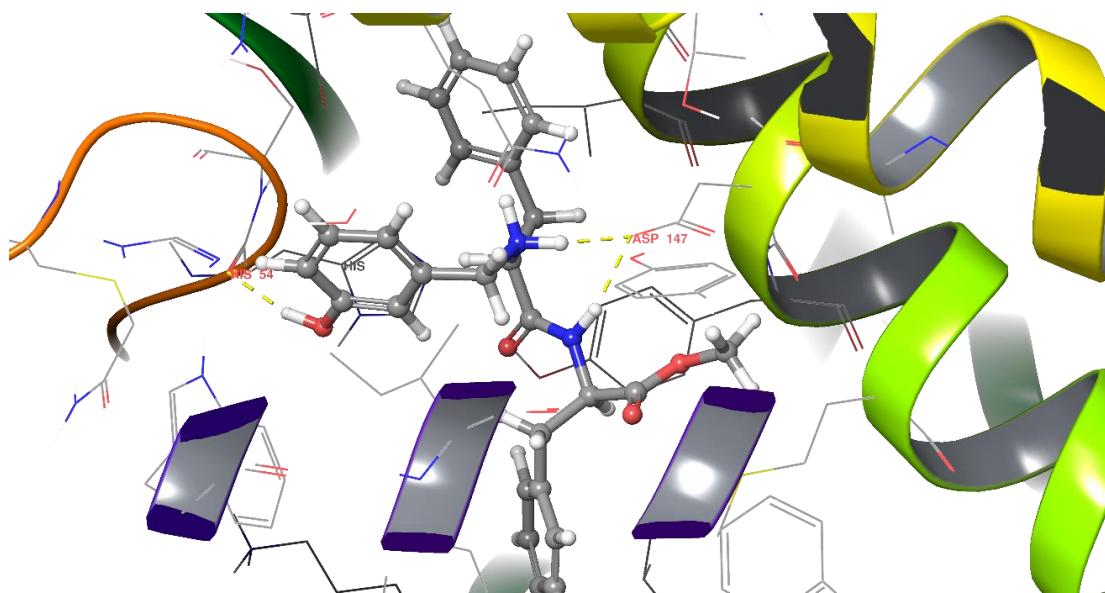
Figure 17. The docking of B29 inside Mu-opioid receptor (PDB: 5C1M) suggested that interactions with Asp147 and His54 may be important. Moreover, π-π stacking with His297 was also predicted.

As the *N*-substituted dipeptide scaffold provides six sites for stereospecific substitution, it should be possible to develop compounds with desired selectivity on a particular therapeutic class. For example, as shown in Figure 17 docking studies on opioid receptors suggest strong binding of B27 and B29 to opioid receptors (compounds that have been inactive on tested cancer cell lines). Overall, the existence of an expanded set of therapeutic targets, including anti-cancer activities, CNS disorders and pain relief, enables optimization separately for each therapeutic or target class.

6. Library II: Piperazine-2,5-diones (DKPs)

This chapter describes the synthesis of a piperazine-2,5-dione library, followed by a cheminformatics analysis of library properties with respect to the efficiency parameters described in the introduction part. Then, the results of bioactivity studies are given, especially assays on cancer cell lines and inhibition of BRSK1 enzyme has been discussed. Finally, a comparative docking study on Rho-kinase is detailed-as a test of the reliability of the method.

6.1. Synthesis

6.1.1. Synthetic protocol for DKPs

Though the synthesis of DKPs is possible to carry out in various ways[59,60,102], the synthesis via cyclization of dipeptide remains the most common protocol. Moreover, the synthesis via cyclization of dipeptide also enables the introduction of stereospecific substituents, a

benefit not easily available in other methods. For example, for the synthesis of a library similar to the one we considered, Qiao et al.[70] used an *N*-alkylation route using haloacetylhalide, which limited the possibility of diverse substitutions to only one carbon. To achieve substitutions on the other carbon, one would require a corresponding haloacid halide, which is not easily accessible. On the other hand, for the synthesis of cysteine based DKPs, Szardenings et al.[67,68] used dipeptide-route via simple solid phase synthesis to generate stereospecific *S*-cysteine and *R*-cysteine based similar DKPs with other substitutions on the other carbon and one nitrogen. However, this route also necessitated the use of protected amino acids along with usual protection-deprotection steps as in solid phase peptide synthesis.

In line with our efforts to use diversity oriented synthesis for chemoprospecting, the *N*-substituted dipeptide ester library worked as a precursor for the DKP library. Further, our approach enabled introduction of stereospecific substitutions at both α-carbons, and protection-deprotection steps. Thus, our approach achieved both advantages over previous protocols

6.1.2. Synthesis of library using microwave assisted cyclization

The most common protocol for synthesis of DKPs is the cyclization of dipeptide esters. However, the reaction requires the transformation of the orientation of the amide bond from trans to energetically less favourable cis orientation. Hence, a high activation energy barrier necessitates long times at high temperatures for completion.

The use of microwave radiation as a source of energy in organic synthesis is not new. The absorption of microwaves takes place in the polar covalent bonds of a molecule, resulting into rapid increase in rotational and translational energies of a molecule. This in turn increases the probability of successful collisions or folding of reactant molecules. Hence, microwave assisted reactions usually complete much faster than the reactions with only conventional heating do. As the cyclization of *N*-substituted dipeptide esters is also a reaction that requires high temperature and long time, we evaluated the use of a microwave-assisted protocol to achieve the cyclization in faster and efficient manner.

One of such microwave assisted protocols[52] for cyclization of boc-protected dipeptide esters appeared quite promising due to its simplicity and use of only water as a solvent without any other reagent. The protocol used boc-protected dipeptide esters. The Boc-deprotection occurs *in situ*, generating a 1°-amine group as nucleophile, which then reacts

with carbonyl carbon of the ester group, and DKP is formed with the alkoxy group of ester leaving as an alcohol. However, the protocol also had a few limitations/problems:

1) Most of diketopiperazines have no substitution on any nitrogen (denoted as R^3 in the scheme in the mentioned article). In case of substituted ones, the substitution is only methyl and not any greater than that.

2) Even though the article mentions, "water as solvent, 200 °C, 300 W, and 5 min reaction time resulted in the best", the table of synthesized compounds mentions the temperature to be 250 °C. (Table-2, step-b in the article).

3) The conditions mentioned were not possible to achieve. The publication mentions 150 psi pressure (~ 10 atm), 250 °C and 250 W. It was observed that 10 atm pressure was reached at 180 °C and the remaining reaction stayed at 172-175 °C. In subsequent reactions, the pressure limit was increased up to maximum 20 atm to achieve 250 °C, however, this time pressure limit reached 20 atm just at 225 °C and the remaining reaction took place at 220-225 °C.
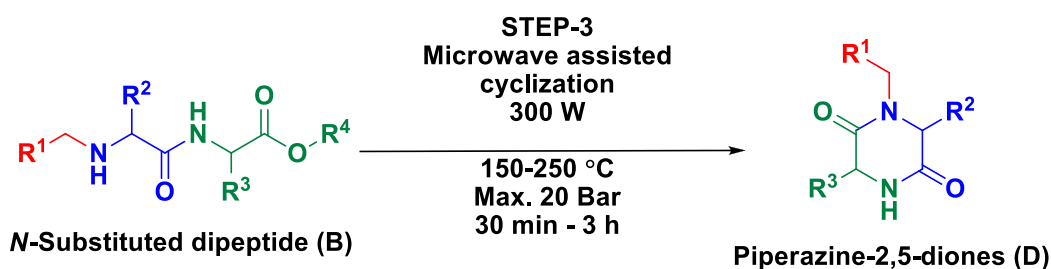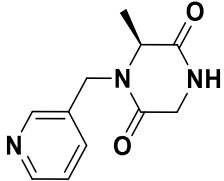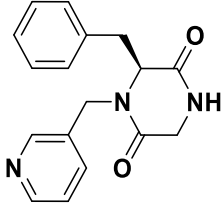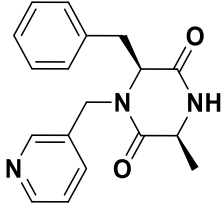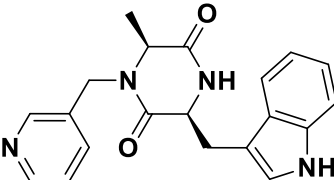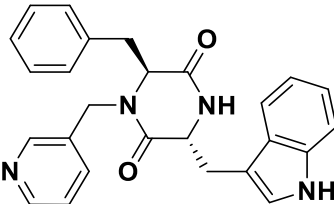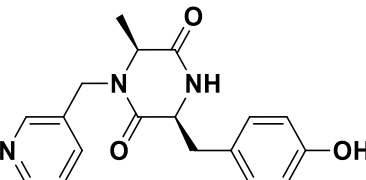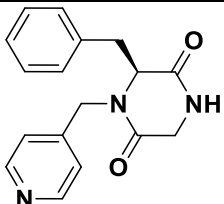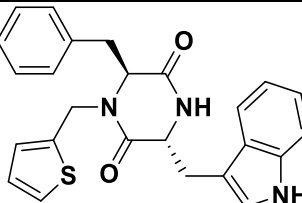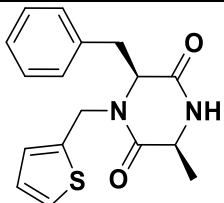
Scheme 6. Microwave assisted cyclization of *N*-substituted dipeptide esters to DKPs
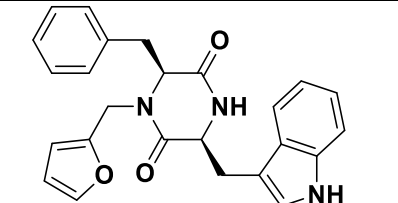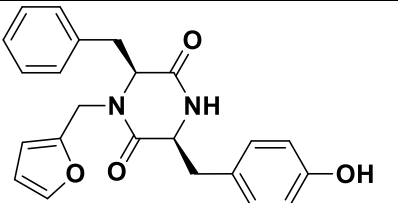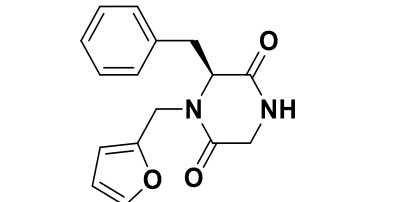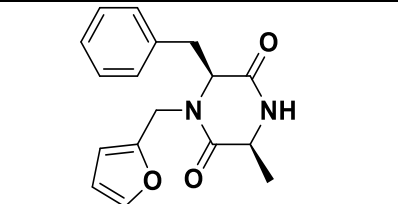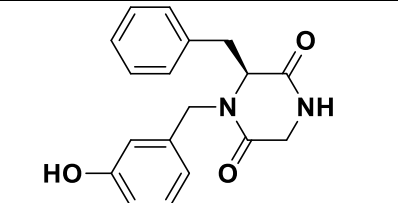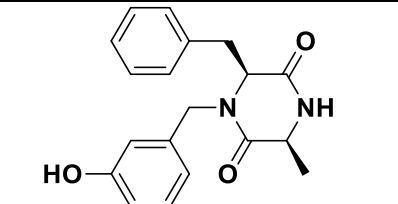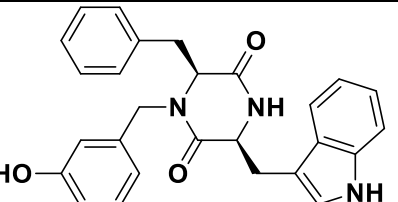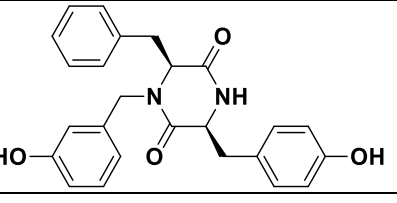
We sought to correct and customize the protocol (Scheme 6) to synthesize diverse DKPs containing larger substitutions with 2°-amine as the nucleophile (Table 10). Moreover, most of our compounds were more or less soluble in water, hence the aqueous work-up procedure could not be used.
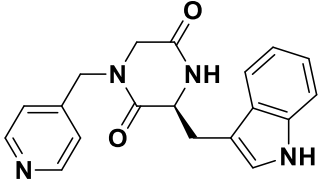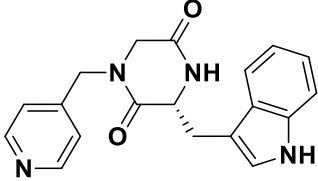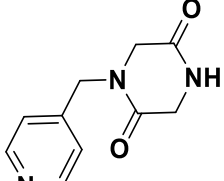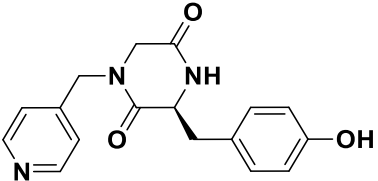
Table 10. Microwave assisted cyclization of *N*-substituted dipeptide esters to DKPs

Step-2 product	Time	Product	Product structure	Isolated Yield %
B1	20 min	D1		78

B2	30 min	D2		76
B3	30 min	D3		60
B4	30 min + 30 min	D4		56
B5	30 min + 30 min + 2 h	D5		47
B6	30 min + 1 h + 30 min	D6		22[a]
B7	30 min + 1 h + 2 h	D7		51
B8	30 min + 30 min	D8		55
B9	1 h + 2 h + 2 h	D9		16[a]
B10	1 h + 2 h	D10		64

B12	1 h + 2 h	D11		22
B13	1 h + 2 h	D12		37[b]
B14	30 min	D13		49[b]
B15	30 min + 30 min	D14		46[b]
B16	30 min + 30 min	D15		59
B18	30 min + 1 h	D16		58
B18	1 h + 1 h	D17		55
B19	1 h + 1 h	D18		51

B21	30 min	D19		61
B22	30 min	D20		65
B23	30 min	D21		72
B24	30 min	D22		59

[a]Degradation was observed.
[b]Very low UV activity may have resulted in low isolated yield.

Overall, the yields were moderate. With the introduction of more and bigger substitutions, the decrease in the yield and increase in the necessary time for reaction completion was notable. A strong steric effect of substitutions in cyclization reaction was thus evident. A detailed study of such effects can be useful to determine synthesizability, and thereby in library design. In order to get a clear idea regarding such effects, theoretical studies were carried out for both stages of the reaction, namely cis/trans isomerization and cyclization led by nucleophilic attack of amino nitrogen on carbonyl carbon of ester group. The theoretical studies have been described in a separate section "Theoretical Studies", from page 77.

6.2. Library properties

6.2.1. Diversifiability

As DKPs are cyclized products of *N*-substituted dipeptide esters, the DKP series has a diversification potential similar to that for NSDs. Both series have the same starting materials, and hence the theoretical chemical space for DKPs and NSDs have the same size. Considering this similarity, the virtual library of total 30118 DKPs with same R^1, R^2 and R^3 as in NSDs was created and all compounds were analyzed for different physicochemical parameters.

6.2.2. Diversity

Like the NSD library described earlier, the DKP virtual library was analyzed with respect to various physicochemical parameters such as molecular weight, AlogP, number of rotatable bonds etc. It can be observed from Figure 18 that the DKP virtual library shows good distribution patterns for different parameters.
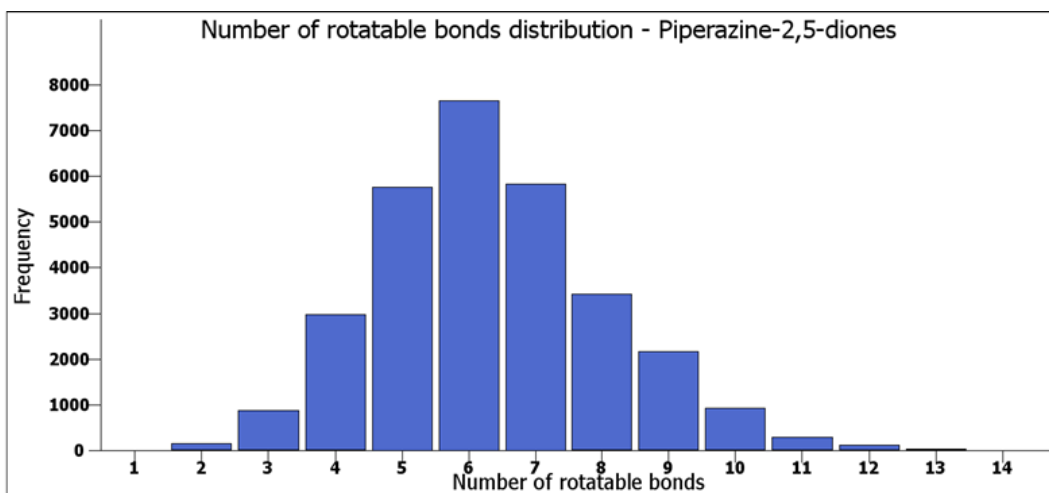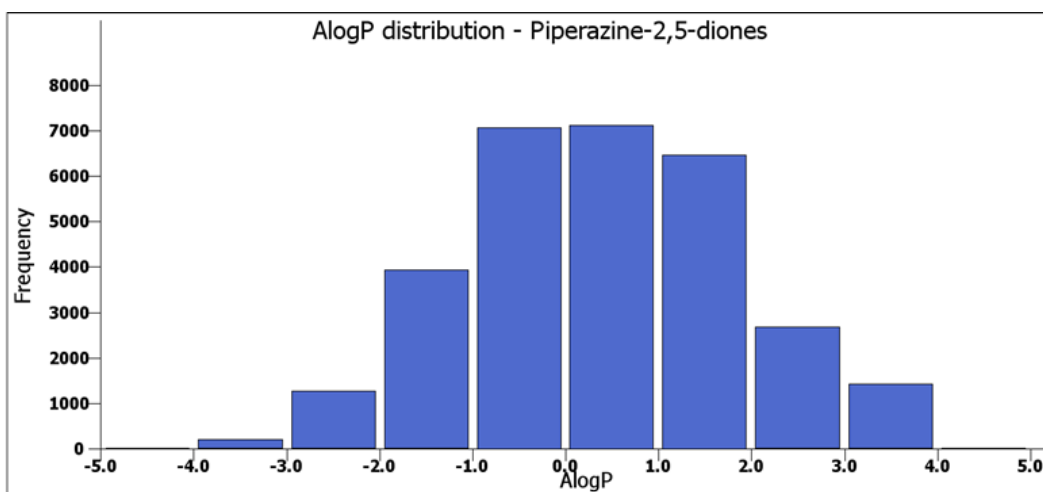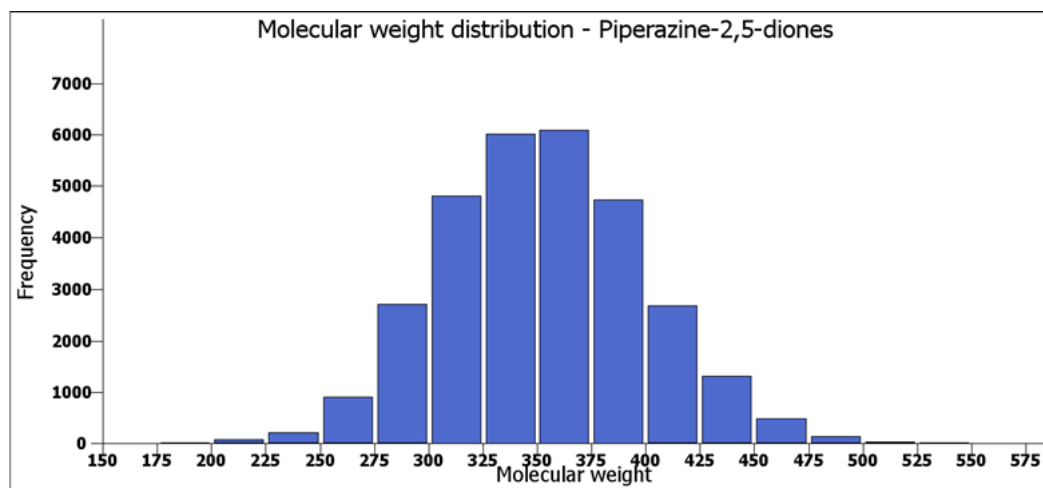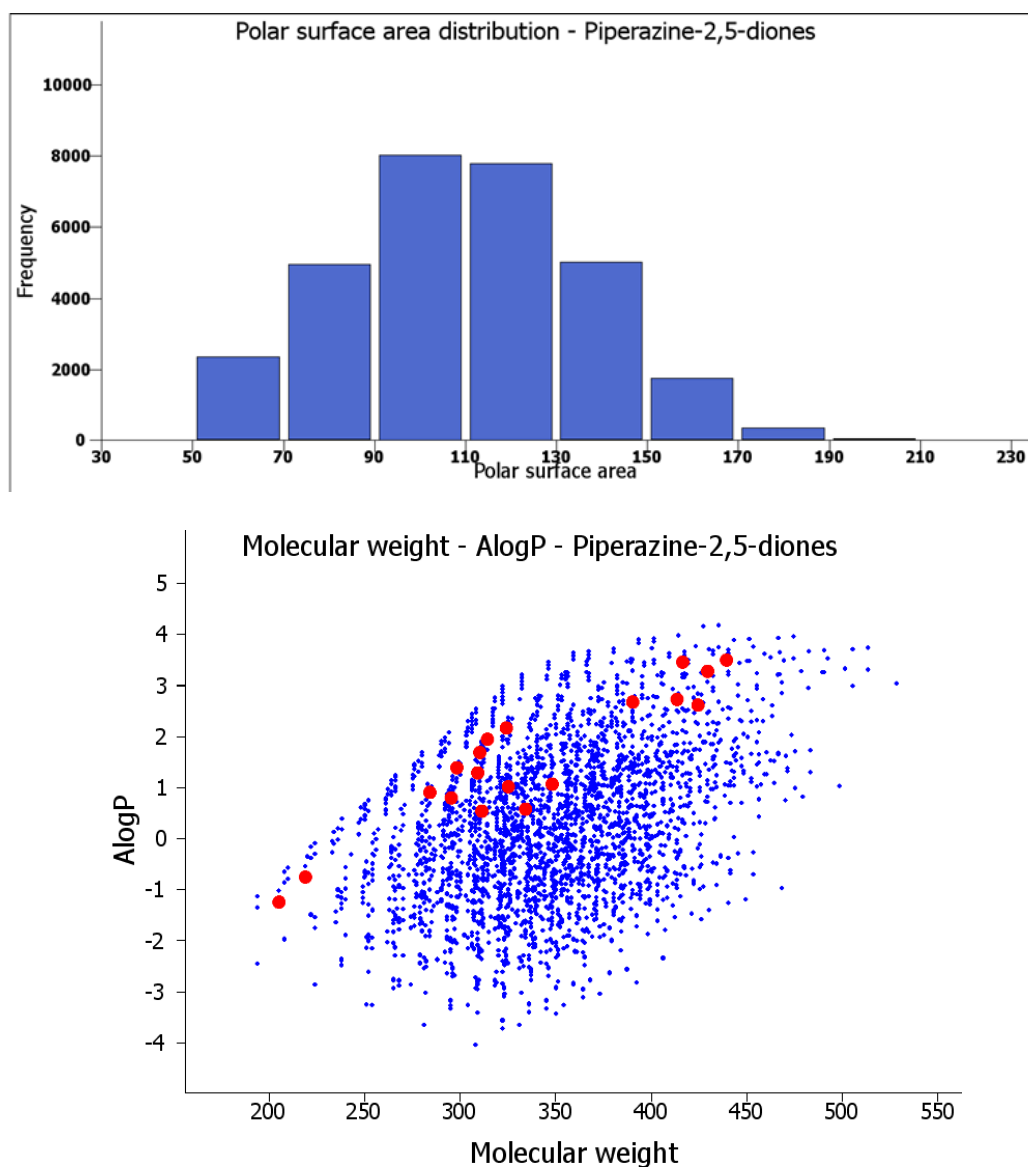
Figure 18. Distribution of a few properties for the virtual library of piperazine-2,5-diones. Red dots denote the synthesized compounds.

6.2.3. Ease of synthesis

The DKP library can be synthesized with a three-step synthesis method from commercially available starting materials, including aldehyde, amino acid and amino acid esters. As described earlier, the first step does not require purification, while the second step and third step require column chromatography. All three steps are quite safe and no harmful reagent is required. Hence, overall the ease of synthesis is quite high.

6.2.4. Cost and availability of starting materials

As this library represents the cyclized product of *N*-substituted dipeptide library, the cost of the starting materials for DKP library remains the same as that for the *N*-substituted dipeptide library.

6.3. Bioactivity studies

The piperazine-2,5-diones library was tested for various bioactivities on multiple platforms such as viability assays on cancer cell lines, anti-microbial assays, anti-oxidant activity assay, and detailed profiling on kinases. One compound showed significant inhibition of a cancer cell line, while another compound showed significant inhibition of a kinase. No other significant activities were observed. The detailed observations of bioactivity assays are included in the appendix.

6.3.1. Inhibition of cancer cell lines

D14 was found to cause loss of viability on A2058 melanoma cell line with 74 % viability loss at 50 µM and $IC_{50} = 43$ µM (Figure 19). However, no other compound showed such significant activities on A2058 or other cell lines.

D14 - 74 % viability loss of A2058 melanoma cell line at 50 µM

D8 - 92 % inhibition of BRSK1 at 100 µM
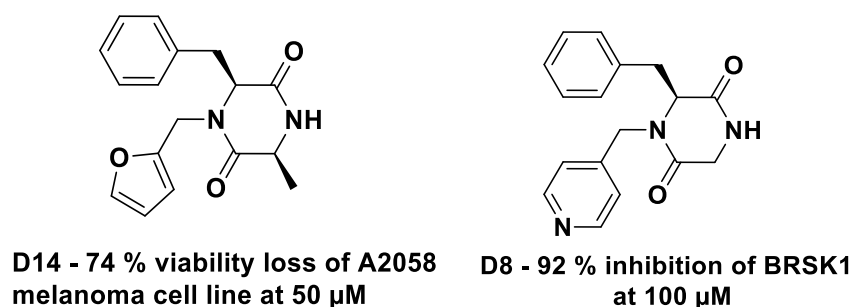
Figure 19. Active compounds from piperazine-2,5-diones series

6.3.2. BRSK1 inhibition

The kinase profiling of a DKP compound **D8** showed 92 % inhibition of BRSK1 at 100 µM. This was not reproduced to that extent in follow-up testing, which showed instead 46 % inhibition and thus an IC_{50} near 100 µM. This revealed BRSK1 (also known as SAD1 or SAD-b) as a possible target that can be explored by the DKP series and its further development. As **D8** is a small sized molecule with molecular weight less than 300, it can be reasonably considered a fragment type hit for further BRSK1 oriented drug-development (Figure 19).

BRSK1 (Brain Specific Kinase 1) is an isoform of brain specific kinases (also known as SAD kinases). It has been shown[103] to be important for neuronal polarization. BRSK1 is activated[104] by LKB1 (Liver Kinase B1, also known as STK11) by phosphorylation in cerebral cortex. Once activated, BRSK1 then phosphorylates[105] microtubule associated proteins, which finally elicit polarization. BRSK1 has also been associated with presynaptic vesicle clustering and axon termination[106] as well as with regulation[107] of neutrotransmitter release. Recently, the relationship between tau-dependent neurogeneration by BRSK1[108] has also been established.

Apart from its role in polarization of neurons and presynaptic regulations, BRSK1 has been recognized[109–111] as an important regulator of centrosome replication, which also explains its role in the check-point response to the DNA damage[112] induced by UV or chemical mutagenesis. Such an important role in cell-cycle makes BRSK1 as an interesting drug target for cancer drug development.

With such an important role, from a pharmaceutical point of view, BRSK1 is a relatively new yet interesting target. BRSK1, as modifier of the PTEN/AKT pathway and the methods of use were patented[113] in 2006, which expires by 2026. Considering 10-15 years as a typical drug-development time-line, 9 years from now is also a convenient window for BRSK1 targeted drug development

6.4. Docking studies of synthesized compounds to Rho kinases

6.4.1. Rho kinases and their biological importance

Rho-associated protein kinase (ROCK) belongs to the AGC kinase family of serine-threonine kinases, expressed throughout human body as two subtypes, namely ROCK1 and ROCK2. More specifically, ROCK1 is expressed more in liver, kidney, testis and spleen, while ROCK2 is expressed more in brain, skeletal muscles and other tissues.[114,115]

Being an important cell-regulator and wide distribution throughout the human body, ROCK has been shown to contribute to various pathophysiological conditions. The biological roles of ROCK[116] in various body-functioning such as in the functioning of skeletal and smooth muscles[117–120], determination of cell-permeability[121,122] insulin signaling[123], neurological disorders[124] has been well documented. A very important role of ROCK has been established in the cardiovascular system[125–128] where it is involved in various cellular processes such as nitric oxide synthesis, atherosclerotic lesion, angiotension generated

hypertrophy, cardiovascular inflammation, hypertension, stroke etc. Similarly, ROCK is also associated with cancers[129,130] contributing to cell-permeability and invasion.

With such a diverse activity profile, ROCK has attracted much attention in recent years as an important drug target. Many ROCK specific inhibitors[131–135] belonging to diverse chemical classes have also been reported. Among them, Fasudil has been approved for treatment of cerebral vasospasm.

6.4.2. Similarity of synthesized DKP derivatives with patented Rho inhibitor piperazines

ROCK shares significant similarity[136] with other AGC kinases such as protein kinase A and specific inhibitors of Rho kinase have also been studied[137] for their activity on PKA using Rho kinase mutant model in PKA as surrogate models.

The literature survey revealed that *N*-substituted piperazine derivatives[135] have been developed and patented as specific Rho kinase inhibitors.

Patented Piperazine derivatives   **Synthesized DKP derivatives**
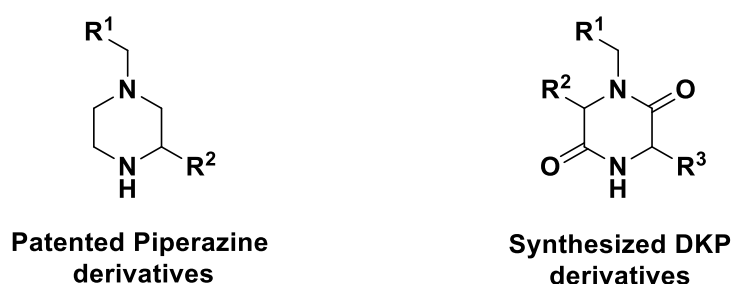
Figure 20. Similarities between patented piperazine derivatives and synthesized DKP derivatives

Because of the structural similarity between patented piperazine derivatives and synthesized DKP derivatives (Figure 20), we considered the possibility of Rho kinase inhibition by synthesized library. Accordingly, we decided to carry out docking studies as a primary indicative tool for possible inhibition. However, in order to determine the predictivity of docking results, we decided to carry out an initial docking study of known Rho kinase inhibitors and PKA inhibitors on different structures of Rho kinase, PKA, and their hybrid structures.

6.4.3. Initial docking of known inhibitors on Rho, PKA and hybrid structures

Following protein structures from the PDB were included for the docking studies:

- ROCK1 structures: 2ESM, 2F2U, 3NCZ, 3V8S, 4W7P
- 5 fold mutant of Rho kinase on PKA (PKA5R) structures: 2GNH, 2GNI, 2GNF

- 3 fold mutant of Rho kinase on PKA (PKA3R) structures: 2GNJ, 2GNL
- PKA structures: 1STC, 1SVG, 1Q8W

Along with above structures, two homology models were also created using Schrodinger software suite and were included in docking.

- Homology models: PKA (1Q8W) based on ROCK1 (3NCZ), ROCK1 (3NCZ) based on PKA (1Q8W)

As known inhibitors, following ligands were extracted from one of their native protein-ligand complexes and were subjected to LigPrep application prior to docking.

- Fasudil, Hydroxyfasudil, Y27632, H1152P, 3NC from PDB ID: 3NCZ, 3ND from PDB ID: 3NDM, Indazole compound 18 from PDB ID: 3V8S, RKI1342 from PDB ID: 3TV7, YB-15-QD37 from PDB ID: 4W7P, Staurosporine

The docking results were studied to compare the predicted docking pose vis-à-vis the crystal structure pose of the compound in the same enzyme, considering the crystal structure pose as a most energetically favourable pose. The results are summarized in Figure 21.

- Although only a small fraction of docking experiments reproduced the experimental binding pose in its entirety (dark green), somewhat over half reproduced at least the hinge binding interactions (dark and light green), and a clear majority predicted at least some hinge binding (yellow and greens).
- The failure of staurosporine to dock successfully in any target except its parent structure (and here only with XP precision docking) is notable. Because staurosporine, with its extended planar and aromatic structure, significantly expands the binding site,[138] docking requires adequate prediction of flexibility prediction for the target, as has been noted previously.[139]
- The target 1SVG also usually failed to predict hinge binding. Here, due to definition of the pocket grid based on the extent of the native ligand, the more extended inhibitor of 1SVG led to a pocket that included more residues distant from the hinge. Several of these created a hydrophobic/aromatic site that the docking algorithm ranked higher than hinge binding interactions.

Figure 21. Docking of known inhibitors on Rho kinase, PKA and hybrid structures. Each target structure (in rows) was used with (top half of row) and without (bottom half of row) minimization. Inhibitors with multiple possible charge or protonation states have multiple columns.

- Although the minimization of the target molecule often significantly changed the outcome, there is no overall correlation with respect to the quality of the prediction.

- The proton assignment could be decisive: Unfavorable assignment of the protonation state of indazole-18 search prevented correct docking.

The homology models performed diversely: the ROCK1 model (based on 1Q8W was as good as the best ROCK1 structures, while the PKA homology model was quite poor.

6.4.4. Docking of synthesized compounds with PKA, Rho and hybrid structures

From the above study, it was apparent that protein structures 1SVG and 1STC structures, as parameterized, were unreliable for docking. Hence, docking of synthesized compounds was carried out on all structures except these two. A standard protocol to use of energy minimized protein structures was followed. Moreover, all native ligands extracted from protein structures were also included for the docking studies in order to make a possible comparison with docking scores of synthesized compounds.

The docking results showed that ligands 3ND and 3NC (from native ligands) were quite good binders to most structures. Known inhibitors of Rho kinase also docked well. In comparison, the synthesized compounds showed relatively poor docking score. Only a few compounds such as D18 showed docking score within around 1-2 units range of the scores for 3NC and 3ND, while for others, the docking scores were poorer.

Overall, the docking results were not very positive for the synthesized compounds with respect to binding with Rho kinase.

7. Library III: Tartaric acid bisamides (TABs)

This chapter first describes different methods used for synthesis of tartaric acid bisamides. This synthesis also led to a faster protocol for cyclic tartrimide derivatives. This is followed by cheminformatics analysis of library properties as per the efficiency parameters described in the introduction part. A brief account on bioactivity studies is also included.

7.1. Synthesis

7.1.1. Designing the synthetic route to tartaric acid bisamides (TABs)

As the target structure was a bisamide with two different amides at both ends, reaction of both carboxylic acid centers with two different amines was necessary. However, it was not

possible to react both carboxylic acid centers with two different amines simultaneously. Hence, it was necessary to use a multi-step approach.
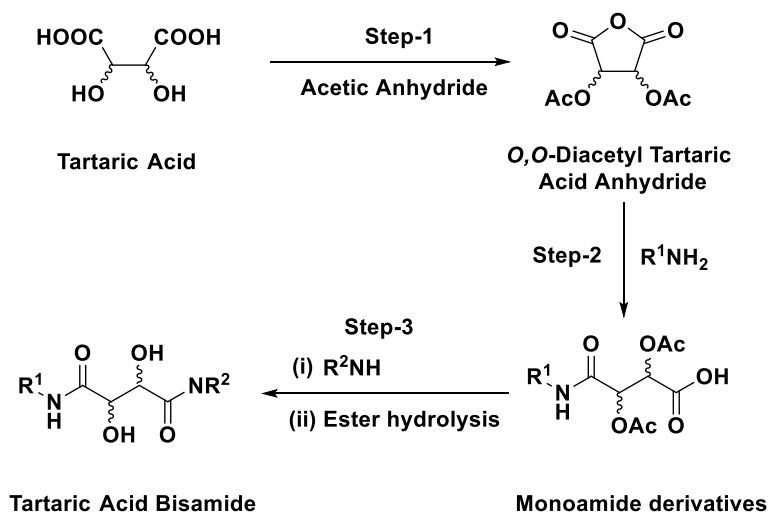
Scheme 7. Synthetic route to tartaric acid bisamides

The general approach for synthesis was a three step scheme (Scheme 7), where the first step involved synthesis of *O,O*-diacetyl tartaric acid anhydride. The reaction of the anhydride with an amine would give *O,O*-diacetyl monoamide derivatives, which upon reaction with another amine followed by alkaline hydrolysis would afford the desired tartaric acid bisamides.

7.1.2. Synthesis of *O,O*-diacetyl tartaric acid anhydride (E)
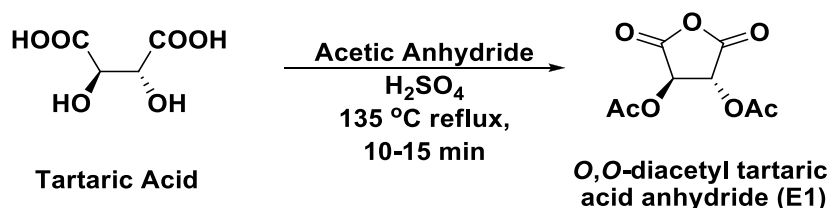
Scheme 8. Synthesis of *O,O*-diacetyl tartaric acid anhydride

Following a reported protocol[140,141], *O,O*-diacetyl tartaric acid anhydride was synthesized by refluxing tartaric acid with acetic anhydride and catalytic amount of sulfuric acid for 10-15 min in 69 % yield (Scheme 8). The pure product was isolated on gram scale after crystallization followed by washing with benzene and ether.

7.1.3. Synthesis of *O,O*-diacetyl tartaric acid monoamide (F)
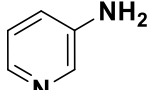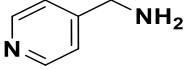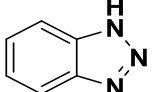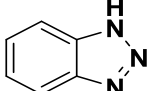
The anhydride was reacted with amines to afford *O,O*-diacetyl tartaric acid monoamide. The use of 2 equivalents of amines[141,142] resulted in salt formation due to reaction of excess

65

amine with the free carboxylic acid group of monoamide product. Therefore, only slight excess of anhydride (1.05 equivalent) was used (Scheme 9). Initially, the 1°-amines were used to generate corresponding 2°-amides. However, later on when the follow-up steps could not succeed with 2°-amides, 3°-amides were synthesized via reactions of 2°-amines with the anhydride (Table 11).
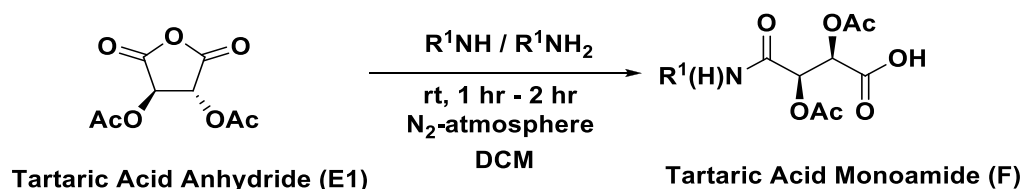
Scheme 9. Synthesis of tartaric acid monoamides

Table 11. Synthesis of *O,O*-diacetyl tartaric acid monoamides

Product	R^1NH / R^1NH_2	Reaction conditions	Yield
F1		rt, 2 hr	88 %
F2		rt, 1 hr	92 %
F3		rt, 1 hr	98 %
F4		rt, 1 hr	95 % (unstable product)
F5		rt, 1 hr	98 %
F6		1 day	No reaction
F6		60 °C, 6 hr	No reaction

7.1.4. Attempts of peptide coupling

In order to synthesize tartaric acid bisamides from the monoamides, first peptide coupling was tried, as shown in Scheme 10. The same protocol, which was used for *N*-substituted dipeptide library, was initially used here. However, the reactions were not successful as the crude reaction mixture showed degradation and no discernible product formation. The result remained the same with other coupling reagents like EDC.
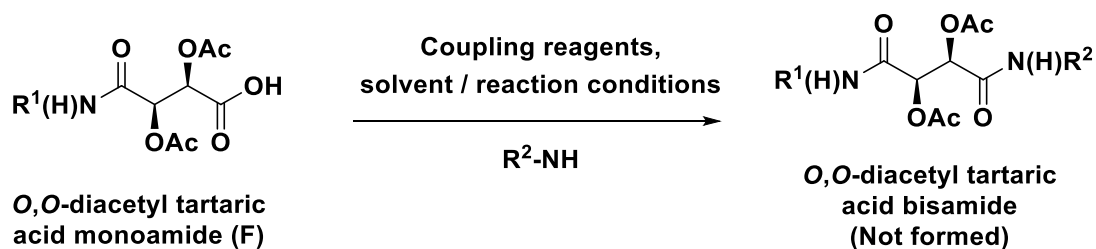
Scheme 10. Attempts of peptide coupling with monoamide

It appeared that the monoamide might be sensitive to the peptide coupling protocols as the reaction with tryptophan methyl ester (often used in the *N*-substituted dipeptide series) was also unsuccessful.

Table 12. Peptide coupling attempts with *O*,*O*-diacetyl tartaric acid monoamides

No.	R^1-N(H)	R^2-NH	Reaction conditions
1			HBTU (1.2 equiv), rt, TEA (4.8 equiv), ACN, 30 min
2			DCC (2.1 equiv), 80 °C, ACN, 3 h
3			DCC (2.1 equiv), rt, ACN, 14 hr
4			EDC (2.0 equiv), rt, ACN, 14 hr
5			HATU (1.2 equiv), rt, TEA (4.8 equiv), ACN, 1 h
6			HATU (1.2 equiv), 60 °C, TEA (4.8 equiv), ACN, 1 h
7			HBTU (1.2 equiv), rt, TEA (4.8 equiv), ACN, 30 min
8			HBTU (1.2 equiv), rt, TEA (4.8 equiv), ACN, 30 min
9			DCC (2.0 equiv), rt, ACN, 14 hr
10			HBTU (1.2 equiv), rt, DIPEA (4.8 equiv), ACN, 30 min

Initially, the peptide coupling reactions were carried out with only monoamides generated from 1°-amines. After the efforts with 2°-amides proved to be unsuccessful, monoamides from 2°-amines like morpholine were synthesized. However, peptide-coupling reactions on these were also unsuccessful (Table 12).

7.1.5. Attempts of peptide coupling after deprotection

The failure of peptide coupling reactions was unexpected. However, we found an example in literature where the tartaric acid monoamides had been reported[143] to undergo ketene formation in presence of thionyl chloride. We hypothesized that activation of the carboxylic acid group with coupling reagents might also lead to ketene formation, which might be a reason for the failure of the protocol because a ketene intermediate could lead to different reaction pathways. If that were the case, the deacetylated monoamide might show some different result. Hence, a hydrolysis reaction was performed as shown in Scheme 11.

O,O-diacetyl tartaric acid monoamide → NH₄OH (aq), RT, 2 hours → **Ammonium salt of tartaric acid monoamide (Impure)**
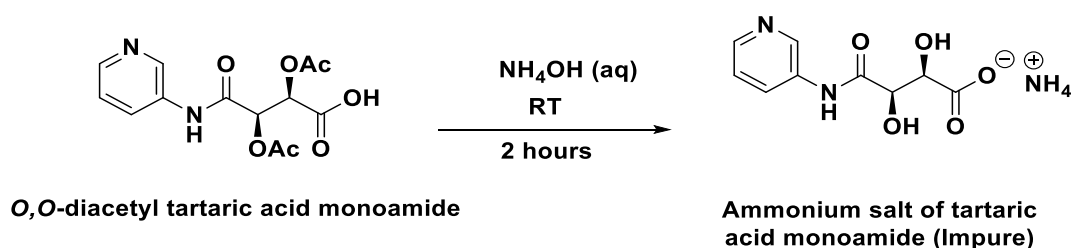
Scheme 11. Hydrolysis of diacetyl tartaric acid monoamide

The O,O-diacetyl monoamide was hydrolyzed using 28 % aq. ammonium hydroxide[144] solution. The hydrolysis reaction was not clean. The product was highly polar which did not elute out of silica or alumina column. Thus, it was an unsuccessful attempt of isolating deacetylated monoamide product.

7.1.6. Attempts of acid-chloride synthesis from 2°-amide – tartrimide formation
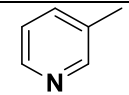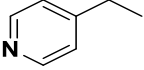
As the peptide coupling was unsuccessful, and the deacetylation product could not be isolated, formation of the acid chloride intermediate was chosen as the strategy for the required amide bond formation. Hence, the reaction of monoamide was carried out with thionyl chloride at reflux[145,146], followed by addition of a secondary amine, morpholine (Scheme 12).
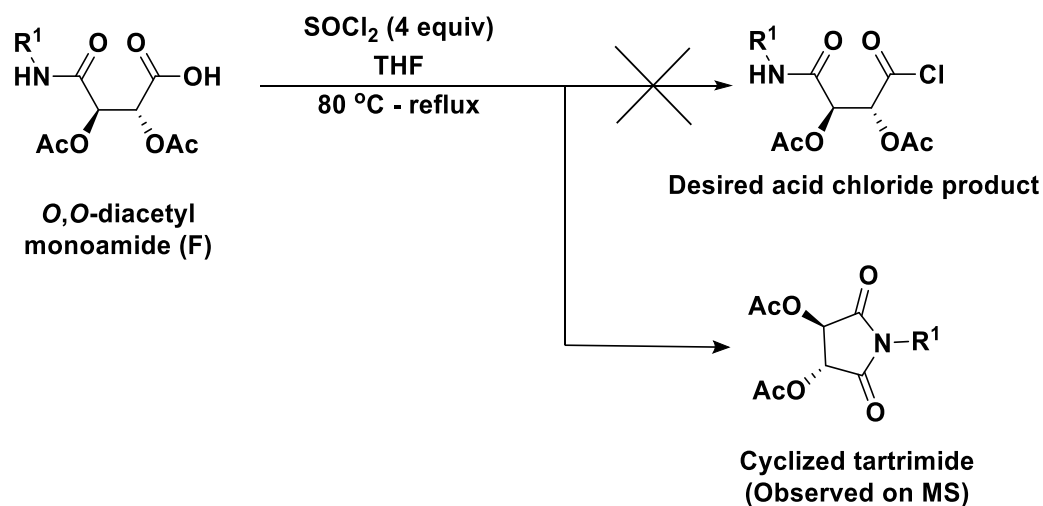
Scheme 12. Scheme for acid chloride synthesis, which resulted in tartrimide formation

However, when the reaction was carried out, the mass spectrum of the crude reaction did not show any peak for desired product (Table 13). Instead, a single peak for dehydrated product was observed, which implied that instead of acid chloride formation, the monoamide might have cyclized in the presence of thionyl chloride via ketene intermediate.[143] Later, this was confirmed in a different protocol using thionyl chloride as solvent, and a new series of tartrimide was developed (Please see section 7.1.8)

Table 13. Attempts for synthesis of acid chlorides from monoamides

No.	R^1	Reaction time	Remarks
1		15 min	Peak for cyclized tartrimide observed on MS, not isolated
2		30 min	Peak for cyclized tartrimide observed on MS, not isolated

The result indicated that the 2°-amides were susceptible to cyclization and tartrimide formation. Therefore, it was preferable to use 3°-amides generated via reaction of anhydride with 2°-amine for the acid chloride route.

7.1.7. Tartaric acid bisamide (G) synthesis via acid-chlorides

After it was realized that monoamides with fully substituted amide groups were better starting materials than the monoamides with 1°-amine, a modified acid chloride route (Scheme 13) was used for synthesis of target bisamides. Instead of taking 4 equiv. of thionyl chloride in THF, thionyl chloride was used as solvent, and the reactions were carried out at room temperature. The use of thionyl chloride as solvent resulted in side-

reactions that generate a yellow colour. In order to minimize it, small scale reactions were carried out (only to monitor the completion, not to isolate product – first four entries in Table 14) and the optimum duration for the acid chloride formation was found to be 30 min. On completion, the reaction mixture was evaporated to dryness at room temperature using argon flow. After complete evaporation of thionyl chloride, dry THF was added as solvent for the next step, followed by amine.
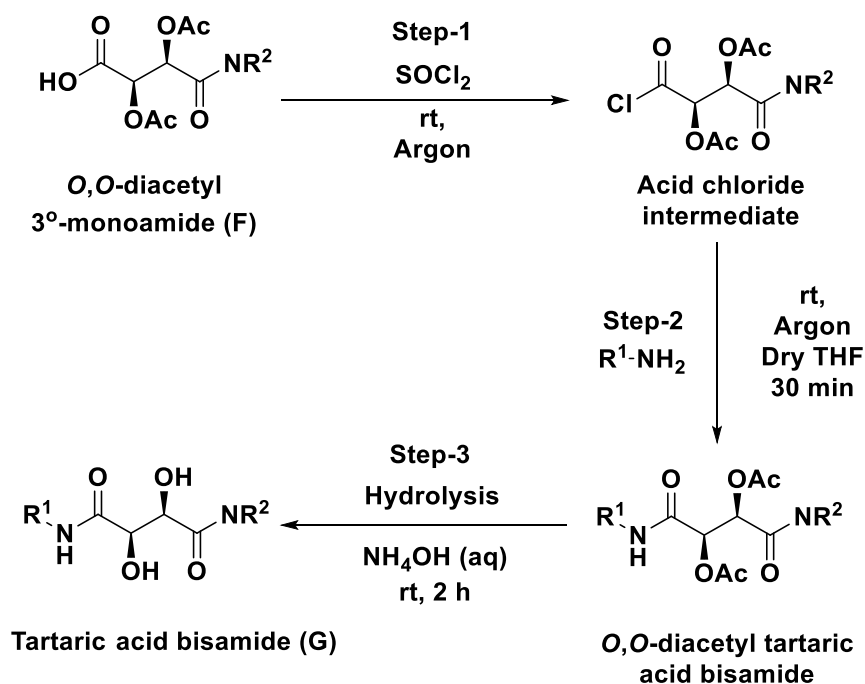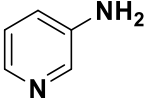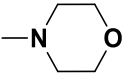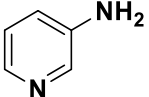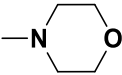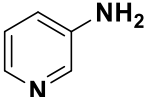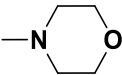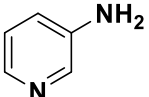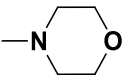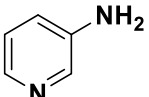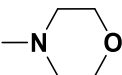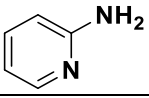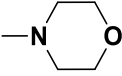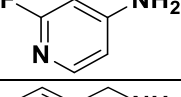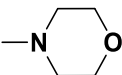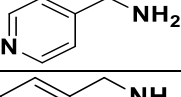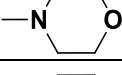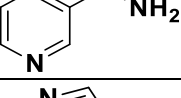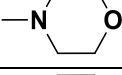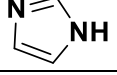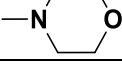
Scheme 13. One-pot three-step method for synthesis of tartaric acid bisamides via acid chlorides

Initially, an attempt was made to isolate the diacetyl bisamide products (step-2 products) with column chromatography. However, on isolation, it was realized that the thionyl chloride evaporation causes deacetylation of one or both acetyl groups, resulting in a mixture of four compounds (with both acetyl groups present, with either of two acetyl groups and with both acetyl groups removed), which was not possible to separate. Hence, the isolation and purification of pure product could be done only after hydrolysis.

As per this protocol, after the first step, the solvent is evaporated with argon flow to dryness. Then the second step is carried out with new solvent and reagents. After the second step, the solvent is again evaporated to dryness under reduced pressure, followed by addition of new solvent, which is again evaporated after reaction completion. It was only after the third step that the crude reaction mixture was subjected to isolation and

purification. Thus, this protocol works satisfactorily as a one-pot three-step method for synthesis of tartaric acid bisamides.

Table 14. Synthesis of tartaric acid bisamide derivatives via acid chloride route

No.	R^1-NH$_2$	NR^2	Step-1 duration	Overall Yield	Remarks
G1	3-aminopyridine	morpholino	15 min	Not isolated	Reaction incomplete
G1	3-aminopyridine	morpholino	30 min	Not isolated	Reaction complete
G1	3-aminopyridine	morpholino	1 hr	Not isolated	Reaction complete
G1	3-aminopyridine	morpholino	2 hr	Not isolated	Reaction complete
G1	3-aminopyridine	morpholino	30 min	27 %	
G2	2-aminopyridine	morpholino	30 min	21 %	
G3	2-fluoro-4-aminopyridine	morpholino	30 min	31 %	
G4	pyridin-4-ylmethanamine	morpholino	30 min	Not isolated	Highly polar product
G5	pyridin-3-ylmethanamine	morpholino	30 min	Not isolated	Highly polar product
G6	imidazole (NH)	morpholino	30 min	No reaction	

7.1.8. Tartrimide (H) synthesis

Tartrimides are an important class of tartaric acid derivatives, used as an intermediate for synthesis of natural products[147,148] and bioactive molecules.[149–151] Dicaffeoyl- or Digalloyl derivatives of tartrimide have also been reported as HIV integrase inhibitors.[152] Generally tartrimides are prepared by reaction of amines with *O*-substituted or unsubstituted tartaric acid derivatives in toluene at high temperature for long time.[153] Another common protocol is three-step synthesis[154–156] via *O,O*-diacyl tartaric acid anhydrides by sequential reaction with acid chloride, amine and again acid chloride, which is a multi-step method where the cyclization reaction takes 5-6 hours or more.
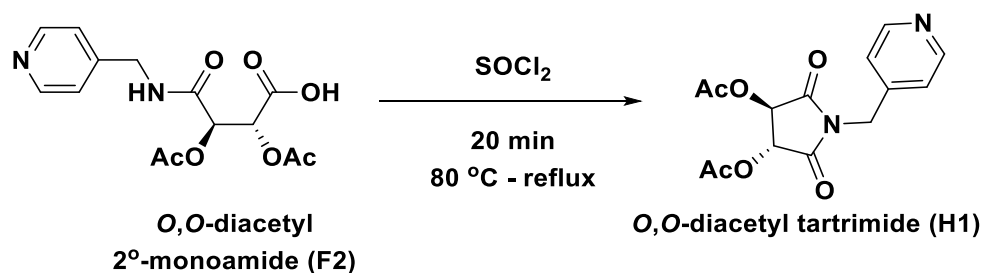
Scheme 14. Synthesis of tartrimide from 2º-monoamides
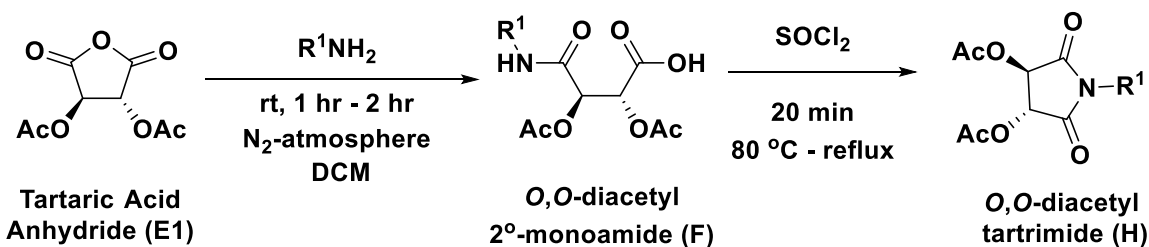
Scheme 15. One-pot two step synthesis of tartrimides from tartaric acid anhydride

Table 15. Tartrimide synthesis

Compound	R^1	Isolated Yield
H1		69 %
H2	H3C—⟨⟩—	62 %
H3		47 %
H4	H3CO—⟨⟩—	65 %

As discussed in section 7.1.6, reaction of *O,O*-diacetyl 2º-monoamides with thionyl chloride at higher temperature was observed to provide cyclic tartrimide based on MS analysis within 15 minutes, however the product could not be isolated. Considering the importance of the series, we sought to develop the observation into an easier and faster protocol for tartrimide synthesis.

As reaction of *O,O*-diacetyl 2º-monoamide with 4 equiv. of thionyl chloride in THF at 80 ºC provided only a tiny amount of tartrimide (detectable only on MS), it clearly suggested need of more thionyl chloride for completion of the reaction. Therefore, a reaction (Scheme 14) was carried out using thionyl chloride as the solvent, which gave the desired tartrimides

in just 20 minutes. However, an attempt of isolation by evaporation of thionyl chloride resulted in deacetylation and gave mixture of products. Hence, the products were isolated by precipitation in pentane, which provided product **H1** in good yields.

An attempt to integrate the reaction with monoamide synthesis step as a one-pot synthesis was also successful as shown in Scheme 15. Reactions of tartaric acid anhydride were carried out with 1º-amines. Upon completion, solvent DCM was evaporated with nitrogen flow up to dryness, and thionyl chloride was added as the new solvent. Upon reflux for 20 min at 80 ºC, each reaction mixure (entries **H2** to **H4** in Table 15) was poured in pentane. The precipitates were filtered and washed with pentane, and then dried in air. The reaction mixtures that did not precipitate were subjected to crystallization in pentane at -20 ºC, and the crystals were filtered followed by washing with cold pentane and drying in air. Overall, the tartrimides were obtained in good yields in shorter time with easy isolation.

7.2. Library properties

As discussed in the introduction (chapter 2), the properties of library-compounds should be reasonably favourable for their use in drug discovery process. Further, as mentioned in section 3.4.3, TABs were considered rather fragment-like, amenable to extension via substituted fragments on both variable positions. Hence, only a small virtual library (Figure 22) with 20 amine varieties for R^1 and 6 amine varieties of R^2 was created and analyzed for distribution of properties.

7.2.1. Diversifiability

TABs offer two places for substitutions, which makes them diversifiable fragments. Further, considering the huge varieties of commercially available amines, the possible diversification potential remains high. Moreover, it is also possible to derivatize the vicinal hydroxyl groups in order to introduce further diversification.

7.2.2. Diversity

As for NSDs and DKPs, the diversity of the virtual library of TABs fragments was also analyzed with respect to various physicochemical parameters such as molecular weight, AlogP, number of rotatable bonds and polar surface area. It can be observed from Figure 23 that the virtual library shows promising distribution patterns for various parameters, especially considering TABs as a series of diverse fragments.
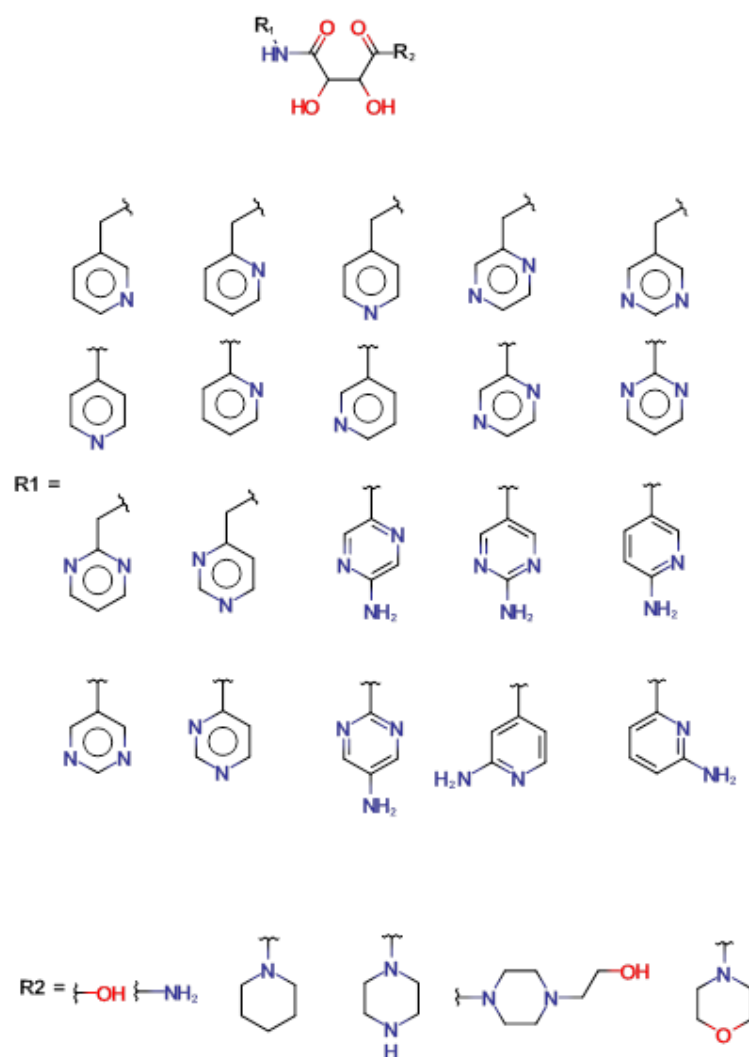
Figure 22. Virtual library generation for TABs

7.2.3. Ease of synthesis

As described in *N*-substituted dipeptide and DKP libraries, ease of synthesis is an important factor for chemoprospecting. The synthesis of TABs requires 5 steps. The first step is short and easy to scale-up. As the product is isolated with precipitation, no column chromatography is required. The second to fifth steps can be carried out as one-pot synthesis, which makes them possible to integrate with a parallel synthesis assembly. The overall yield is medium. Products are usually highly polar, which makes isolation with column chromatography difficult. Reverse phase chromatography, however, may be very beneficial. The overall ease of synthesis is high.

Molecular Weight Distribution - Tartaric acid bisamides

AlogP distribution - Tartaric acid bisamides

Rotatable bond distribution - Tartaric acid bisamides
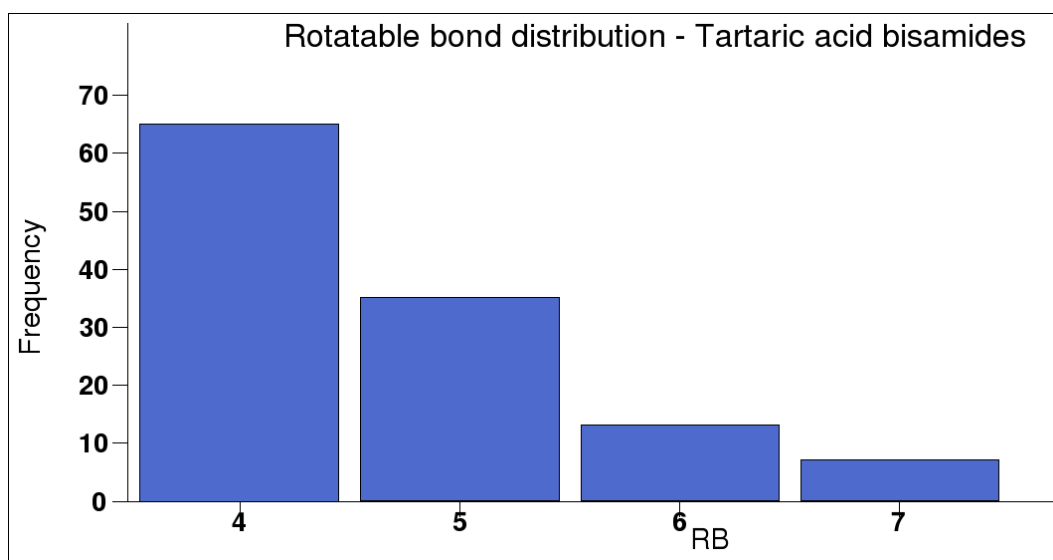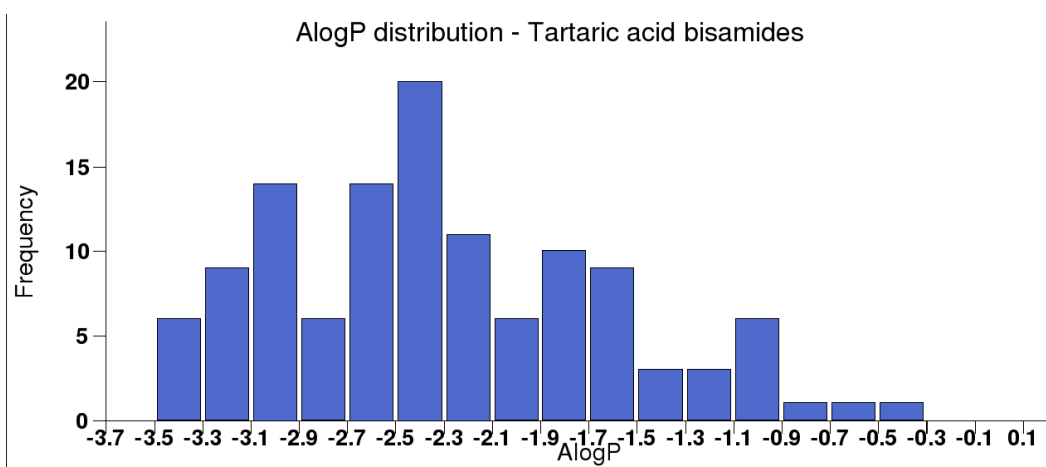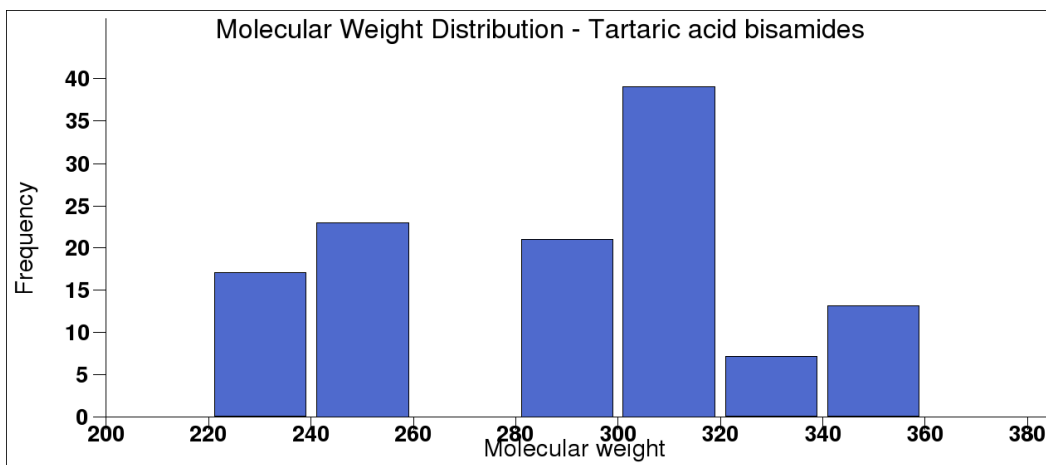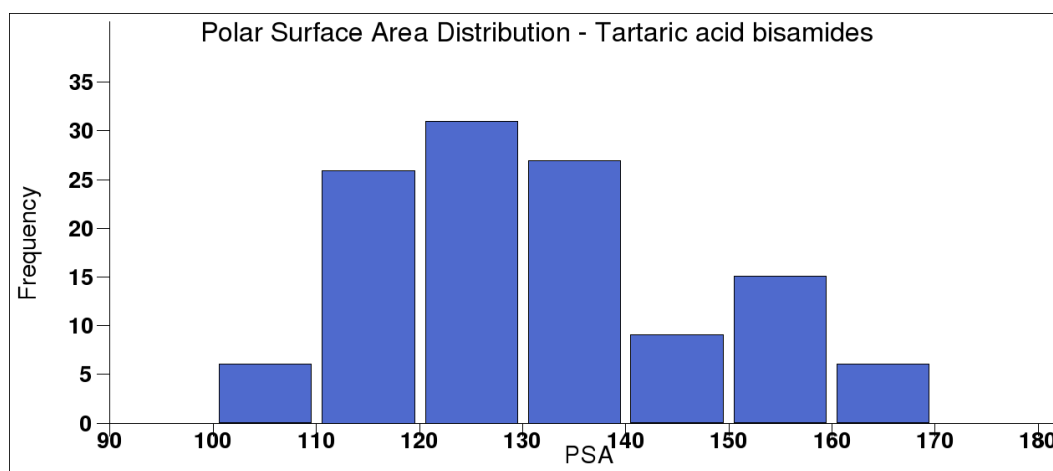
Figure 23. Distribution of a few properties for the virtual library of TABs

7.2.4. Cost and availability of starting materials

The main starting materials for TABs library are tartaric acid and amines. Tartaric acid is quite cheap and commercially available. However, unnatural isomers of tartaric acid are not cheap in pure enantiomeric forms. On the other hand, amines are commercially available in wide varieties, and most of them are cheap.

7.3. Bioactivity studies

The TABs library was tested on assays such as anti-microbial assays, viability assays on cancer cell-lines and anti-oxidant activity assay at MarBio and was also subjected to kinase profiling, but no compound showed any significant activity. The tartrimide series was synthesized only recently, and will be tested on various platforms in future.

The details of tests for bioactivity are included in the appendix.

Theoretical Studies

As described in section 6.1.2, microwave assisted cyclization of dipeptide esters to piperazine-2,5-diones revealed strong substitution effects resulting in significantly higher times required for reaction completion, and relatively low yields for substrates with bigger substitutions. The objective of the theoretical studies was to rationalize such substitution effects in terms of energetics. This rationalization would enable analysis of such effects for different substitution patterns along with stereochemical considerations with a potential applicability to determine synthesizability and thereby compound selection. The theoretical studies were carried out using density functional theory (DFT) on both stages: the cis/trans isomerization of dipeptide esters and cyclization by attack of nucleophilic amine nitrogen to electrophilic carbonyl carbon of ester moiety.

8. Cis/trans isomerization in NMA and GGMe

This chapter will introduce theoretical studies on cis/trans isomerization in secondary amides. Therefore, the discussion in this chapter will focus on primary understanding of the secondary amide geometries in terms of stationary points such as optimized minimum energy trans and cis geometries as well as transition state geometries using a small secondary amide prototype N-methylacetamide (NMA) and a peptidic prototype Glycylglycine methyl ester (GGMe). The detailed cis/trans isomerization process will be discussed in terms of minimum energy reaction paths via transition states and corresponding geometry changes for NMA and GGMe, and the applicability to bigger substituted systems will be established. This chapter will prepare the reader for the discussion on cis/trans isomerization in substituted peptidic systems.

8.1. Introduction

8.1.1. Cis/trans isomerization in secondary amides

Due to partial double bond character of the C-N bond in amide bond, the free rotation around the bond is hindered by a relatively high energy barrier, resulting in cis and trans isomers where trans is energetically favoured over cis.[157–160] A majority of cis peptide bonds have been observed to be tertiary amide bonds (i.e. preceding prolyl residues) as the relative stability of corresponding cis and trans conformations would be similar.[160–162] Conversely, secondary (i.e. non-prolyl) cis peptide bonds are rare but they do occur[162–164] and play a key role in various biological functions, such as chemo-mechanical cycle of motor proteins[165], protein folding[166–168] and catalytic activity[169] of enzymes such as cyclophilin A. Moreover, cis/trans isomerization is also involved in cascade dissociation of peptide cation radicals[170], and hence is important for peptide sequencing. From a synthetic chemistry point of view, cis/trans isomerization is an essential step for cyclization reactions of peptides, such as formation of piperazine-2,5-diones.[171] In the drug discovery area, cis/trans isomerization is also linked with efficacy and selectivity of peptidomimetics, hence attempts have been made to tailor the flexibility of the rotamers and overall conformations by introducing constraints such as intramolecular hydrogen bonding and/or steric bulk.[172,173]

8.1.2. Theoretical studies

While cis/trans isomerization has been a subject of interest for chemists since 1950s[174–179], effective theoretical studies have become possible only rather recently. QM methods on secondary cis/trans isomerizations have been employed to explain and predict rotational barrier values in agreement with experimental values[180,181], and studies have been carried

out to understand different phenomena such as role of conjugation[182] and solvent effects with molecular dynamics[183]. The theoretical studies with *ab initio* methods on secondary amide prototypes[184–186] have highlighted the importance of pyramidalization of amide nitrogen along with two transition state geometries. These studies described two transition states interchanging though a high-energy saddle point of 2nd order along with a reaction coordinate path involving the more stable transition state, however a discussion on changing planarity and hybridization state of amide nitrogen was lacking. Following this, Mantz et al. described[187] cis/trans isomerization using force-field methods by generation of an ensemble of transition state geometries, which described plausible hybridization change for amide nitrogen around ω dihedral values between 90º-110º. However, the observations from force-field methods indicated a single path for cis/trans isomerization passing through both transition states via the high-energy saddle point of 2nd order, requiring further clarity on geometry changes during the isomerization. Hence, a fresh relook with assessment on the reaction path(s), corresponding geometry changes and the relevance/importance of the high-energy 2nd order saddle point can provide more details about geometry changes.

In addition, it has been observed experimentally[188,189] that moieties on both sides of amide bond may affect the cis/trans isomerization process. Therefore, it seemed interesting to replicate such studies on bigger moieties, such as dipeptides or their derivatives, to assess the effect of peptide backbone on isomerization. Juxtaposing observations on a small amide prototype such as *N*-methylacetamide (NMA) with that on a model dipeptide scaffold would provide an excellent foundation for such studies. A **dipeptide ester scaffold**, owing to the use in pro-drugs and peptidomimetics[190–193] and as a precursor of piperazine-2,5-diones via cyclization where cis/trans isomerization is an important step, can serve as a valuable model dipeptide scaffold not only for understanding of isomerization in peptide chains, but also in peptidomimetic drug design and synthetic chemistry. Overall, such studies can provide a realistic view on cis/trans isomerization in peptidic systems that may serve as a basis for further studies with larger systems.

In this context, our theoretical assessment on cis/trans isomerization in NMA and a dipeptide ester, glycylglycine methyl ester (GGMe) is presented here (Figure 24).
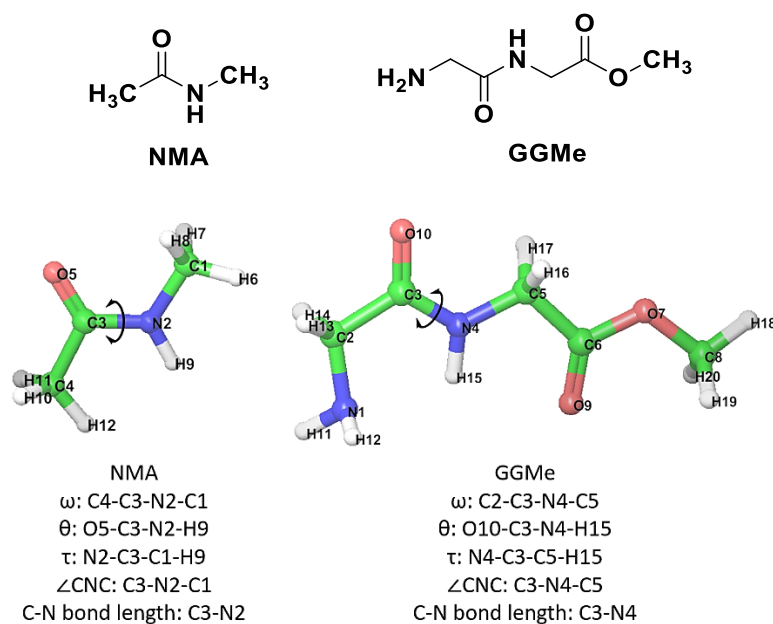
Figure 24. NMA and GGMe structures, atom numbers and different measures

8.2. Stationary points in NMA and GGMe

8.2.1. Optimized trans and cis geometries

The optimized minimum energy geometries of NMA and GGMe were found to be in an extended form, albeit with the "peptide backbone" slightly offset from complete planarity (Figure 25). A weak N-H-N hydrogen bond (2.24 Å) between N-terminal nitrogen and the hydrogen attached to amide nitrogen appeared to contribute to further stabilization of trans geometry of GGMe as evident from shorter C-N amide bond length than that in NMA (Table 16). The optimized trans and cis geometries of NMA appear more planar than those reported previously at SCF and MP2 levels though the relative stabilities of trans and cis geometries are in excellent agreement.[186]

Table 16. Optimized minimum energy geometries of NMA and GGMe in cis and trans form

Structure	Cis/Trans	ω	Ө	C-N bond length (Å)	∠CNC	Relative energy[a]	ΔG[b]
NMA	Trans	178.7	176.8	1.366	121.7	0.0	0.0
NMA	Cis	5.3	3	1.370	127.1	2.3	2.2
GGMe	Trans	178.3	180.7	1.354	121.5	0.0	0.0
GGMe	Cis	0.1	-0.1	1.367	127	4.6	4.5

[a] Gas phase energy relative to the minimum energy trans geometry, calculated in kcal/mol at B3LYP/6-311G-3DF-3PD++ level.
[b] Relative free energy at 298.15 K relative to minimum energy trans geometry, calculated in kcal/mol at B3LYP/CC-PVTZ(-F)++ level.
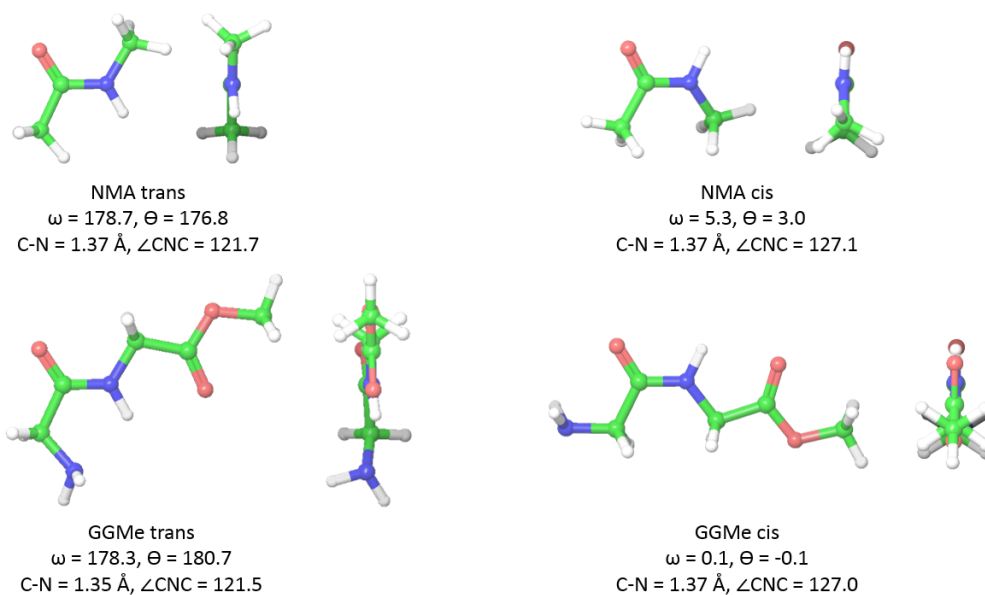
Figure 25. Trans (left) and cis (right) optimized energy minimum geometries of NMA (above) and GGMe (below).

8.2.2. Transition state geometries (TSGs) for NMA and GGMe

For trans to cis isomerization, the rotation of amide bond in terms of ω dihedral would be from ~180º to ~0 (= 360º), which can occur in both directions, i.e. 180º - 270º - 360º (positive direction) or 180º - 90º - 0º (negative direction). Similarly, the rotation from cis to trans in terms of ω dihedral would be from ~0º to ~±180º, which can occur either 0º - 90º - 180º (positive direction) or 0º - (-90º) - (-180º) (negative direction). Owing to the symmetry of NMA and GGMe, it would be sufficient to describe the isomerization between ~0º to ~180º via positive ω values.

Instead of creating reasonable transition state geometries for transition state search manually, we chose to use generate such reasonable geometries using a relaxed coordinate scan (RCS) protocol along the ω dihedral, and the energy barrier geometries thus obtained were used for transition state search. The RCS along the ω dihedral from trans isomer led to *syn* energy barrier geometry (EBG – an optimized RCS geometry near the energy barrier) near ω = ~60º, while the trajectory starting from the cis isomer led to *anti* EBG near ω = ~120º. The *syn* and *anti* terms represent the orientation of the lone electron pair of pyramidalized nitrogen with respect to C=O bond. The transition state geometries (TSGs – obtained with QST transition state search, with one imaginary frequency) were found to be in agreement with previous studies[186,187], and also extremely similar to the EBGs, confirming the validity of the EBGs (Table 17). Henceforth, EBG$_{syn}$ will refer to the *syn*

EBGs, TS$_{syn}$ will refer to *syn* TSGs, EBG$_{anti}$ will refer to the *anti* EBGs and TS$_{anti}$ will refer to *anti* TSGs.

Table 17. Comparison of energy barrier geometries and transition state geometries

Geometry	ω	θ	C-N bond length (Å)	∠CNC	τ	Relative energy[a]	ΔG[b]	Global RMSD
NMA - EBG$_{anti}$	122.6	60.3	1.455	112.2	-32.5	18.4	16.3	0.0008
NMA - TS$_{ant}$	122.6	60.7	1.455	112.2	-32.4	18.6	16.3	
NMA - EBG$_{syn}$	59.7	114.2	1.449	115.3	29.0	21.8	19.6	0.0027
NMA - TS$_{syn}$	60	115.1	1.449	115.2	29.2	21.9	19.6	
GGMe - EBG$_{anti}$	118.3	56.6	1.449	113.1	-31.9	19.5	17.4	0.0031
GGMe - TS$_{anti}$ Rota I	118.1	56.1	1.449	113.1	-31.8	19.6	17.4	
GGMe - EBG$_{syn}$	61	110.3	1.451	117.2	26.7	25.4	23.5	0.0131
GGMe - TS$_{syn}$ Rota I	60.7	106.5	1.449	117.8	25.4	25.3	23.5	

[a] Gas phase energy relative to the minimum energy trans geometry, calculated in kcal/mol at B3LYP/6-311G-3DF-3PD++ level.
[b] Relative free energy at 298.15 K relative to minimum energy trans geometry, calculated in kcal/mol at B3LYP/CC-PVTZ(-F)++ level.
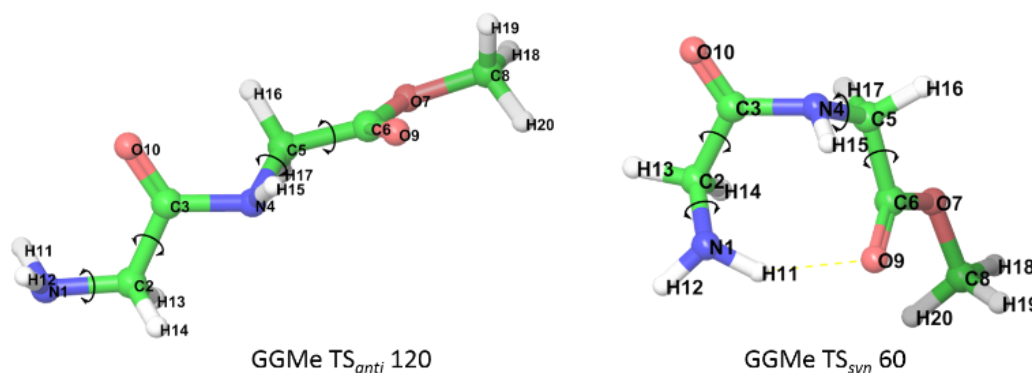
Figure 26. Rotations around the rotatable bonds generate local saddle points corresponding to the transition state rotamers.

As a small molecule with only methyl groups beyond the amide moiety, NMA does not have asymmetric rotatable bonds, or substitutions that would generate multiple local minima rotamers for a coordinate (such as ω) pertaining to cis/trans isomerization, hence a single geometry can represent a particular optimized state such as trans, cis, or a transition state. However, this is not the case for bigger systems such as GGMe as multiple rotatable bonds with substituted carbons outside amide moiety (Figure 26) give rise to multiple local transition state rotamers. Accordingly, reasonable rotamer geometries were generated from TS$_{anti}$ Rotamer I and TS$_{syn}$ Rotamer I, and subjected to transition state search protocols. Thus, a total of 14 rotamers of TS$_{anti}$ and TS$_{syn}$ (7 of each type) were chosen (Table 18 and

Figure 27). Rotamer I for TS*anti* and rotamer VII for TS*syn* were found to be most stable among them, and will be used for further discussion on transition state geometries (Figure 28 and Table 19). It is also important to note that there may exist more transition state rotamers of GGMe.

Table 18. Rotamers of TS*anti* near ω = ~120 (upper part) and of TS*syn* near ω = ~60 (lower part).

GGMe TS*anti* 120 rotamers	Relative Energy[a]	ΔG[b]	Dihedral 3-2-1-11	Dihedral 1-2-3-4	ω	Dihedral 3-4-5-6	Dihedral 4-5-6-7	ϴ
I	19.6	17.4	57.4	-179.3	118.1	158.2	158.8	56.1
II	20.4	24.0	149.4	173.1	115.4	158.8	158.4	54.3
III	24.4	22.0	-156.4	77.6	118.5	161.1	157.1	55.1
IV	21.7	18.9	-178.4	-40.4	122.7	156.0	160.8	58.6
V	20.5	18.2	70.6	-176.8	115.8	-76.0	-158.1	46.3
VI	24.6	30.0	-162.2	89.8	116.6	-70.5	-159.5	46.4
VII	23.5	45.0	-57.8	47.3	112.8	92.1	-159.7	52.3

GGMe TSsyn 60 rotamers	Relative Energy[a]	ΔG[b]	Dihedral 3-2-1-11	Dihedral 1-2-3-4	ω	Dihedral 3-4-5-6	Dihedral 4-5-6-7	ϴ
I	25.3	23.5	-86.0	62.2	60.7	-89.5	174.2	106.6
II	24.4	22.4	58.4	60.3	63.6	-93.1	172.7	115.7
III	23.5	21.5	-32.3	-166.5	63.4	-98.7	175.7	116.1
IV	25.1	23.0	-171.4	-8.9	64.5	-124.4	-175.4	116.4
V	26.2	24.0	-86.2	42.3	56.6	-161.0	-155.8	110.8
VI	28.5	26.5	-83.8	60.7	62.8	-88.0	-23.2	107.5
VII	23.3	20.8	56.5	179.0	61.3	-95.3	172.3	115.9

[a] Gas phase energy relative to the minimum energy trans geometry, calculated in kcal/mol at B3LYP/6-311G-3DF-3PD++ level.
[b] Relative free energy at 298.15 K relative to minimum energy trans geometry, calculated in kcal/mol at B3LYP/CC-PVTZ(-F)++ level.

GGMe TS$_{anti}$ 120 I

GGMe TS$_{anti}$ 120 II

GGMe TS$_{anti}$ 120 III

GGMe TS$_{anti}$ 120 IV

GGMe TS$_{anti}$ 120 V

GGMe TS$_{anti}$ 120 VI

GGMe TS$_{anti}$ 120 VII

GGMe TS$_{syn}$ 60 I

GGMe TS$_{syn}$ 60 II

GGMe TS$_{syn}$ 60 III

GGMe TS$_{syn}$ 60 IV

GGMe TS$_{syn}$ 60 V

GGMe TS$_{syn}$ 60 VI

GGMe TS$_{syn}$ 60 VII
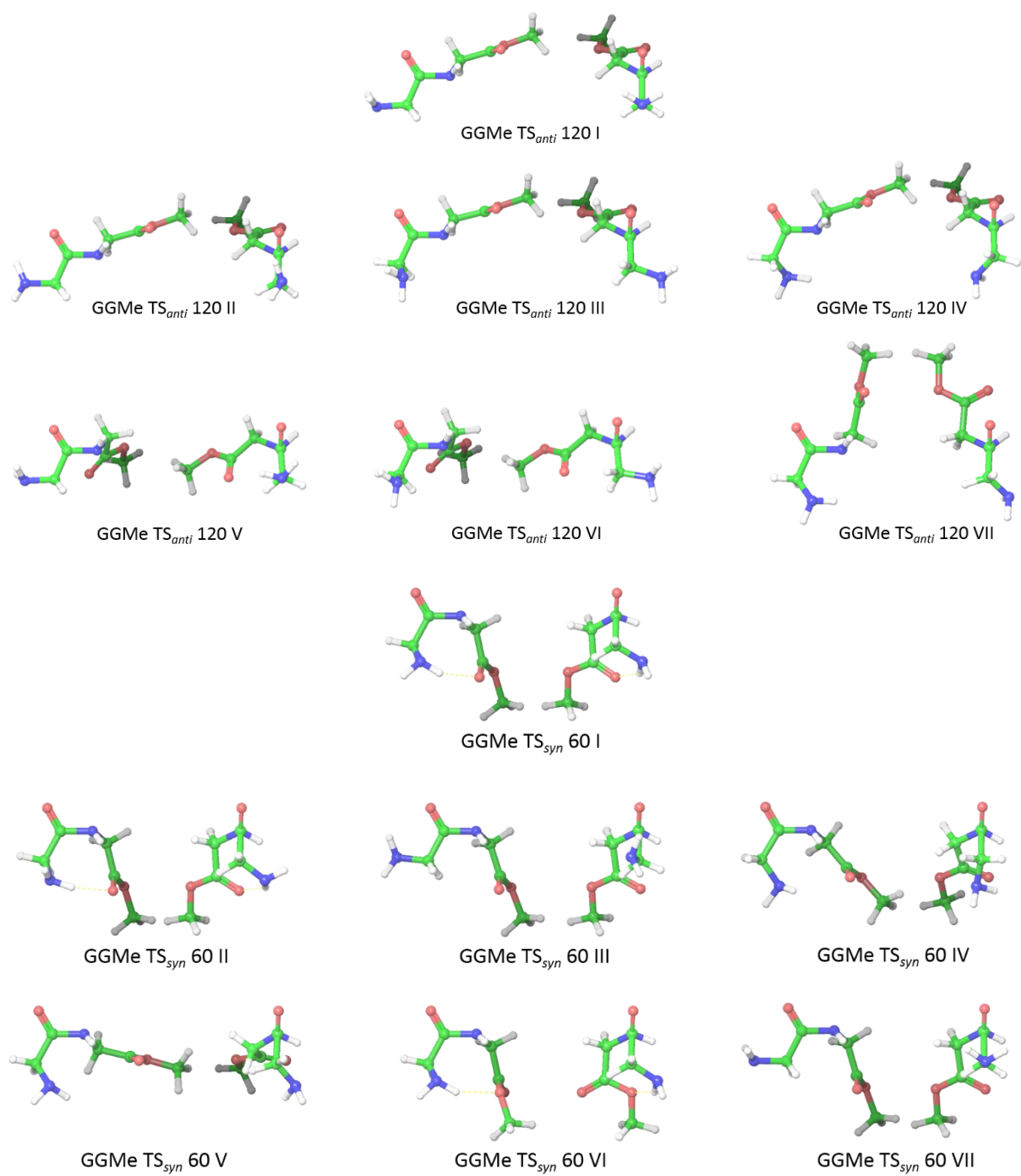
Figure 27. Transition state rotamers for GGMe. Seven rotamers for TS$_{anti}$ 120 (upper) and seven rotamers for TS$_{syn}$ 60 (lower). Each geometry is projected from side and from N-terminal end.
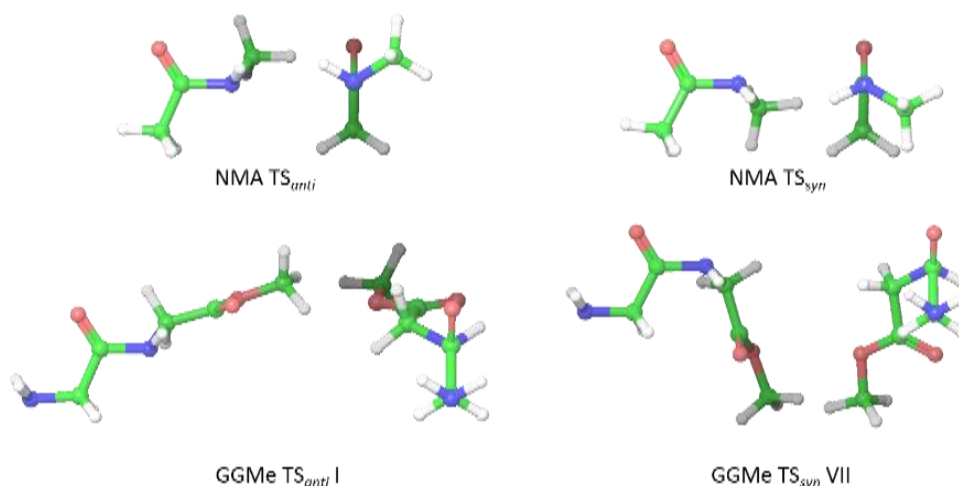
Figure 28. Transition state geometries for NMA and GGMe in gas phase.

Table 19. Comparison of transition state geometries of NMA and GGMe

Geometry	ω	Θ	C-N bond length (Å)	∠CNC	τ	Relative energy[a]	ΔG[b]
NMA TS$_{anti}$	122.6	60.7	1.455	112.2	-32.4	18.6	16.3
NMA TS$_{syn}$	60	115.1	1.449	115.2	29.2	21.9	19.6
GGMe TS$_{anti}$ I	118.1	56.1	1.449	113.3	-32.0	19.5	17.4
GGMe TS$_{syn}$ VII	61.3	115.9	1.450	116.3	29.3	23.3	20.8

[a] Gas phase energy relative to the minimum energy trans geometry, calculated in kcal/mol at B3LYP/6-311G-3DF-3PD++ level.

[b] Relative free energy at 298.15 K relative to minimum energy trans geometry, calculated in kcal/mol at B3LYP/CC-PVTZ(-F)++ level.
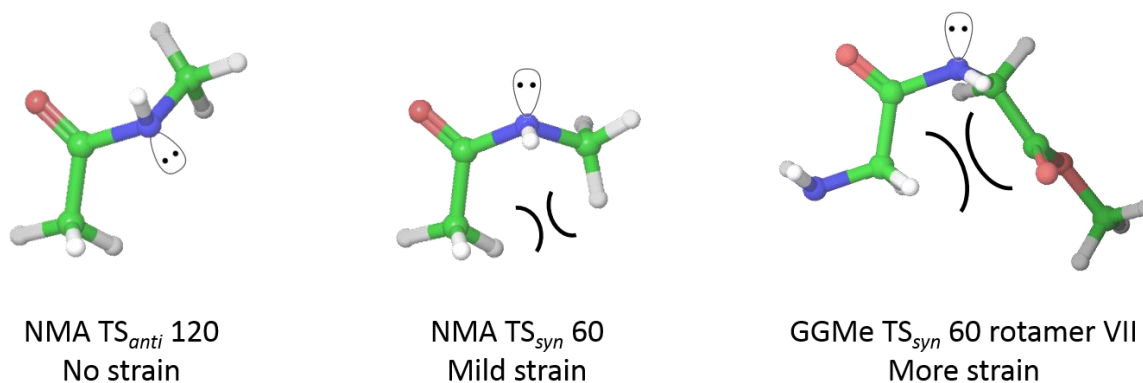
Figure 29. Steric bulk caused by atoms neighbouring the amide moiety, and repulsion between C=O bond and lone N-electron pair lead to significant strain in *syn* geometries.

Interestingly, the pyramidal character of amide nitrogen in *syn* and *anti* TSGs differs in terms of bond angle ∠CNC and improper dihedral angle τ (Table 19). The values for ∠CNC were observed to be ~112-113º for *anti* TSG, which is close to expected value of 109º for

sp^3 hybridized pyramidal nitrogen, but for *syn* TSG, the values were higher, at ~115-116°. Similarly, while the values of τ for *anti* TSG remained at ~32°, close to expected values of 33-34° for sp^3 pyramidalized nitrogen, those for *syn* TSG were at ~29°. The strained geometry for *syn* TSG clearly reflects the effect of steric bulk due to the proximity of atoms in neighbourhood of amide moiety (Figure 29). This strain, together with the repulsion between C=O bond and lone pair of electron, results in significant destabilization of the geometry. This destabilization makes *syn* TS more unstable by ~3-4 kcal/mol compared to *anti* TS.

8.2.3. Molecular orbitals

Although molecular orbital analysis on transition state geometries is complicated by the mixing of several molecular orbitals, the molecular orbital HOMO-10 on *syn* RBGs (and *syn* TSGs as well) of GGMe provided a relatively clear view. It is apparent from Figure 30 that the angle between two planes containing C=O π orbital and orbital with N-lone pair is 90° – implying lack of overlap. This further confirms the complete pyramidalization of the amide nitrogen at the energy barrier and explains the increase in C-N bond length to 1.45-1.46 Å compared to 1.35-1.37 Å in energy minimized cis or trans geometries.

GGMe TS$_{syn}$ 60 l
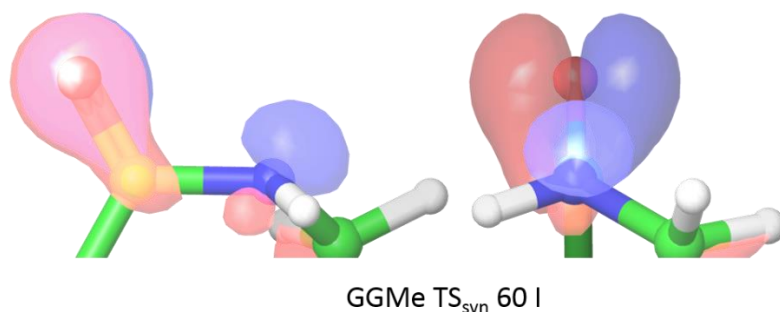
Figure 30. The molecular orbital HOMO-10 of TS$_{syn}$ of GGMe showing 90° angle between two planes of C=O π orbital and orbital with N-lone pair.

8.3. Geometry changes during cis/trans isomerization

8.3.1. Reaction paths and geometry changes in cis/trans isomerization

The most important feature of the cis/trans isomerization is the progression of changes of the effective atomic orbital hybridization of the amide nitrogen. Cis and trans geometries have a planar sp^2 amide nitrogen, but a pyramidal sp^3 hybridization state characterizes both transition states. Previous studies[186,187] suggested an inversion of sp^3 pyramidal nitrogen in terms of interchange between *syn* and *anti* transition states via a high-energy 2nd order

saddle point with $\omega = \pm 90°$. However, no reaction coordinate path connecting the two transition states via the saddle point with planar nitrogen was established. Based on the geometry changes shown by our intrinsic reaction coordinate (IRC) study, we describe here a significantly different and more detailed account of reaction paths with corresponding geometry changes driving the cis/trans isomerization (Figure 31).
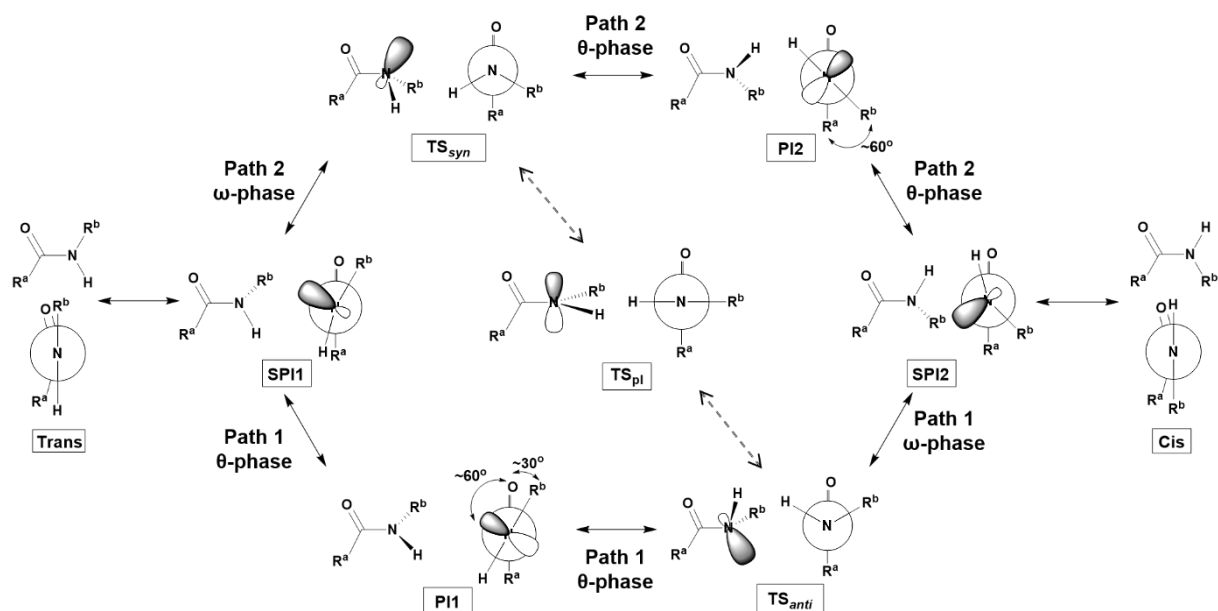
Figure 31. Both paths of cis/trans isomerization and relevant geometries. R^a and R^b represent substitutions attached to the carbonyl carbon and amide nitrogen respectively. For simplicity, the rotation is shown only for positive ω values between 0° (cis) and 180° (trans).

8.3.2. Two reaction paths

Cis/trans isomerization is a function of both ω and Θ, and both dihedrals are similar for trans and cis geometries (close to $\pm 180°$ and 0° respectively). However, the values of both dihedrals differ substantially for the transition states i.e. $\omega = \sim 60°$, $\Theta = \sim 120°$ for TS_{syn} and $\omega = \sim 120°$, $\Theta = \sim 60°$ for TS_{anti}. This implies that along the paths to the transition states, the two dihedral angles do not change linearly with respect to each other (i.e. rotation about the amide C-N bond must involve loss of planarity of at least one of the bonded atoms). This results in two different paths via two different transition states for the isomerization. Accordingly, Figure 31 details the two distinct paths, path 1 via TS_{anti} and path 2 via TS_{syn}. Because TS_{anti} is more stable transition state, path 1 is energetically more favourable than path 2; however, it is also important to note that (as will be discussed later) the relative stability of the *syn* and *anti* transition states may vary depending on the neighbouring substitutions and conformational state of the overall geometry.

87

The isomerizations begin via common reaction paths, characterized mostly by change in ω, until a bifurcation point is reached. During the common paths, Ө values remain close to 0 or ±180 (for isomerization from cis or trans geometries respectively). This preserves the extended π-bonding orbital overlap, in synchrony with atomic orbital hybridization changes at the nitrogen atom. These change progressively from sp^2 towards sp^3 as amide nitrogen geometry changes from planar to semiplanar/half-pyramidal at the bifurcation point (SPl1 for trans or SPl2 for cis). Here, the paths diverge, and ω and Ө change disproportionately along the two paths to complete the pyramidalization of nitrogen at the two respective transition states. Progressing past the transition states, the situation reverses, and the path previously characterized by a relatively greater variation of ω now shows greater variation of Ө and vice versa, continuing until the other semiplanar/half-pyramidal bifurcation point is reached. So overall, each reaction path is composed of two phases, with one phase characterized mostly by changing ω (henceforth referred to as "ω-phase") and the other phase characterized mostly by changing Ө and very little change in ω (henceforth referred to as "Ө-phase").

As the nitrogen becomes pyramidal during the isomerization, the two paths differ qualitatively with respect to changes in the planarity of nitrogen as described in terms of improper dihedral τ (τ is explained in Figure 24). The ω-phase proceeds from the bifurcation point to the transition state through continued progression from half- to full pyramidal nitrogen geometries. On the other hand, the Ө-phase proceeds from the bifurcation point to the transition state via inversion of the nitrogen geometry, i.e. first with loss of half-pyramidal geometry to become planar ($\tau = \sim0$) and then progressing to a full pyramidal state with inverted geometry (from positive to negative values of τ and vice versa).

8.3.3. The ω phase

The ω phase describes the reaction coordinate paths between SPl1 and TS$_{syn}$ and between SPl2 and TS$_{anti}$. As mentioned earlier, this phase is characterized by changes of both ω and Ө, although ω changes dominate until close to the transition state, when both change at similar rates.

8.3.4. The Ө-phase

The Ө-phase describes the reaction coordinate paths between SPl1 and TS$_{anti}$ and between SPl2 and TS$_{syn}$. This phase is driven almost exclusively by changes in Ө with small or no

change in ω. A drastic change in geometry is observed during this phase as the amide nitrogen undergoes inversion, seen as a flip in the hydrogen position. During inversion, the geometry of nitrogen changes from pyramidal at the transition state, moves through a planar state and becomes semi-pyramidal with inverted geometry at the bifurcation point. Correspondingly, its atomic orbital hybridization state changes from the sp^3 state, passes through an sp^2 state (Pl1 or Pl2) and regains a partial sp^3 state with an inverted geometry. The intermediate geometries with planar nitrogen (Pl1 and Pl2) show the *p*-orbital of nitrogen with its lone pair of electrons in a *gauche*–like conformation with respect to C=O bond. This corresponds well with a chemically intuitive expectation of the lowest energy geometry for the planar nitrogen state, although other interactions, such as a steric interaction between R^a and R^b in Pl2, may also play a role. This is also in stark contrast to the high energy 2nd order saddle point (TS_{pl}) mentioned in previous studies with an eclipsed geometry. Instead, the geometries with planar nitrogen Pl1 and Pl2 have partial orbital overlap and hence are stabilized geometries compared to respective transition states.

GGMe TS$_{syn}$ 60 I   GGMe SPl2 60 I
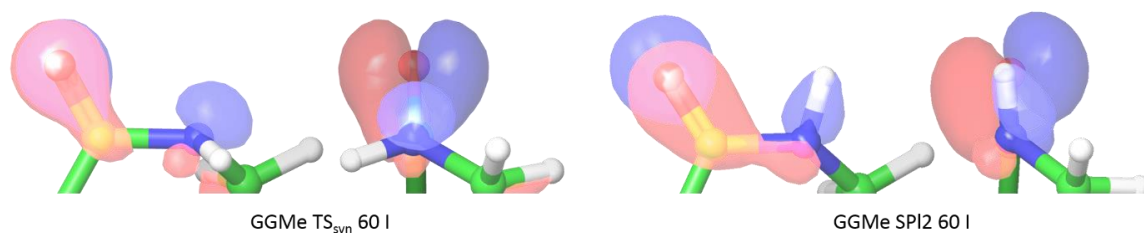
Figure 32. The molecular orbital HOMO-10 of rotamer GGMe TS*syn* 60 I shows 90º angle between two planes of C=O π orbital and orbital with lone N-electron pair, whereas the same orbital of corresponding SPl2 bifurcation point geometry showing restored overlap between the two orbitals.

As the hydrogen-flip (pyramidal inversion) approaches the bifurcation points (SPl1 and SPl2), the overlap of orbitals continues to increase and stabilizes the geometries in both paths (Figure 32). Conversely, in the reverse direction, reaching the transition state from a bifurcation point via Ɵ-phase requires progressively higher energy geometries as decreasing overlap progresses through inversion to the transition state with little or no change in ω. This, in turn, explains why RCS on ω dihedral from trans geometry leads to *syn* EBG, while from cis geometry leads to *anti* EBG in its attempt to find local minimum for given ω coordinate near bifurcation points.
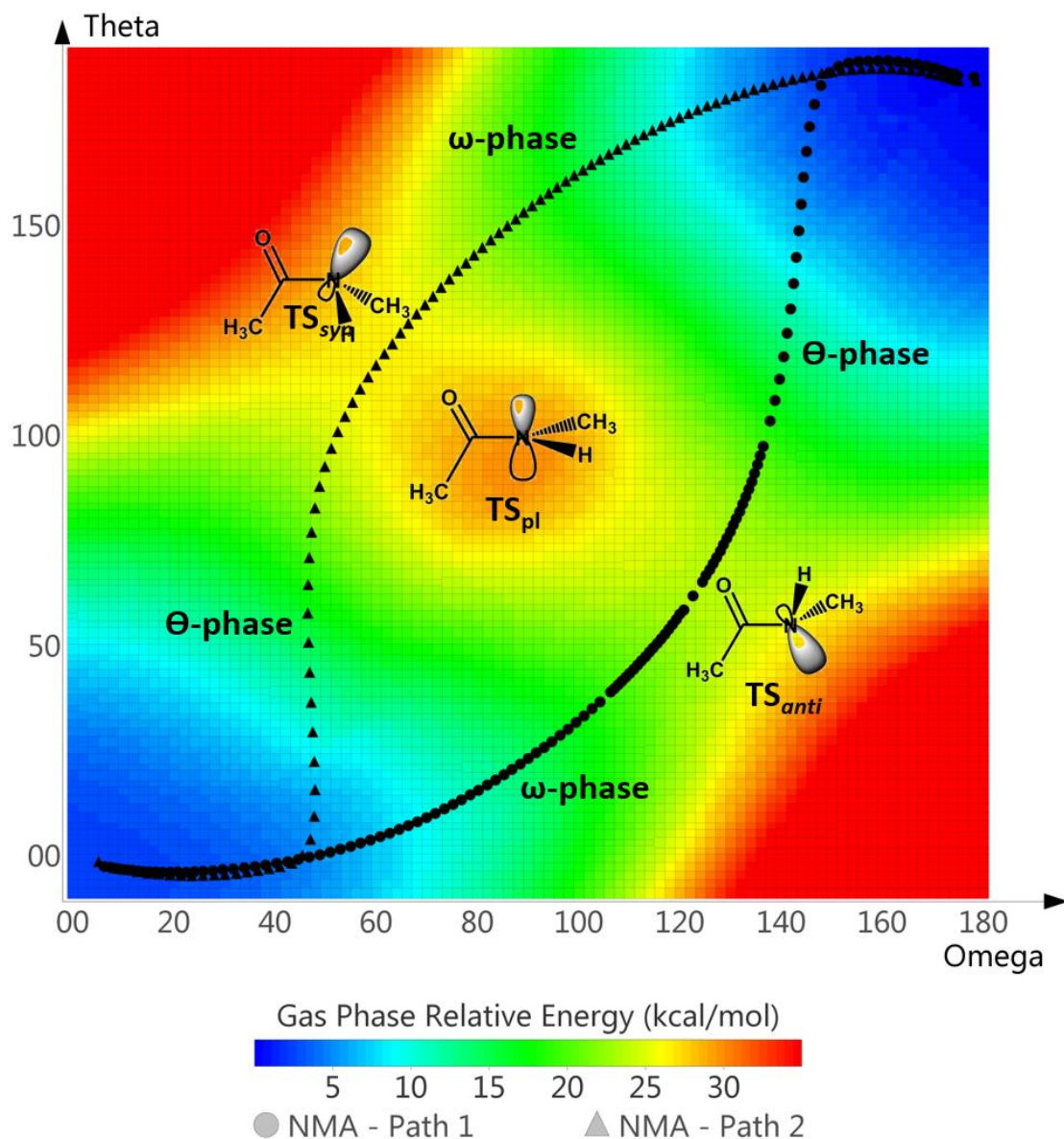
Figure 33. Contour map for ω and Θ dihedral coordinates for cis/trans isomerization of NMA. The black marks represent IRC geometries from path 1 and path 2. The minimized coordinate scan geometries for the creation of contour were generated using MMFFs force-field, while energy values for the generated geometries were calculated at B3LYP/6-31G**++ level.

8.4. Cis/trans isomerization in NMA

The IRC study showed that cis/trans isomerization in NMA precisely follows the scheme described above (Figure 31) and its corresponding geometry changes. Figure 33 illustrates the IRC paths for cis/trans isomerization in NMA and important geometric measurements of the relevant geometries have been tabulated in Table 20. The Newman projections of transition states and intermediate geometries display the geometry changes clearly (Figure 34).

90

Table 20. Important measurements of cis/trans isomerization path geometries of NMA.

Title	C-N bond length (Å)	∠CNC	ω	Θ	τ	Relative energy[a]	ΔG[b]
NMA – trans	1.366	121.68	178.72	-176.78	2.56	0.00	0.0
NMA – cis	1.370	127.14	5.27	-3.08	-4.32	2.3	2.2
NMA – TS$_{anti}$	1.455	112.20	122.64	60.75	-32.42	18.7	16.3
NMA – Pl1	1.383	120.07	143.50	147.61	-2.18	7.7	7.4
NMA – SPl1	1.377	120.33	149.57	184.80	16.59	2.4	1.3
NMA – TS$_{syn}$	1.449	115.20	60.00	115.18	29.25	21.9	19.7
NMA – Pl2	1.395	122.02	46.83	42.13	-0.92	12.9	12.7
NMA – SPl2	1.386	122.77	44.43	-2.29	-23.68	4.8	3.6

[a] Gas phase energy relative to the minimum energy trans geometry, calculated in kcal/mol at B3LYP/6-311G-3DF-3PD++ level.
[b] Relative free energy at 298.15 K relative to minimum energy trans geometry, calculated in kcal/mol at B3LYP/CC-PVTZ(-F)++ level.
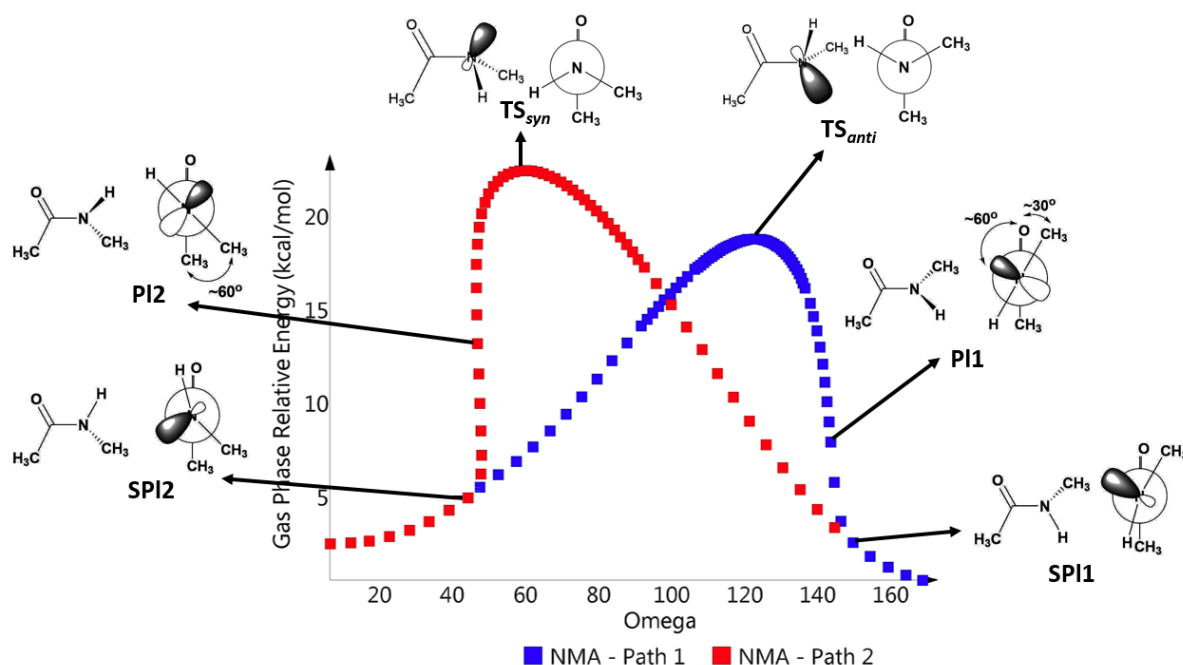
Figure 34. Plot showing the geometries and corresponding energy changes for both isomerization paths for NMA with change in ω dihedral.

During the Θ-phase, as a result of nitrogen inversion and hydrogen flip in path 1 (via TS$_{anti}$), the Θ dihedral shows a drastic change, from ~60° to ~180°, while ω changes only from ~120° to ~150°. For path 2 (via TS$_{syn}$), the change in ω during nitrogen inversion is even smaller, from ~60° to ~45°, while Θ changes from ~120° to ~0°. The geometries Pl1 and Pl2 are significantly more stable (by ~9 kcal/mol for path 1 and by ~7 kcal/mol for path 2) than the corresponding transition states due to partial restoration of overlap unlike the high-energy saddle point of 2nd order TS$_{pl}$ (Figure 33). As the geometry approaches bifurcation

point geometries SPl1 and SPl2, it remains within 2-5 kcal/mol of the energy minimum geometry.

From Figure 33, path 1 and path 2 geometries along the common path might appear to follow slightly different paths in terms of ω and Ѳ. However, a closer look reveals that the geometries differ only in the orientation of one or both methyl groups attached to the amide moiety, with no real energy barrier for interconversion (Figure 35). In other words, the geometries along the common path are easily interconvertible between path 1 and path 2, essentially implying the common path.

NMA geometries at ω = ~30
Path 1: ω = 29.7, Relative energy = 3.8
Path 2: ω = 30.1, Relative energy = 3.2

NMA geometries at ω = ~159
Path 1: ω = 159.2, Relative energy = 1.12
Path 2: ω = 159.9, Relative energy = 1.04
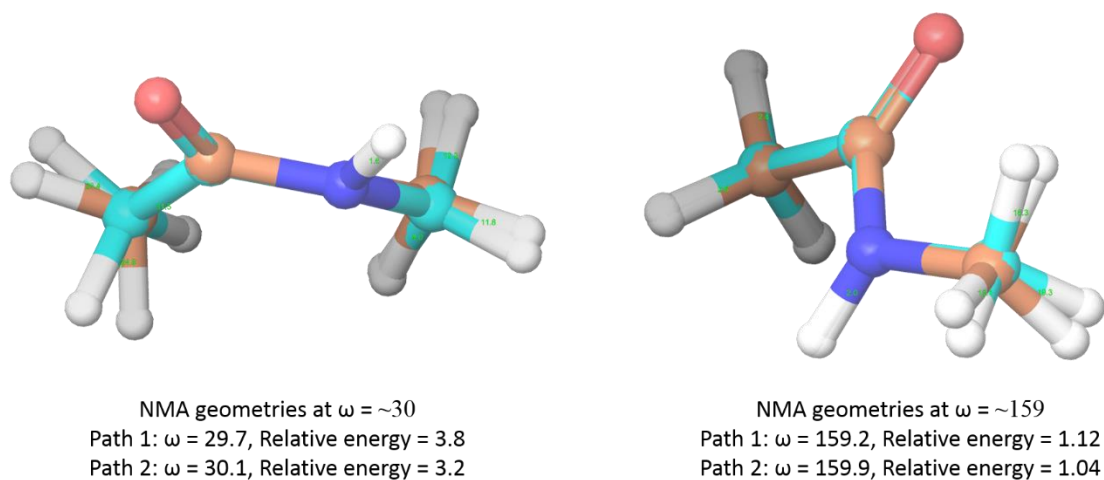
Figure 35. Path 1 and path 2 geometries of NMA along the common path show easily interconvertible geometries.

8.5. Cis/trans isomerization in GGMe and applicability to peptidic systems

As discussed in section 8.2.2, larger systems such as GGMe and other peptidic systems contain asymmetric rotatable bonds or substitutions, and corresponding energy landscapes of (much) greater dimensionality. Thus, transition state saddle points (such as TS_{syn} and TS_{anti} with energy maxima along ω pertaining to cis/trans isomerization) occur multiply for the multiple geometries of the larger system. This greatly complicates the study of cis/trans isomerization paths. The energy surface with respect to ω and Ѳ dihedrals differs for each rotamer and may involve coupled rotations of rotatable bonds. However, the rotamers may be classified according to the degree to which they interfere (couple) with the isomerization transition; coupling may arise from electronic influence on the peptide bond itself. As the simplest form of a dipeptide, GGMe introduces only rotamers with relatively weak coupling to ω rotation. Thus, cis/trans isomerization may be studied by considering multiple rotamers individually.
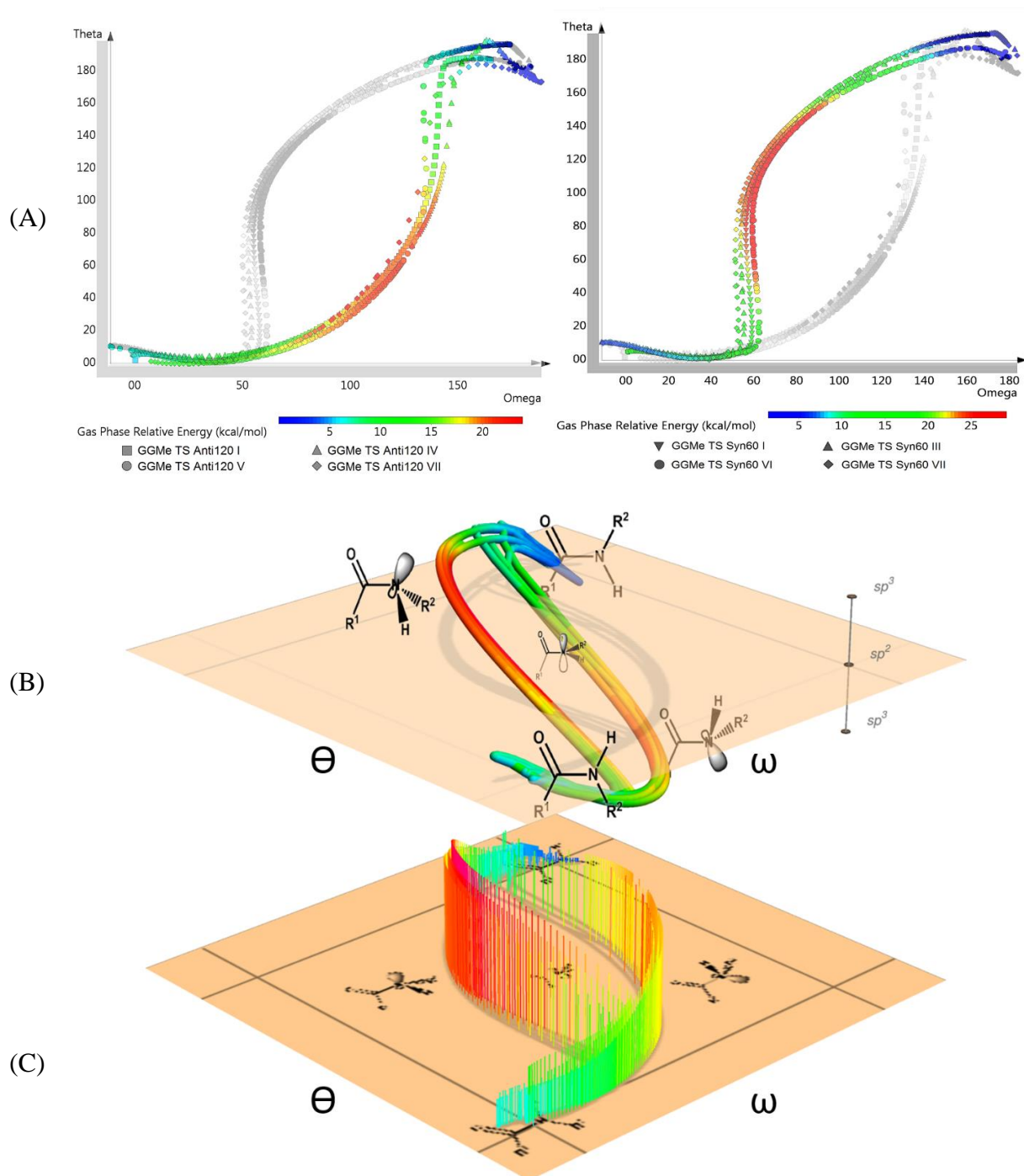
Figure 36. (A) Four path-1 (left) and four path-2 (right) IRC geometries generated from TS$_{anti}$ and TS$_{syn}$ rotamers of GGMe respectively are plotted in terms of ω and Θ dihedrals. In order to facilitate the comparison, the grey-scale path-2 IRC geometries have been superimposed on coloured path-1 IRC geometries, and vice versa. (B) All eight IRC paths are plotted in terms of ω, Θ and τ dihedrals, showing change in planarity of nitrogen during isomerization with clear inversion of nitrogen in Θ-phases of both paths. (C) The plot shows relative energy change for all IRC paths along with change in ω and Θ dihedrals.

Because the IRC path generation method follows only the two decreasing energy paths from the transition state, and because rotamer interconversion would usually involve an additional energy barrier, IRC generation from a specific TS rotamer will lead to a specific pair of trans and cis rotamers. The process identifies the geometries and reaction paths for isomerization for each GGMe rotamer, and so describes TS_{syn}, TS_{anti}, Pl1, Pl2 SPl1 and SPl2 for each specific rotamer. The set of isomerization paths thus generated represents a sampling of the full energy landscape with greatly simplified dimensionality, suitable for study of the range of variation created by the rotamer set.

In this study, eight of the fourteen rotamers generated for TS_{anti} and TS_{syn} (see Figure 36) were chosen for IRC analysis (four for each TS type). The analysis confirms that each transition state rotamer leads to a pair of trans and cis rotamers and path specific to the chosen TS rotamer. As illustrated in Figure 36, the overall geometry and corresponding energies of all rotamers differ significantly because of the conformational differences, but the IRC paths show very similar geometry changes within the amide moiety throughout the isomerization path.
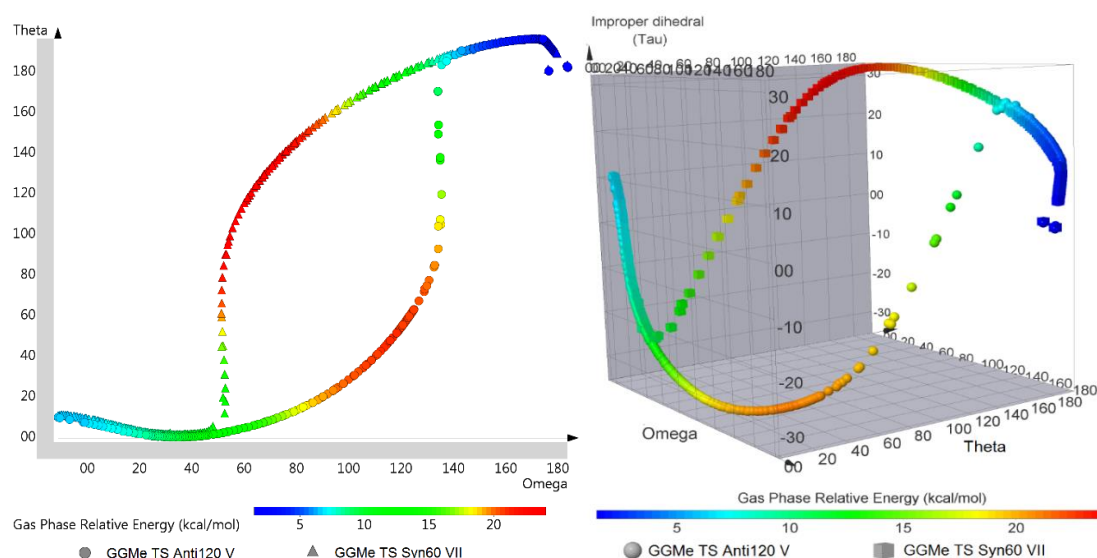
Figure 37. Two IRC paths connecting the same trans and cis rotamers of GGMe show both the IRC paths following the described isomerization scheme on peptidic systems.

Significantly, as shown in Figure 37, it was also observed that for a corresponding TS_{syn} (or TS_{anti}) rotamer pair, the IRC paths lead to identical pairs of trans and cis rotamers, replicating the full reaction path as seen for NMA. Thus, the GGMe set of IRC paths shows all the salient characteristics of the NMA path described above: a common path until a bifurcation point, splitting into ω- and Ө- phases leading to transition states, continuation

to another bifurcation point of the opposite isomer via Θ and ω phases on two separate paths, and completion of the isomerization via a common path.

9. Cis/trans isomerization in substituted GGMe derivatives

Based on the understanding of cis/trans isomerization in secondary amides in the previous chapter, this chapter now describes effects of substitution patterns on the energetics of minimum energy geometries as well as of cis/trans isomerization.

After detailed study of isomerization paths and geometry changes in the previous chapter, it was evident that a small prototype such as NMA has limited applicability to represent larger systems, especially with respect to structural factors and energetics, owing to existence of multiple rotamers and asymmetry around the molecular backbone in larger systems. On the other hand, a peptidic prototype such as GGMe can suitably represent the isomerization and the complexities arising out of a bigger molecular backbone, and can be used as the primary scaffold to generate models structures with different substitution patterns in order to assess and compare the steric effect of the substitution patterns.

As different side chains and substitutions have different extent of steric effect, the study of the role of steric effect becomes complex. Further, bigger and bulkier side chains and substitutions lead to more number of atoms and interactions with potentially more asymmetry and flexibility, and hence substantially higher CPU time for QM studies along with interference of other factors such as resonance and hydrogen bonding. Hence, such a study requires a representative substitution-group that can serve as a small and symmetrical yet bulky group, devoid of any resonance or hydrogen bonding effects. With the the A value > 4[194,195] and υ_{ef} value $= 1.2$[196,197], *tert*-butyl group can be as an ideal choice for the bulky substitution as it fits all these criteria. In synthetic chemistry, it is a common practice to use *tert*-butyl group as a bulky substitution to analyse the effect of steric bulk.[198–200]

9.1. Model structures
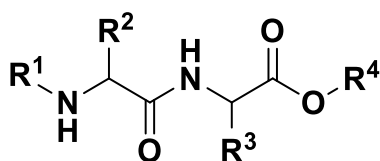
Figure 38. Model structure and the substitution positions

As shown in Figure 38, there are four available positions for substitutions. Among the four, two are alpha carbons of amino acids and are thus chiral centers. In line with natural amino acids, the default configuration of substitution at alpha carbons was kept as the S-configuration. However, when both chiral centers had substitutions (i.e. neither R^2 nor R^3 was hydrogen), the R^2 configuration was kept constant (S), and both isomers with different configurations (R and S) of R^3 were included. Hence, along with compounds with the natural S-configuration, two compounds with unnatural R-configuration at the alpha carbon were also included. R^4 forms a distant moiety, which would not be a part of "conventional" dipeptide structure, and may not be expected to exert significant steric effect on the peptide structure, so only one variation of R^4 was included. Thus, a total of 11 model structures with different substitution patterns of *tert*-butyl groups were generated as described in Table 21.

Table 21. Model structures and their substitutions

Compound	R^1	R^2	R^3	R^4
N0000 (GGMe)	H	H	H	Me
N0001	H	H	H	*t*-Bu
N0010	H	H	*S-t*-Bu	Me
N0100	H	*S-t*-Bu	H	Me
N0110	H	*S-t*-Bu	*S-t*-Bu	Me
N0120	H	*S-t*-Bu	*R-t*-Bu	Me
N1000	*t*-Bu	H	H	Me
N1010	*t*-Bu	H	*S-t*-Bu	Me
N1100	*t*-Bu	*S-t*-Bu	H	Me
N1110	*t*-Bu	*S-t*-Bu	*S-t*-Bu	Me
N1120	*t*-Bu	*S-t*-Bu	*R-t*-Bu	Me

9.2. Comparison of trans isomers using additive principle

In order to assess the magnitude of steric factors in terms of energy (^{st}E), the energy values calculated based on additive principle[194] for *tert*-butyl substituted structures (^{cal}E) were compared with their direct energy estimates for their minimum energy geometries (^{opt}E).
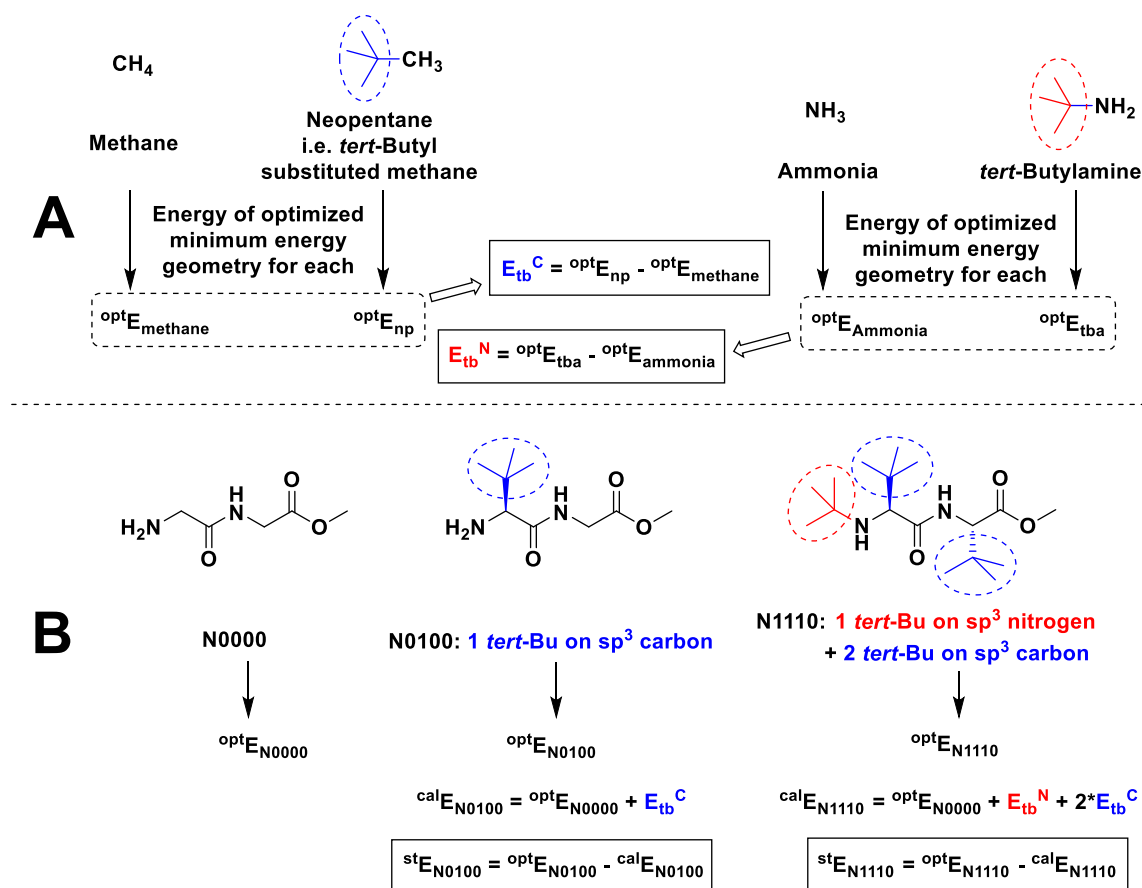
Figure 39. Using additive principle to assess the magnitude of steric bulk in terms of energy explained with examples of N0100 and N1110.

Table 22. Calculation of *tert*-Bu additive energy in the gas phase

Compound	^{opt}E	Attachment point	*tert*-Bu additive energy kcal/mol (E_{tb})
NH_3	-35508.975	sp^3 nitrogen	E_{tb}^N = -98709.131
tert-Bu-NH_2	-134218.106		
CH_4	-25437.531	sp^3 carbon	E_{tb}^C = -98712.134
tert-Bu-CH_3	-124149.665		

Table 23. Calculation of *tert*-Bu additive energy in the water phase

Compound	^{opt}E	Attachment point	*tert*-Bu additive energy kcal/mol (E_{tb})
NH_3	-35514.577	sp^3 nitrogen	E_{tb}^N = -98707.646
tert-Bu-NH_2	-134222.223		
CH_4	-25436.175	sp^3 carbon	E_{tb}^C = -98711.962
tert-Bu-CH_3	-124148.137		

As explained in Figure 39A, the energy difference between methane and neopentane was taken to represent the additive energy value of *tert*-butyl substitution on an sp^3-carbon (E_{tb}^C) while the energy difference between ammonia and tert-butylamine was taken to

represent the additive energy value of *tert*-butyl substitution on pyramidal sp^3-nitrogen of amino group (E_{tb}^N). Both values were calculated for both gas and water phases (Table 22 and Table 23). Then, as described in Figure 39B, energy value for optimized minimum energy geometry of N0000 (GGMe) i.e. $^{opt}E_{N0000}$ was taken as the basis for the calculations of other compounds having *tert*-butyl substitutions on alpha carbon or amino nitrogen. For each compound, the additive energy value (^{cal}E) was determined by addition of corresponding additive energy values of *tert*-butyl group (E_{tb}) – depending on the atom of attachment (carbon or nitrogen), number of *tert*-butyl groups and the phase (gas or water). The steric effect (^{st}E) was calculated as the difference between ^{cal}E and the energy values for optimized minimum energy geometries of respective compounds (^{opt}E) in respective phase i.e. gas or water phases (Table 24 and Table 25).

Table 24. Calculation of steric energy in gas phase

Compound	^{opt}E kcal/mol	No. of *t*-Bu- on sp^3 Nitrogen	No. of *t*-Bu on sp^3 carbon	^{cal}E kcal/mol	^{st}E kcal/mol
N0000t	-333806.071			-333806.071	0
N0010t	-432516.46		1	-432518.205	1.745
N0100t	-432515.901		1	-432518.205	2.304
N0110t	-531225.709		2	-531230.339	4.63
N0120t	-531225.494		2	-531230.339	4.845
N1000t	-432514.375	1		-432515.202	0.827
N1010t	-531224.575	1	1	-531227.336	2.761
N1100t	-531222.518	1	1	-531227.336	4.818
N1110t	-629932.315	1	2	-629939.47	7.155
N1120t	-629931.92	1	2	-629939.47	7.55

Table 25. Calculation of steric energy in water phase

Compound	Solvation Energy (kcal/mol)	^{opt}E kcal/mol	No. of *t*-Bu- on sp^3 Nitrogen	No. of *t*-Bu on sp^3 carbon	^{cal}E kcal/mol	^{st}E kcal/mol
N0000t	-16.33	-381811.506			-381811.506	0
N0010t	-16.85	-480519.643		1	-480523.468	3.825
N0100t	-14.86	-480518.077		1	-480523.468	5.391
N0110t	-16.1	-579224.507		2	-579235.43	10.923
N0120t	-12.66	-579225.518		2	-579235.43	9.912
N1000t	-11.16	-480518.478	1		-480519.152	0.674
N1010t	-15.97	-579226.192	1	1	-579231.114	4.922
N1100t	-13.47	-579222.324	1	1	-579231.114	8.79
N1110t	-11.49	-677932.254	1	2	-677943.076	10.822
N1120t	-11.34	-677929.486	1	2	-677943.076	13.59
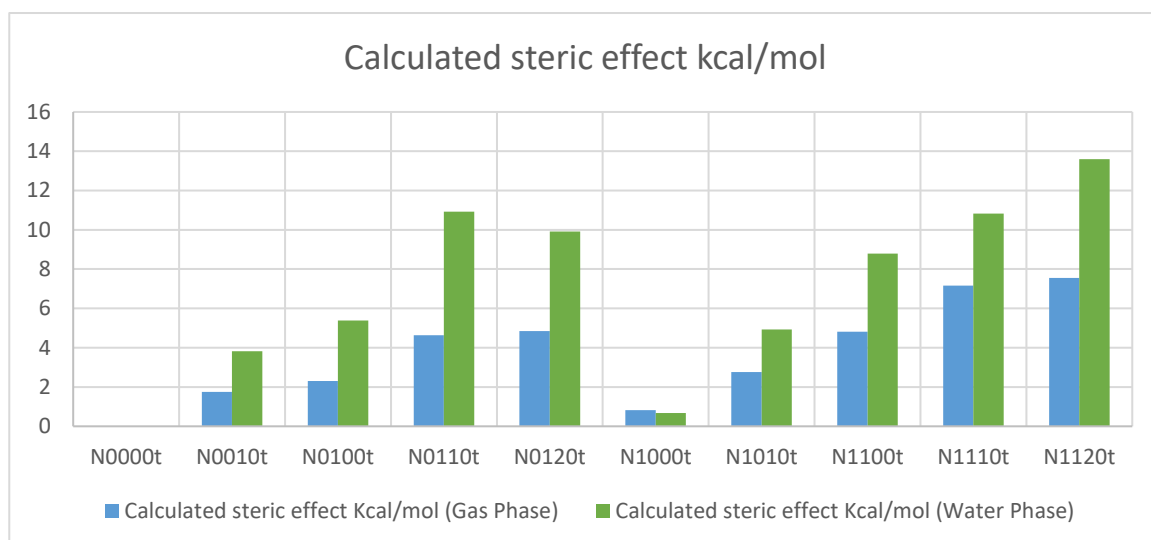
Figure 40. Calculated steric effect for trans-isomers in gas phase and water phase using additive principle

The trend, as shown in Figure 40, suggests that adding a bulky substitution to an alpha carbon results in a considerable steric effect. The bulky substitution on the N-terminal alpha carbon has slightly larger impact than that on the C-terminal alpha carbon. A bulky substitution on the N-terminal amino nitrogen also leads to a steric effect that is consistently present across corresponding pairs, becoming stronger with bulk on alpha carbons. This steric effect seen in terms of energies is also reflected in strained geometries, and is stronger when using the polarizable continuum model (the water phase).

9.3. Comparison of cis isomers

9.3.1. Relative energy compared to minimum energy trans-geometry

Table 26 and Table 27 describe the relative energies of minimum energy cis geometries in gas and water phases respectively, along with corresponding equilibrium constant estimates. Interestingly, the relative energy values of minimum energy geometries of most cis isomers (compared to minimum energy geometry of trans isomers) were found to be higher in the gas phase than in water phase (Figure 41). This may be attributed partially to the fact that the trans-isomers in water phase already have much higher energy than corresponding gas phase structures as revealed in steric factor calculations with additive principles for trans isomers. Further, the solvation energy values for most cis isomers are higher (i.e. more negative) than corresponding trans isomers by 2-3 kcal/mol. This indicates better stabilization/solvation of cis isomers than trans isomers in the water phase,

which may, in turn, be attributed to the higher dipole moment of cis isomers compared to trans isomers, as evident from Figure 42.

Table 26. Relative energies of minimum energy cis-isomer in gas phase

Compound	Relative energy[a]	ΔG^{b}	Equilibrium constant $k_c{}^{c}$
N0000c	4.63	4.51	4.91E-04
N0001c	4.31	4.24	7.83E-04
N0010c	5.23	4.78	3.16E-04
N0100c	5.12	4.76	3.22E-04
N0110c	6.17	5.82	5.42E-05
N0120c	4.96	4.76	3.22E-04
N1000c	5.02	5.25	1.41E-04
N1010c	5.67	5.78	5.84E-05
N1100c	5.17	4.92	2.49E-04
N1110c	7.52	7.62	2.61E-06
N1120c	7.52	7.65	2.48E-06

[a] Gas phase energy relative to the minimum energy trans geometry, calculated in kcal/mol at B3LYP/6-311G-3DF-3PD++ level.
[b] Relative free energy at 298.15 K relative to minimum energy trans geometry, calculated in kcal/mol at B3LYP/CC-PVTZ(-F)++ level.
[c] Equilibrium constant estimated at 298.15 K based on ΔG.

Table 27. Relative energies of minimum energy cis-isomer in water phase

Compound	Solvation Energy (kcal/mol)	Relative energy[a]	ΔG^{b}	Equilibrium constant $k_c{}^{c}$
N0001c	-18.53	1.881	1.98	3.54E-02
N0000c	-19.43	2.035	2.01	3.36E-02
N0010c	-17.85	2.967	2.73	9.96E-03
N0100c	-16.88	2.107	2.23	2.33E-02
N0110c	-15.67	1.385	1.66	6.08E-02
N0120c	-14.7	2.797	2.18	2.53E-02
N1000c	-19.86	2.788	2.57	1.30E-02
N1010c	-18.63	3.345	3.14	4.95E-03
N1100c	-14.35	3.919	3.69	1.96E-03
N1110c	-11.87	7.666	7.88	1.68E-06
N1120c	-11.76	4.86	4.94	2.39E-04

[a] Water phase energy relative to the minimum energy trans geometry, calculated in kcal/mol at B3LYP/6-311G-3DF-3PD++ level.
[b] Relative free energy at 298.15 K relative to minimum energy trans geometry, calculated in kcal/mol at B3LYP/CC-PVTZ(-F)++ level.
[c] Equilibrium constant estimated at 298.15 K based on ΔG.
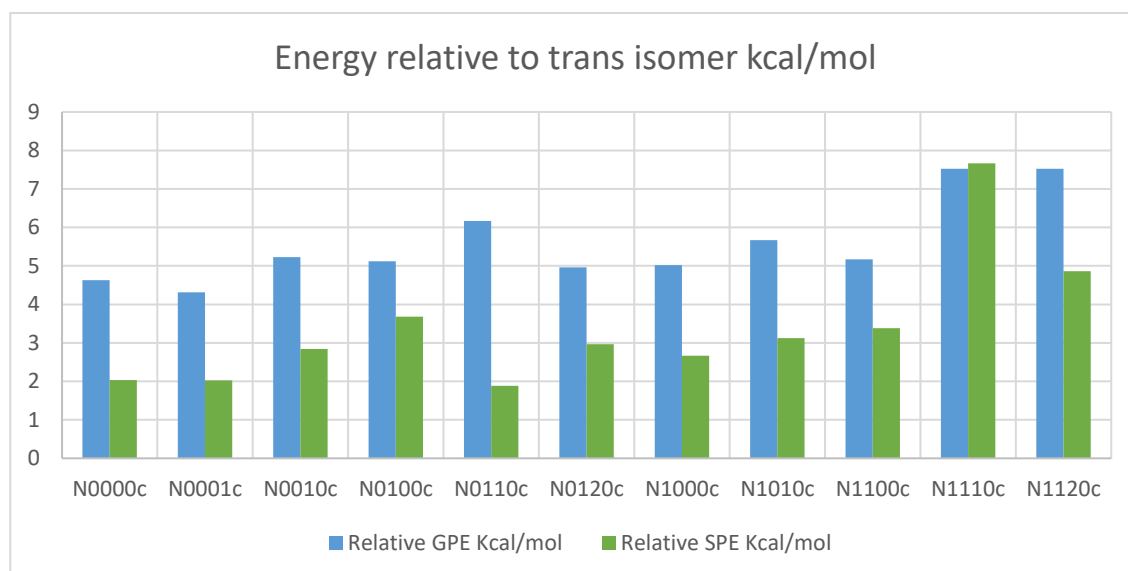
Figure 41. Energies of minimum energy geometries of cis isomers relative to that of trans-isomers
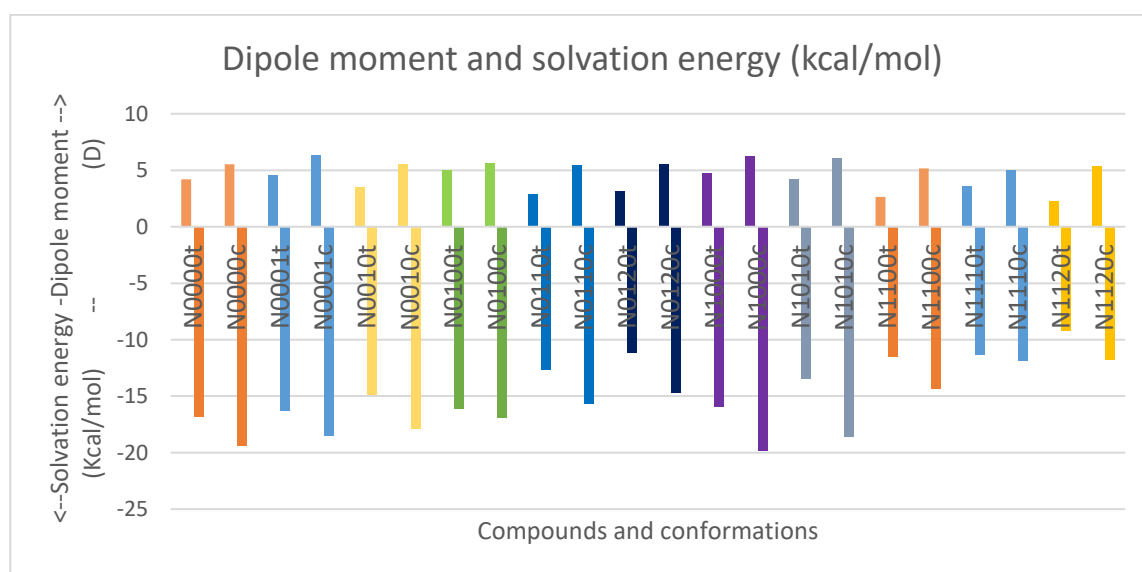
Figure 42. Comparison of dipole moment and solvation energy for trans and cis isomers

9.3.2. Steric factor effect calculation

Table 28 and Table 29 describe the calculation of steric effect on cis-isomers based on the additive principle in gas and water phases respectively. The steric effect calculation based on the additive principle shows a trend similar to that for the trans-isomers (Figure 43) as discussed in section 9.2.

Table 28. Steric factor effect on the cis-isomer based on the additive principle in the gas phase

Compound	^{opt}E kcal/mol	No. of t-Bu- on sp^3 Nitrogen	No. of t-Bu on sp^3 carbon	^{cal}E kcal/mol	^{st}E kcal/mol
N0000c	-333801.441			-333801.441	0
N0010c	-432511.233		1	-432513.575	2.342
N0100c	-432510.785		1	-432513.575	2.79
N0110c	-531219.543		2	-531225.709	6.166
N0120c	-531220.53		2	-531225.709	5.179
N1000c	-432509.352	1		-432510.572	1.22
N1010c	-531218.91	1	1	-531222.706	3.796
N1100c	-531217.348	1	1	-531222.706	5.358
N1110c	-629924.792	1	2	-629934.84	10.048
N1120c	-629924.4	1	2	-629934.84	10.44

Table 29. Steric factor effect on the cis-isomer based on the additive principle in the water phase

Compound	^{opt}E kcal/mol	No. of t-Bu- on sp^3 Nitrogen	No. of t-Bu on sp^3 carbon	^{cal}E kcal/mol	^{st}E kcal/mol
N0000c	-333819.556			-333819.556	0
N0010c	-432526.761		1	-432531.518	4.757
N0100c	-432526.055		1	-432531.518	5.463
N0110c	-531233.207		2	-531243.48	10.273
N0120c	-531232.806		2	-531243.48	10.674
N1000c	-432525.775	1		-432527.202	1.427
N1010c	-531232.932	1	1	-531239.164	6.232
N1100c	-531228.49	1	1	-531239.164	10.674
N1110c	-629934.673	1	2	-629951.126	16.453
N1120c	-629934.711	1	2	-629951.126	16.415
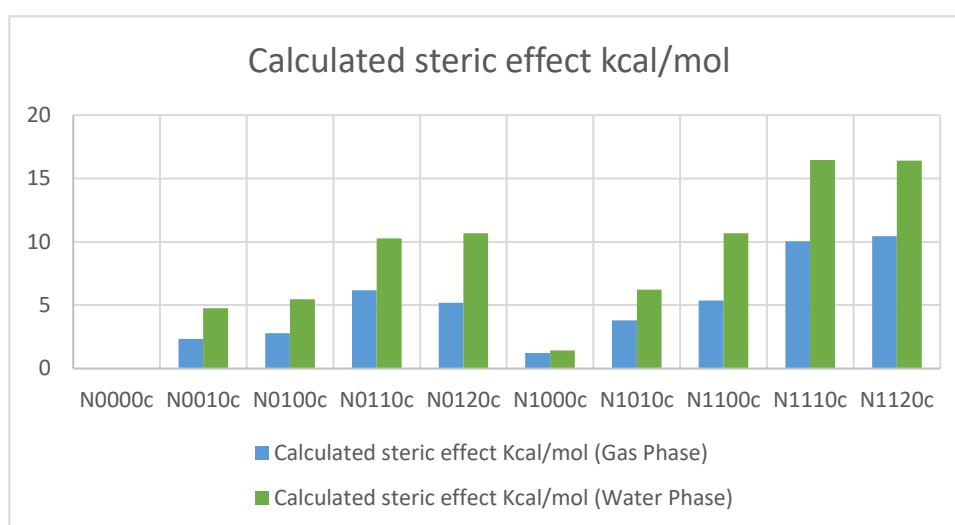
Figure 43. Steric effect for cis-isomers in the gas and water phases using the additive principle

9.4. Energy barrier comparison

The calculation of energy barriers for substituted derivatives of GGMe is a complex task. As we have seen in case of GGMe (chapter 8), the existence of multiple rotamers makes it difficult to identify every possible transition state. In case of GGMe, however, the lack of substitutions on α-carbons still simplifies the calculation significantly, as the isomerization study along the rotation coordinate ω between 180º and 0º suffices due to the symmetry. This becomes more complex with the introduction of substitutions and stereospecificity at α-carbons, which introduces asymmetry along the rotation coordinates, such as the ω-dihedral, resulting in different energy barriers for rotation along the positive vs negative directions of rotation (see section 8.2.2), with simultaneous existence of multiple rotamers. Therefore, the number of possible transition states conformations increases enormously.

In order to simplify this complex problem, it was necessary to use a specific method to enable comparison among same type of geometries for different compounds. Such a comparison can provide information about the overall effect of substitution patterns, if not for specific conformers. From the examples of NMA and GGMe, it became evident (see Table 17) that the energy barrier geometries obtained with step wise RCS provides a reasonably accurate estimation of transition states and corresponding energies. Therefore, we decided to find energy barrier geometries (EBGs) to estimate energy barrier values. For compounds without substitutions on α-carbon, the calculation in one direction (between 0º and 180º, or between 180º (= -180º) and 360º (= 0º) was sufficient. However, for other substituted compounds, calculation in both directions was carried out. Correspondingly, from trans isomers, *syn* EBGs were obtained close to ω = ±60º and from cis isomers, *anti* EBGs were obtained close to ω = ±120º.

Table 30 and Table 31 describe energy barrier values corresponding to *syn/anti* EBGs in gas phase and water phase respectively. It is evident that the energy barrier value strongly depends on the direction of rotation for compounds with substitution on chiral α-carbons, and in such cases, the energy difference between same types of energy barrier geometries in two different directions of rotation can be significant, from 1 to 6 kcal/mol.

Table 30. Energy barrier values corresponding to *syn* and *anti* EBGs generated from optimized minimum energy trans and cis geometries, respectively, in the gas phase.

Cpd	Anti120GP		Anti-120GP		Syn60GP		Syn-60GP		ΔG_{eff}	Rate constant k^c
	ΔE^a	ΔG^b	ΔE^a	ΔG^b	ΔE^a	ΔG^b	ΔE^a	ΔG^b		
N0000	19.53	17.46	19.53	17.45	25.46	23.53	25.26	23.56	17.45	1.01E+00
N0001	19.45	17.40	19.45	17.38			25.37	23.47	17.38	1.13E+00
N0010	24.22	22.48	19.34	17.42	25.22	23.75	27.99	25.85	17.42	1.06E+00
N0100	20.96	18.38	20.51	18.15	26.30	24.36	25.26	22.94	18.15	3.08E-01
N0110	25.64	23.98	19.76	18.02	23.98	22.37	28.18	26.43	18.02	3.84E-01
N0120	19.99	17.73	24.92	23.00	27.33	25.13	29.92	28.30	17.73	6.24E-01
N1000	19.55	18.01	19.82	18.27			25.09	22.61	18.01	3.89E-01
N1010	24.33	23.44	20.10	18.72	24.95	23.68	27.80	26.05	18.72	1.18E-01
N1100	21.54	19.17	20.59	18.58	26.61	24.35	25.40	23.64	18.58	1.48E-01
N1110	26.08	24.54	20.15	18.60	28.29	27.29	34.39	32.74	18.60	1.45E-01
N1120	21.43	19.36	26.18	23.42	30.79	29.04	23.68	22.48	19.36	3.98E-02

[a] Gas phase energy relative to the minimum energy trans geometry, calculated in kcal/mol at B3LYP/6-311G-3DF-3PD++ level.
[b] Relative free energy at 298.15 K relative to minimum energy trans geometry, calculated in kcal/mol at B3LYP/CC-PVTZ(-F)++ level.
[c] Reaction rate constant estimated with Eyring equation from ΔG_{eff}.

Table 31. Energy barrier values corresponding to *syn* and *anti* EBGs generated from optimized minimum energy trans and cis geometries, respectively, in the water phase

Cpd	Anti120WP		Anti-120WP		Syn60WP		Syn-60WP		ΔG_{eff}	Rate constant k^c
	ΔE^a	ΔG^b	ΔE^a	ΔG^b	ΔE^a	ΔG^b	ΔE^a	ΔG^b		
N0000	19.18	16.81	19.18	16.83	24.29	21.70			16.81	2.95E+00
N0001	19.27	17.14	19.22	17.14			26.50	24.31	17.14	1.69E+00
N0010	25.40	23.65	19.03	17.08	29.42	27.05	24.23	21.98	17.08	1.86E+00
N0100	22.45	19.95	20.66	18.68	23.31	21.13	26.09	24.04	18.68	1.25E-01
N0110	25.13	23.41	19.32	17.84	22.23	20.59	20.47	18.76	17.84	5.23E-01
N0120	26.44	24.12	26.55	24.23	23.39	20.82	26.51	24.48	20.82	3.38E-03
N1000	22.30	19.56	22.35	19.68			23.86	21.02	19.56	2.85E-02
N1010	26.60	24.56	19.00	17.29	29.60	27.45	24.29	22.06	17.29	1.31E+00
N1100	22.54	19.97	23.71	21.35	26.50	24.03	23.04	21.02	19.97	1.43E-02
N1110	29.46	27.62	24.53	23.10	29.24	28.14	30.37	28.81	23.10	7.25E-05
N1120	23.15	21.02	27.71	25.62	26.65	24.73	23.78	22.37	21.02	2.42E-03

[a] Water phase energy relative to the minimum energy trans geometry, calculated in kcal/mol at B3LYP/6-311G-3DF-3PD++ level.
[b] Relative free energy at 298.15 K relative to minimum energy trans geometry, calculated in kcal/mol at B3LYP/CC-PVTZ(-F)++ level.
[c] Reaction rate constant estimated with Eyring equation from ΔG_{eff}.

However, considering the minimum value energy barrier as the "real" barrier, it appears that bulky substitution patterns only marginally affect the energy barrier values in gas phase, as the most substituted compounds N1110 and N1120 show an increase in the

energy barrier of just 2 kcal/mol compared to N0000 (GGMe). On the other hand, the effect of bulky substitution patterns becomes significant in the water phase, and can show an increase in energy barrier by up to 5 kcal/mol. Interestingly, a *syn* EBG was found to be more stable than both *anti* EBGs in case of compound N0120 in water phase.

Overall, bulky substitutions affect cis/trans isomerization in two ways: by making the cis isomer more unstable and thereby shifting the equilibrium towards the trans isomer, and by increasing the reaction barrier height and thereby decreasing the reaction rate. As both effects discourage trans to cis isomerization, compounds with bulky substitutions are expected to take longer and/or require higher temperatures for the same. Moreover, as trans to cis isomerization is an essential step in cyclization of dipeptides to piperazine-2,5-diones, it is certain that the slow isomerization would also result in decelerated dipeptide cyclization.

10.   Cyclization of dipeptide esters to piperazine-2,5-diones

10.1. Water catalyzed cyclization mechanism

When the amide bond of the dipeptide ester acquires a cis orientation, the nucleophilic nitrogen of the N-terminal amino group can come in proximity of the electrophilic carbonyl carbon of the C-terminal ester group. The reaction then occurs in two steps. In the first step, the nucleophilic amino group attacks the electrophilic carbonyl carbon of the ester group and generates an alcohol intermediate. In the second step, the alkoxy part of the ester group leaves as a corresponding alcohol. For synthesis of our biofocussed chemoprospecting library of piperazine-2,5-diones, the cyclization of dipeptide esters (see section 6.1.2) was carried out using water as a solvent, without any other reagent.

As described by Xia et al.[201], it is likely that water plays an important role as a catalyst in the cyclization, as explained in Scheme 16. Both steps of the reaction occur via two different six-membered transition states with direct involvement of a water molecule. In the first step, a water molecule accepts a hydrogen atom from amino nitrogen, and donates a hydrogen atom to the carbonyl oxygen, creating an alcohol intermediate. In the second step, a water molecule accepts a hydrogen atom from the carbonyl carbon and donates a hydrogen atom to the leaving alkoxy group to generate corresponding alcohol and product piperazine-2,5-dione. During the course of the reaction, the carbonyl carbon changes from sp^2 to sp^3 hybridized state, resulting in a tetrahedral geometry in the first step, and regains sp^2 hybridized state in the second step. The authors observed that the first step is rate

determining, with its transition state (TS1) having the highest energy for the complete cyclization reaction. Further, the cyclization reaction up to TS1 is reversible, capable of regenerating the trans isomer of *N*-substituted dipeptide ester. However, once the reaction intermediate (INTMDT) is formed, its stabilization is achieved through removal of leaving group (alkoxy part of ester group) and formation of piperazine-2,5-diones via the second transition state (TS2). TS2 is more stable than TS1, and does not affect the reaction rate significantly. Therefore, for analysis of substitution effects on reaction, it is sufficient to perform theoretical studies on TS1 only.
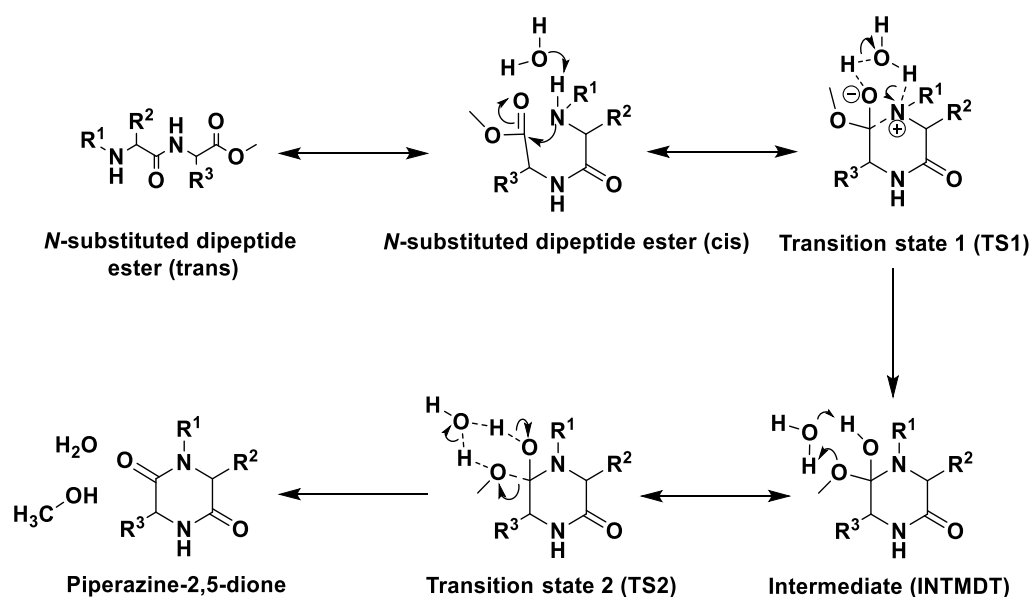
Scheme 16. Water catalyzed cyclization of a dipeptide ester to a piperazine-2,5-dione

10.2. Model structures
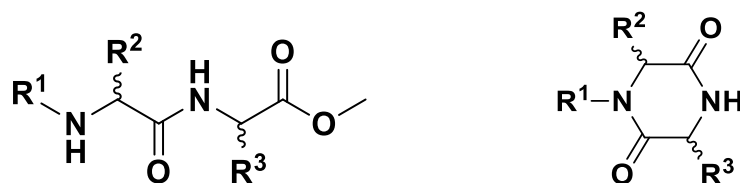
Figure 44. Dipeptide ester and corresponding piperazine-2,5-dione scaffolds with corresponding substitution positions

With intramolecular cyclization, the dipeptide ester scaffold cyclizes to piperazine-2,5-dione scaffold (Figure 44). In order to study the cyclization reaction, the model structures used for trans-cis isomerization studies were used along with their analogues with *tert*-Bu groups as methyl group replacements. The pattern of substituents was also kept same. Thus,

a total of 19 model structures with different substitution patterns were generated, as described in Table 32.

Table 32. Model structures and their substitutions

Dipeptide ester compound	Piperazine-2,5-dione compound	R^1	R^2	R^3
N0000	D000	H	H	H
N0010	D001	H	H	*S-t*-Bu
N0100	D010	H	*S-t*-Bu	H
N0110	D011	H	*S-t*-Bu	*S-t*-Bu
N0120	D012	H	*S-t*-Bu	*R-t*-Bu
N1000	D100	*t*-Bu	H	H
N1010	D101	*t*-Bu	H	*S-t-Bu*
N1100	D110	*t*-Bu	*S-t*-Bu	H
N0030	D003	H	H	*S*-Me
N0300	D030	H	*S*-Me	H
N0330	D033	H	*S*-Me	*S*-Me
N0340	D034	H	*S*-Me	*R*-Me
N3000	D300	Me	H	H
N3030	D303	Me	H	*S*-Me
N3300	D330	Me	*S*-Me	H
N3330	D333	Me	*S*-Me	*S*-Me
N3340	D334	Me	*S*-Me	*R*-Me
N1110	D111	*t*-Bu	*S-t*-Bu	*S-t*-Bu
N1120	D112	*t*-Bu	*S-t*-Bu	*R-t*-Bu

As the reaction site of the rate-determining step (C-N bond formation) is "crowded" with local substitutions and a water molecule, we expected that methyl substitutions already should be sufficient to demonstrate the role of steric factor, requiring significantly less computational resources compared to *tert*-butyl group calculations. Therefore, while the geometries for trans, cis isomer, reaction intermediate and corresponding piprazine-2,5-dione were optimized for all 19 model structures, the transition state search and reaction coordinate studies were performed mostly with model structures having methyl group substitutions, and only two model structures with *tert*-butyl substitutions were used. For the synthesized compounds, the steric bulk size of most of the amino acid side chains is intermediate, between the bulk of methyl and *tert*-butyl groups, hence the observations for model structures from this study can be correlated to synthesized compounds. As the reaction occurs in water phase, all the calculations were performed for water phase only.

Table 33. Energetics of water catalyzed cyclization for dipeptide esters

NSD	DKP	Cis - Relative energy[a]	ΔG_{cis}[b] kcal/mol	TS1 - Relative Energy[a]	ΔG_{TS1}[b] kcal/mol	INTMDT - Relative Energy[a]	ΔG_{INT}[b] kcal/mol	DKP - Relative Energy[a]	ΔG_{DKP}[b] kcal/mol
N0000	D000	2.03	1.98	15.90	41.10	11.70	15.02	-5.50	-15.09
N0030	D003	2.03	1.44	17.15	37.42	11.58	15.61	-6.30	-16.06
N0300	D030	2.84	2.61	17.73	43.84	12.67	17.04	-5.17	-14.90
N0330	D033	3.59	3.34	18.94	44.49	12.14	16.47	-6.39	-16.06
N0340	D034	3.68	3.46	17.48	43.74	13.70	17.85	-5.75	-15.51
N0010	D001	2.97	2.73			16.19	20.63	-6.37	-15.74
N0100	D010	2.11	2.23			12.19	16.66	-7.97	-17.00
N0110	D011	1.39	1.66			16.54	21.29	-6.94	-16.03
N0120	D012	2.80	2.18			18.89	22.57	-6.25	-16.04
N3000	D300	2.67	2.10	18.29	43.31	13.42	17.15	-6.05	-16.08
N3030	D303	3.13	2.90	17.93	43.56	13.80	17.82	-6.91	-17.30
N3300	D330	3.38	2.79	19.56	44.63	15.42	19.23	-6.36	-16.65
N3330	D333	4.17	3.72	22.09 (S)	47.49	16.24	20.30	-6.58	-16.87
				22.75 (R)	47.98	15.81	19.65		
N3340	D334	2.82	2.47	20.88 (S)	46.33	15.80	19.85	-6.97	-17.51
				22.95 (R)	43.06	17.20	21.22		
N1000	D100	2.79	2.57			26.06	29.67	1.69	-8.66
N1010	D101	3.35	3.14			32.07	36.55	1.71	-8.62
N1100	D110	3.92	3.69			30.35	34.31	4.07	-6.74
N1110	D111	7.67	7.88	52.84 (S)	79.54	42.28	47.23	15.16	4.55
				48.14 (R)	74.54	43.67	49.14		
N1120	D112	4.86	4.94	45.28 (S)	71.98	40.89	45.84	10.62	0.41
				58.16 (R)	86.14	38.98	43.59		

[a] Water phase energy relative to the minimum energy trans geometry, calculated in kcal/mol at B3LYP/6-311G-3DF-3PD++ level.

[b] Relative free energy at 298.15 K relative to minimum energy trans geometry, calculated in kcal/mol at B3LYP/CC-PVTZ(-F)++ level.

10.3. Cyclization energetics

The relative energies of relevant geometries for water catalyzed cyclization are shown in Table 33. It is evident that even a small moiety such as the methyl group causes a significant steric effect, destabilizing the geometries, especially transition state TS1. This destabilization is observed to become stronger with increasing number of substitutions. From D000 (unsubstituted piperazine-2,5-dione) to D333 (all 3 methyl substitutions, with two *S*-methyl), the cis, TS1 and intermediate (INTMDT) geometries are further destabilized, by 2, 6 and 5 kcal/mol respectively, discouraging the cyclization reaction at each transformation step. For the structure with triply substituted *tert*-butyl group, the TS1

energy barrier values were found to be unrealistically high, at more than 45 kcal/mol, demonstrating the extent of destabilization with bigger and bulkier substitutions.

Overall, the theoretical studies clearly demonstrate the strong effect of steric factor for the cyclization reaction. These results were correlated with the experimental results obtained for DKP library synthesis.

11.   Rationalization of experimental results

As described in section 6.1.2, the experimental results had shown a strong steric effect on two reaction parameters: yield and reaction time. The theoretical studies of both stages, cis/trans isomerization and cyclization to piperazine-2,5-dione also demonstrated that increasing bulk of substitution exerts steric effect in both reaction stages.

Because the theoretical studies were carried out on model compounds, it was necessary to categorize the library synthesis reactions according to the most relevant model structures. The model compounds subjected to theoretical studies had two type of substitutions: methyl and *tert*-butyl, and most of the substitutions in the synthesized compounds are moderate in terms of bulk; often greater than methyl, but smaller than *tert*-butyl. Therefore, for the calculation, the categorization was done using following criteria.

1) All compounds had substitution on one nitrogen. All substitutions on nitrogen were greater than methyl, but still not as bulky as *tert*-butyl, so only D3xx categories were considered.

2) The compounds with at least one hydrogen substitution were categorized into relevant D3x0 or D30x categories depending on the other substitutions.

3) The compounds with at least one methyl substitution were categorized into relevant D330 or D303 or D333, depending on the other substitutions.

4) The compounds with all substitutions bigger than hydrogen or methyl were classified in a new category D333+, indicating that all substitutions are greater than methyl.

Using this categorization protocol, library synthesis reactions were categorized in 5 categories as shown in Table 34. For each category, average reaction time and yield were calculated.

Table 34. Categorization of synthesized compounds for correlation

R¹	R²-amino acid	R³-amino acid methyl ester	Time (min)	Isolated Yield %	Relevant comparison	Avg time	Avg yield %
(pyridin-3-yl)	Glycine	Glycine	20	78	D300	25	75
(pyridin-4-yl)	Glycine	Glycine	30	72	D300		
(pyridin-4-yl)	Glycine	D-Tryptophan	30	65	D303	30	61.6
(pyridin-3-yl)	Glycine	L-Tryptophan	30	61	D303		
(pyridin-4-yl)	Glycine	L-Tyrosine	30	59	D303		
(pyridin-3-yl)	L-Alanine	Glycine	30	76	D330	42	59.5
(pyridin-3-yl)	L-Phenyl alanine	Glycine	30	60	D330		
(pyridin-4-yl)	L-Phenyl alanine	Glycine	60	55	D330		
(furan-2-yl)	L-Phenyl alanine	Glycine	30	49	D330		
(3-hydroxyphenyl)	L-Phenyl alanine	Glycine	60	59	D330		
(pyridin-3-yl)	L-Phenyl alanine	L-Alanine	60	56	D333	130	53.6
(thiophen-2-yl)	L-Phenyl alanine	L-Alanine	180	64	D333		
(furan-2-yl)	L-Phenyl alanine	L-Alanine	60	46	D333		
(3-hydroxyphenyl)	L-Phenyl alanine	L-Alanine	90	58	D333		
(pyridin-3-yl)	L-Alanine	L-Tryptophan	180	47	D333		
(pyridin-3-yl)	L-Alanine	L-Tyrosine	210	51	D333		
(3-hydroxyphenyl)	L-Phenyl alanine	L-Tryptophan	120	55	D333+	140	47.6
(furan-2-yl)	L-Phenyl alanine	L-Tyrosine	180	37	D333+		
(3-hydroxyphenyl)	L-Phenyl alanine	L-Tyrosine	120	51	D333+		

Table 35. Correlation of theoretcal energetics with reaction time and yield

Title	Title	Cis - Relative SPE Kcal/mol	TS1 - Relative Energy	INTMDT - Relative Energy	DKP - Relative Energy	Average time	Average yield %
N3000	D300	2.669	18.294	13.418	-6.052	25	75
N3030	D303	3.126	17.93	13.804	-6.909	30	61.6
N3300	D330	3.383	19.564	15.424	-6.359	42	59.5
N3330	D333	4.173	22.09 (S)	16.245	-6.579	130	53.6
			22.75 (R)	15.81			
	D333+					140	47.6
N1110	D111	7.666	52.84 (S)	42.281	15.16		
			48.14 (R)	43.67			

As shown in Table 35, the categorization of available library synthesis examples provided a clear trend of increasing substitution effect with more and bigger substitutions. It is evident that the substitution effect is reflected for both stages, cis/trans isomerization as well as cyclization, and it discourages the overall cyclization reaction in both stages. The high activation energies for each category also explains the need of carrying out the reaction at high temperature, and the increase in the required time. The decrease in the yield is secondary to the increase in the required time, and may be attributed to plausible side reactions such as hydrolysis and degradation of the ester with increase in time. Thus, the theoretical studies satisfactorily rationalize the strong substitution effects observed during synthesis of piperazine-2,5-dione library. From a larger perspective, such theoretical studies can provide valuable information on synthesis parameters for compound selection.

Concluding Discussion

12. Biofocussed chemoprospecting: synergizing best approaches in early phase drug discovery

As a part of this PhD research work, a "Biofocussed chemoprospecting" approach – a hybrid approach for design and synthesis of hit-finding libraries was developed. The essence of the approach is to use diverse, yet "bio-like" compounds for efficient hit-finding. Biofocussing enhances hit-rates; upfront chemical synthesis enables the choice of efficient protocols and diversification potential, and the screening approaches enable discovery of novel therapeutic targets and mechanisms. Integration of informatics and theoretical methods promise to aid the optimization of property profiles for chemical library design.

Congruent to the research goals described in chapter 4, three diverse libraries based on biomolecules such as linear and cyclic dipeptides, and tartaric acid were designed. Virtual libraries corresponding to the scaffolds were generated, and their physicochemical properties and drug likeness were analysed computationally.

This enabled property filtering and optimization of qualities such as diversity of physicochemical properties, drug likeness, ease of synthesis, low cost and "bio-likeness" for efficient selection of compounds. The compounds providing optimum diversity of scaffolds, substitutions and physicochemical properties were selected.

Using conventional or novel protocols, easy and short synthetic routes for the library synthesis were designed, optimized, and the libraries were synthesized.

The synthesized libraries were tested for various biological activities. Many of the synthesized compounds showed significant bioactivities on different targets based assays and cell based assays.

The *N*-substituted dipeptide esters showed the most significant bioactivities, including loss of viability on cancer cell lines as well as activities on opioid and neurokinin receptors and other enzymes such as COX-1,2 and prolyl oligopeptidase. While a primary SAR has been generated for the loss of cancer cell-viability, the specific drug target and mechanism remain to be identified conclusively. The primary results from SILAC studies have provided a good starting point in this direction. Considering that the normal ratio between MIC (minimum inhibitory concentration) for cell-based assays and assays on isolated target (such as enzymes) usually varies from 100 to 1000 fold, a nanomolar level activity may be expected for the target corresponding to the viability loss. The library also provides

an opportunity to develop novel molecules with selective activity on other targets such as opioid and neurokinin receptors. The multiple optimization opportunities imply significant minimization of the investment-risk for commercial partners. Thus, the next logical step is to secure the IPRs with a high value patent. The university has claimed the rights and patent filing process is in progress, which is expected to be complete before thesis defense.

The piperazine-2,5-diones have provided interesting hits for anti-cancer activity and BRSK1 enzyme inhibition. Though these hits could not be developed further within the constraints of this PhD research, they will be important starting points for future research.

With smaller libraries and screening, the tartaric acid bisamides and related compounds are yet to show significant activities. However, the compounds will be tested on broader screening platforms to explore the potential bioactivities.

In sharp contrast to conventional approaches, which use libraries comprising thousands of compounds synthesized at a very high cost, the libraries synthesized following the biofocussed chemprospecting approach provided hits from small libraries at cheap cost with hit-rates comparable to target-based approaches, with additional potential to identify novel drug targets and mechanisms.

In addition, the theoretical QM studies on cis/trans isomerization and cyclization of secondary amides using density functional theory were carried out. These studies highlighted the geometry changes, including the pyramidalization and inversion of nitrogen in cis/trans isomerization, and rationalized the observed substituent effects in cyclization reaction. Such studies can be integrated into future efforts to include synthetic efficiency as a parameter of library design.

The success of the approach can be attributed essentially to the <u>efficient library design</u> as an outcome of <u>focus on bio-likeness</u> and <u>optimized diversity</u> – the core ideas of the biofocussed chemoprospecting approach.

13.   Scientific impact and future directions

Overall, this PhD research work has been highly interdisciplinary in nature giving diverse medicinal chemistry experience, from computer aided drug design (CADD) and cheminformatics to library design and synthesis to bioactivity assays to proteomics analysis, and has also included theoretical QM studies. This has generated multiple opportunities for research publications.

Two manuscripts have been published:

1) The docking studies of known inhibitors of Rho kinase and PKA carried out as a part of this research (section 6.4.3) have already been published as a part of a publication.[202] (See appendix, **Paper 1**)

2) A publication describing the biofocussed chemoprospecting approach in detail with outline of bioactivities has been published recently.[203] (See appendix, **Paper 2**)

We expect to submit up to five more manuscripts for publication and potentially a patent from this PhD research work as following:

Three manuscripts describing theoretical chemistry studies:

1) Density Functional Studies on Cis/Trans Isomerization in Secondary Amides: Reaction Paths, Nitrogen Inversion and Relevance to Peptidic Systems (chapter 8) – Ready for submission. (See appendix, **Paper 3**)

2) Impact of bulky substitutions on minimum energy geometries and cis/trans isomerization of secondary amides (chapter 9).

3) Rationalization of strong steric effects observed in water catalyzed cyclization of dipeptide esters to piperazine-2,5-diones using density functional methods (chapter 10 and 11).

After a decision on patent filing for the *N*-substituted dipeptide ester series, the following two manuscripts can be submitted:

1) Anticancer activity of *N*-substituted dipeptide esters

2) Opioid and/or neurokinin receptor activity of *N*-substituted dipeptide esters (This will require a small study, which we hope to complete before thesis defense)

The research carried out on tartaric acid bisamides and related compounds has also created opportunities for two publications, although these require more work before submission:

1) One-pot synthesis of tartaric acid bisamides

2) One-pot synthesis of 2°-monoamides of tartaric acid and cyclic tartrimides, along with their hydrolysis products.

A good indicator of a successful research is how it inspires further research. We hope that it is evident that this PhD thesis has generated wide range of opportunities for further research in both synthetic chemistry and medicinal chemistry.

Experimental Section

14. Computational chemistry

14.1. Computational chemistry – General information

The structure of compounds were generated using MarvinSketch module of JChem suite (ChemAxon, Budapest, Hungary), ChemSketch software (ACD labs, Toronto, Ontario, Canada) or ChemDraw module of ChemBioOffice (Perkin Elmer, Massachusetts, USA). The virtual combinatorial libraries were generated using Markush enumerator plugin of JChem suite. Docking and related studies such as protein preparation, ligand preparation, homology modelling were performed in Maestro module[204] of Schrödinger software suite (Schrödinger, LLC. New York City, USA) using default settings unless specified otherwise. Cheminformatic analysis was carried out using Canvas module of Schrödinger software suite and DataWarrior software (developed by Actelion Pharmaceuticals Ltd, Allschwil, Switzerland). Theoretical QM studies were carried out using Jaguar program in Maestro module of Schrodinger suite on Stallo supercomputer using default settings unless specified otherwise.

14.2. Chemoinformatics analysis of libraries

Virtual libraries for both series were generated using the JChem software suite. Chemoinformatics 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.

14.3. 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 Sigma-Aldrich (St. Louis, MO, USA) and other suppliers (if not available from Sigma-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.

14.3.1. Cost calculation for aldehydes

Aldehyde	Price (NOK)	Amount (g)	MW	mMol	mMol/1.4 (equiv. For reaction)	NOK per 1 mmol reaction
4-pyridaldehyde	745	100	107	934.58	667.56	1.12
3-pyridaldehyde	417	100	107	934.58	667.56	0.62
2-pyridaldehyde	160	25	107	233.64	166.89	0.96
4-hydroxybenzaldehyde	318	50	122	409.84	292.74	1.09
3-hydroxybenzaldehyde	309	100	122	819.67	585.48	0.53
2-hydroxybenzaldehyde	209	100	122	819.67	585.48	0.36
Thiophene-2-carboxaldehyde	103	25	112	223.21	159.44	0.65
Thiophene-3-carboxaldehyde	1090	10	112	89.29	63.78	17.09
Furan-2-carboxaldehyde	235	116[a]	96	1208.33	863.10	0.27
Furan-3-carboxaldehyde	485	5	96	52.08	37.20	13.04
Imidazole-4-carboxaldehyde	724	5	96	52.08	37.20	19.46
4-(1-imidazolyl)-benzaldehyde	1291	5	172	29.07	20.76	62.17
N-methyl-Imidazole-2-carboxaldehyde	351	1	110	9.09	6.49	54.05
N-methyl-Imidazole-5-carboxaldehyde	1137	1	110	9.09	6.49	175.10
4-anisaldehyde	464	100	136	735.29	525.21	0.88
3-anisaldehyde	136	25	136	183.82	131.30	1.04
2-anisaldehyde	391	100	136	735.29	525.21	0.74
4-(*N,N*-dimethylamino)benzaldehyde	269	100	149	671.14	479.39	0.56
Quinoline-3-carboxaldehyde	3279	5	157	31.85	22.75	144.14
Isoquinoline-5-carboxaldehyde	937	1	157	6.37	4.55	205.95
Indole-3-carboxaldehyde	741	25	145	172.41	123.15	6.02
Indole-5-carboxaldehyde	1165	5	145	34.48	24.63	47.30

[a]Price for 100 mL quantity, density = 1.16, therefore 116 g.

14.3.2. Cost calculation for amino acids

Amino acid	Price (NOK)	Amount (g)	MW	mMol	NOK per 1 mmol reaction
Glycine	268	100	75	1333.33	0.20
L-Alanine	642	100	89	1123.60	0.57
L-Serine	892	100	105	952.38	0.94
L-valine	293	25	117	213.68	1.37
L-leucine	170	25	131	190.84	0.89
L-isoleucine	250	10	131	76.34	3.28
L-tyrosine	471	100	181	552.49	0.85
L-phenylalanine	800	100	165	606.06	1.32
L-cysteine	501	100	121	826.45	0.61
L-threonine	464	25	119	210.08	2.21

L-methionine	583	100	149	671.14	0.87
L-tryptophan	348	25	204	122.55	2.84
L-histidine	319	25	155	161.29	1.98
L-asparagine	218	25	132	189.39	1.15
L-glutamine	624	100	146	684.93	0.91
L-aspartic acid	176	100	133	751.88	0.23
L-glutamic acid	204	100	147	680.27	0.30
L-arginine	595	1000	174	5747.13	0.10
L-lysine	370	25	146	171.23	2.16
D-Alanine	1275	25	89	280.90	4.54
D-Serine	659	25	105	238.10	2.77
D-valine	355	5	117	42.74	8.31
D-leucine	986	10	131	76.34	12.92
D-isoleucine	6910	1	131	7.63	905.21
D-tyrosine	952	5	181	27.62	34.46
D-phenylalanine	350	5	165	30.30	11.55
D-cysteine	2779	1	121	8.26	336.26
D-threonine	468	5	119	42.02	11.14
D-methionine	380	5	149	33.56	11.32
D-tryptophan	370	5	204	24.51	15.10
D-histidine	1266	5	155	32.26	39.25
D-asparagine monohydrate	1152	25	150	166.67	6.91
D-glutamine	662	1	146	6.85	96.65
D-aspartic acid	712	25	133	187.97	3.79
D-glutamic acid	510	5	147	34.01	14.99
D-arginine	737	1	174	5.75	128.24
D-lysine	871	1	146	6.85	127.17

14.3.3. Cost calculations for amino acid esters

Amino acid ester	Price (NOK)	Amount (g)	MW	mMol	NOK per 1 mmol reaction	Supplier (other than Aldrich)
Glycine methyl ester HCl	256	100	125.5	796.81	0.32	
L-Alanine methyl ester HCl	1024	25	139.5	179.21	5.71	
L-Serine methyl ester HCl	358	5	155.5	32.15	11.13	
L-valine methyl ester HCl	476	10	167.6	59.67	7.98	
L-leucine methyl ester HCl	444	25	181.6	137.67	3.23	
L-isoleucine methyl ester HCl	522	5	181.6	27.53	18.96	
L-tyrosine methyl ester HCl	430	25	231.6	107.94	3.98	
L-phenylalanine ethyl ester HCl	452	10	229.7	43.54	10.38	
L-cysteine ethyl ester HCl	473	25	185.6	134.70	3.51	
L-threonine methyl ester HCl	3196	5	169.6	29.48	108.41	
L-methionine methyl ester HCl	952	25	199.7	125.19	7.60	

L-tryptophan methyl ester HCl	546	5	254.7	19.63	27.81	
L-histidine methyl ester diHCl	850	25	242	103.31	8.23	
L-asparagine tert-butyl ester HCl	1105	1	224.7	4.45	248.29	
L-glutamine methyl ester HCl	403	0.25	196.6	1.27	316.92	
L-aspartic acid dimethyl ester HCl	719	5	197.6	25.30	28.41	
L-glutamic acid dimethyl ester HCl	544	10	211.6	47.26	11.51	
L-arginine methyl ester diHCl	413	5	261	19.16	21.56	
L-lysine methyl ester diHCl	846	25	247	101.21	8.36	
D-Alanine methyl ester HCl	484	5	139.5	35.84	13.50	
D-Serine methyl ester HCl	451	5	155.5	32.15	14.03	
D-valine methyl ester HCl	793	5	167.6	29.83	26.58	
D-leucine methyl ester HCl	676	5	181.7	27.52	24.57	Chemimpex
D-isoleucine methyl ester HCl	1885	1	145	6.90	273.33	SCBT
D-tyrosine methyl ester HCl	177	5	231.7	21.58	8.20	Chemimpex
D-phenylalanine methyl ester HCl	543	5	215.7	23.18	23.43	
D-cysteine methyl ester HCl	420	1	171.6	5.83	72.07	TCI
D-threonine methyl ester HCl	814	5	169.6	29.48	27.61	Chemimpex
D-methionine methyl ester HCl	308	5	199.7	25.04	12.30	Chemimpex
D-tryptophan methyl ester HCl	1134	5	254.7	19.63	57.77	
D-histidine methyl ester diHCl	386	1	242	4.13	93.41	TCI
D-asparagine methyl ester HCl	3565	5	182.6	27.38	130.19	Oakwood
D-glutamine methyl ester HCl	591	1	196.6	5.09	116.19	Fluorochem
D-aspartic acid dimethyl ester HCl	686	5	197.6	25.30	27.11	Chemimpex
D-glutamic acid dimethyl ester HCl	257	5	211.6	23.63	10.88	Chemimpex
D-arginine methyl ester diHCl	480	1	261	3.83	125.28	TCI
D-lysine methyl ester diHCl	865	5	233	21.46	40.31	

14.4. General docking protocols

14.4.1. Protein preparation and grid generation

The crystal structures were downloaded from the PDB. Using the protein preparation wizard, the target structures were first preprocessed, missing side chains were added if required, water molecules were removed unless they were important as H-bond bridges between two or more side chains. Het-states were generated with the S2 state chosen by default. H-bonds were optimized, followed by restrained minimization. Receptor grids were usually generated with such minimized protein structures unless mentioned specifically for unminimized ones. The active site was identified with the native ligand molecules. In absence of native ligand, the key active site residues such as residues from hinge and DFG loop for kinases were specified. During docking, all rotatable groups were allowed to rotate.

14.4.2. Receptor grid generation for homology models

The model structure (from which the homology model was to be created) was downloaded from PDB after identification from a blast homology search using the target sequence. The incompatible residues were noted and the model was built using "knowledge based model building". After the model was created, the proper loops were subjected to loop refinement. The homology model thus generated was then subjected to H-bond optimization and restrained minimization. Unless specified otherwise, the receptor grid was generated for model structures after minimization. As the homology models are devoid of any native ligand, the active site was specified with key amino acid residues of active sites such as Val123, Asp184 in PKA and its analogues in other kinases. During docking, all rotatable groups were allowed to rotate.

14.4.3. Docking procedure

Native ligands from crystal structures were usually extracted from respective crystal structures and were subjected to LigPrep application along with virtual library of compounds to generate rotamers and charge variants. The docking was carried out without any constraints. For small scale docking studies (ligands < 1000), a virtual screening workflow was used with specifying all receptor grids on which the docking was to be performed. All poses were kept after docking using "high throughput virtual screening" (HTVS) mode and "standard precision" (SP) mode sequentially. The poses were finally subjected to "extra precision" (XP) mode and the best pose was kept for each ligand structure. For large scale docking studies (ligands > 1000), each docking step on each receptor grid was carried out separately and only one pose for each ligand was kept after HTVS, SP and XP modes each.

14.5. QM studies

14.5.1. Generation of structures

The structures were initially generated as 2D structures with trans orientations using MarvinSketch program, ChemAxon and were imported to the Maestro module of Schrodinger suite from which all further studies were performed. The cis geometries were generated by manual adjustment of ω dihedral.

14.5.2. Conformational search

Initial conformers sets for all geometries were generated with the MacroModel[205] conformational search algorithm using MMFFs force-field.

14.5.3. Geometric optimization

This and all further studies were performed using Jaguar[206,207] program. The conformational search conformers were subjected to geometric optimization at B3LYP/6-31G**++ level with maximum grid density and "accurate" accuracy level of SCF. Frequency analyses were carried out to confirm convergence to optimized minimum energy geometries with no imaginary frequencies. For NMA, "tight" convergence criteria were required to locate the optimized geometry. For all structures, only the minimum energy conformers were used for further studies.

14.5.4. Relaxed coordinate scan (RCS)

The optimized cis and trans geometries of NMA and *N*-substituted dipeptide esters were subjected to relaxed coordinate scan (RCS) on ω dihedral coordinate at at B3LYP/6-31G**++ level with maximum grid density and "accurate" accuracy level of SCF. The optimized geometry at each step was used to generate the starting geometry for the next step. Each RCS was performed in three rounds, with 15º, 2º and 0.125º for rounds progressively closer to the rotation barrier. The energy barrier geometries (EBGs) were identified and used as plausible transition state geometries (TSGs) for subsequent transition state searches.

14.5.5. Transition state search

Transition state search was carried out using Quadratic Synchronous Transit (QST) method[208] along reactant-product path at B3LYP/6-31G**++ level with maximum grid density and "accurate" accuracy level of SCF. For transition state search for cis-trans isomerization, EBGs obtained from RCS were used as the plausible transition state input, while optimized trans and cis geometries were used as reactants and products respectively. For transition state search for cyclization reaction, plausible transition state geometries were manually generated from minimum energy conformers of corresponding intermediates. Frequency analyses were carried out to confirm convergence to a single imaginary frequency (corresponding to the vibration connecting reactant and product geometries) for optimized TSGs. From the *syn* and *anti* TSGs of GGMe, 12 more transition state rotamers (six of each type) were created by manual torsional adjustment of rotatable bonds, and were used for QST transition state search to locate other transition state rotamers. Thus, total 14 transition state rotamers of GGMe were generated. For cyclization reaction, both stereoisomers (*S* and *R*) at ester carbonyl carbon of TS1 and intermediate were generated for triply substituted compounds (D111, D112, D333, D334).

14.5.6. Intrinsic reaction coordinates (IRC)

The reaction paths linking the reactant and product geometries via TSGs were established with mass weighted intrinsic reaction coordinate analysis[209] in both forward and reverse directions at B3LYP/6-31G**++ level with maximum grid density and "accurate" accuracy level of SCF. The step-size were varied from 0.1 to 2.0. Initially each IRC study was carried out to generate 20 IRC points in both directions, then, if necessary, further downhill IRC study were continued till clear IRC paths were established.

14.5.7. Single point energy calculations

Accurate single point energies were calculated for all optimized geometries, energy barrier geometries and transition state geometries at the B3LYP/6-311G-3DF-3PD++ level with maximum grid density and "accurate" accuracy level of SCF. Vibrational analyses were carried out at the B3LYP/CC-PVTZ(-F)++ level and the free energy values obtained for 298.15 K were used to calculate free energy changes.

14.5.8. NMA contour generation

The coordinate scan geometries of NMA for the creation of contour were generated using MMFFs force-field along ω dihedral values between 0º - 180º and Θ dihedral values between (-10)º to 190º with 2º increments. Once the geometries were generated, their single point energies were calculated at the B3LYP/6-31G**++ level.

15. Synthetic chemistry – General

Unless mentioned specifically, all reagents and solvents were purchased and 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), while 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 DCM were used as common mobile phase to elute compounds in increasing order of polarity. NMR spectra were recorded at 400 MHz (^{1}H NMR) and at 100 MHz (^{13}C NMR) using Varian spectrometer. The chemical shifts are expressed in ppm relative to TMS or to the literature value of the residual solvent peak. 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. Microwave assisted reactions were carried out using Biotage® Initiator 300 W instrument, keeping fixed hold time ON.

15.1. General procedure for synthesis of *N*-substituted amino acids (Series A)

Representative protocol for A1: Reaction of pyridine-3-carboxaldehyde with glycine

Pyridine-3-carboxaldehyde (1.498 g, 14 mmol, 1.4 equiv.) was added to a stirred solution of glycine (750 mg, 10 mmol, 1 equiv.) in methanol (20 mL), neutralized with finely powdered NaOH (420 mg, 10.5 mmol, 1.05 equiv.). After imine formation for 30 min at room temperature, the solution was cooled on an ice-bath and $NaBH_4$ pellets (494 mg, 13 mmol, 1.3 equiv.) were added. The reaction mixture was stirred under inert N_2 atmosphere for 2 hours at 0 °C followed by acidification with 37 % aq. HCl up to pH = 5.97 i.e. isoelectric point of glycine. The solvent was then distilled off at reduced pressure. The solid mixture was triturated with acetone and filtered. The residue after filtration was dried in air only for 2 min after it was found to be extremely hygroscopic. Hence, the residue (with some acetone) giving 4.78 g crude product. The attempts to remove salts with water-wash resulted in considerable loss of product, hence the product was used in next step without further purification.

15.2. General procedure for synthesis of *N*-substituted dipeptides (Series B)

Representative protocol for B1: Coupling of A1 with glycine methyl ester hydrochloride

A1 (478 mg crude product, 1 mmol, 1 equiv.) was mixed in acetonitrile (3 mL) into which, triethylamine (484.8 mg, 4.8 mmol, 4.8 equiv.) and HBTU (455 mg, 1.2 mmol, 1.2 equiv.) were slowly added sequentially. After 10 min stirring at room temperature, glycine methyl ester hydrochloride (125.5 mg, 1 mmol, 1 equiv.) was added. The reaction mixture was stirred for about 20 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 with silica gel as stationary phase and 0-3 % methanol in DCM as mobile phase to yield 157 mg desired product and 27.2 mg of the cyclic dimer (symmetrical DKP) side product. The product was further purified on preparative TLC using 10 % methanol in DCM as the mobile phase.

15.3. General procedure for synthesis of cyclic dimers (Series C)

Representative protocol for C2: Dimerization of A2

A2 (141 mg crude product, 0.5 mmol, 1 equiv.) was mixed in acetonitrile (2 mL) into which, triethylamine (242 mg, 2.4 mmol, 4.8 equiv.) and HBTU (228 mg, 0.6 mmol, 1.2 equiv.) were slowly added sequentially. The reaction mixture was stirred for 24 hours 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 with silica gel as stationary phase and 0-2 % methanol in DCM as mobile phase to yield 27.8 mg of the cyclic dimer (cyclic dimer) product D2. The product was further purified on preparative TLC using 8 % methanol in DCM as the mobile phase.

15.4. General procedure for microwave assisted cyclization (Series D)

Representative protocol for D1: Microwave assisted cyclization of B1

52 mg of B1 was mixed in 3 mL of water in a 2-5 mL microwave vial. The mixture was heated under microwave irradiation with 300 W power for 20 minutes at 200 °C and 20 atm pressure. After completion, the solvent was evaporated to dryness under reduced pressure and the crude mixture was subjected to column chromatography with silica gel as the stationary phase and 0-4 % methanol in DCM as the mobile phase to yield 35 mg of desired compound. On a small scale, the product was further purified using 10 % methanol in DCM as the mobile phase.

15.5. General procedure for synthesis of tartaric acid anhydrides (E)

Representative protocol for E1: Synthesis of (3*R*,4*R*)-2,5-dioxotetrahydrofuran-3,4-diyl diacetate

A solution of 100 µL conc. sulphuric acid in acetic anhydride (9.29 mL, 98.4 mmol, 4.92 equiv.) was added to a 50 mL single neck RBF containing L-(+)-tartaric acid (3.0 g, 20 mmol, 1.0 equiv.) at room temperature and was stirred until dissolved. The mixture was then heated at 135-137 °C under reflux for 10 min. The reaction mixture showed change in colour to light orange-brown. After 10 min, the reaction mixture was poured into a dry beaker cooled under ice-bath and the product was allowed to crystallize for 1 hour. The crystallized product was filtered, washed with benzene (20 mL) and cold diethylether (20 mL) and then was dried in vacuum to afford 3.02 g product.

15.6. General procedure for synthesis of tartaric acid monoamides (F)

Representative protocol for F1: Synthesis of (2R,3R)-2,3-diacetoxy-4-oxo-4-(pyridin-3-ylamino)butanoic acid

3-aminopyridine (564 mg, 6.0 mmol, 1.0 equiv.) was added portion-wise to a solution of A1 (1.36 g, 6.3 mmol, 1.05 equiv.) in 10 mL DCM at room temperature. The reaction mixture was stirred for 2 hours and monitored on TLC. On completion, the solvent was evaporated under reduced pressure and the crude mixture was washed 3 times with 10 mL ether each, followed by drying under vacuum to provide 2.037 g crude product, which was used for further reactions.

15.7. General procedure for synthesis of tartaric acid bisamides (G)

Representative protocol for G1: Synthesis of (2R,3R)-2,3-dihydroxy-4-morpholino-4-oxo-N-(pyridin-3-yl)butanamide

Under inert argon atmosphere, the solution of F1 (303 mg, 1 mmol, 1 equiv.) in thionyl chloride (2 mL) was stirred in a 10 mL RBF for 30 min at room temperature. The solvent was then evaporated under argon flow up to dryness. Then to the RBF, dry THF was added keeping the inert atmosphere intact and stirred for 5 min. Then, 3-aminopyridine (94 mg, 1 mmol, 1 equiv.) was added to the mixture under argon bed. The reaction mixture was again stirred for 30 min at room temperature and monitored on TLC. On completion, the solvent was evaporated again to dryness under argon flow. Then 28 % aq. ammonium hydroxide solution (2 mL) was added and reaction mixture was again stirred for 2 hours at room temperature and monitored on TLC. After completion, the solvent was evaporated under reduced pressure up to dryness. The crude product was then purified with column chromatography using silica gel 60 (230-400 mesh) as the stationary phase and 0-4 % methanol in DCM as the mobile phase, to give 79 mg of product.

15.8. General procedure for synthesis of tartrimides (H)

Representative protocol for H1: Synthesis of (3R,4R)-2,5-dioxo-1-(pyridin-4-ylmethyl)pyrrolidine-3,4-diyl diacetate

In a 10 mL RBF, 5 mL thionyl chloride was mixed with F5 (65 mg, 0.2 mmol, 1 equiv.) under inert argon atmosphere. Keeping the inert atmosphere intact, the mixture was heated at 80 °C reflux for 20 min. After 20 min, the reaction mixture was slowly poured in a beaker containing 50 mL pentane cooled with ice-bath and allowed to precipitate for 1 hour. After

1 hour, the precipitates were filtered and were washed 3 times with pentane (10 mL each), followed by drying under air to afford 42.3 mg product.

15.9. Spectral data report

15.9.1. Series A: *N*-substituted amino acids

A1: (Pyridin-3-ylmethyl)glycine

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.54 (s, 1H), 8.44 (d, *J* = 4.9 Hz, 1H), 7.88 (d, *J* = 7.8 Hz, 1H), 7.42 (dd, *J* = 7.9, 5.0 Hz, 1H), 3.82 (s, 2H), 3.18 (s, 2H).

^{13}C NMR (100 MHz, DMSO-d_6) δ 175.16, 149.75, 148.15, 137.12, 136.01, 123.75, 54.16, 51.10.

A2: (Pyridin-3-ylmethyl)-L-alanine

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.71 (d, *J* = 2.3 Hz, 1H), 8.60 (dd, *J* = 4.9, 1.6 Hz, 1H), 8.04 (dt, *J* = 7.9, 2.1 Hz, 1H), 7.53 (dd, *J* = 7.9, 4.9 Hz, 1H), 4.31 (d, *J* = 13.1 Hz, 1H), 4.23 (d, *J* = 13.1 Hz, 1H), 3.63 (q, *J* = 7.1 Hz, 1H), 1.54 (d, *J* = 7.1 Hz, 4H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 172.86, 150.16, 149.53, 138.51, 128.72, 124.11, 104.99, 57.90, 15.06.

A3: (Pyridin-3-ylmethyl)-L-phenylalanine – HCl salt

^{1}H NMR (400 MHz, D$_2$O) δ 8.77 (d, *J* = 2.1 Hz, 1H), 8.75 (dt, *J* = 5.9, 1.2 Hz, 1H), 8.57 (dq, *J* = 8.3, 1.5 Hz, 1H), 8.01 (dd, *J* = 8.2, 5.9 Hz, 1H), 7.32 – 7.21 (m, 3H), 7.21 – 7.15 (m, 2H), 4.40 (s, 2H), 4.05 (dd, *J* = 7.2, 6.1 Hz, 1H), 3.25 – 3.11 (m, 2H).

^{13}C NMR (100 MHz, D$_2$O) δ 171.55, 148.53, 142.64, 142.35, 134.22, 130.96, 129.37, 129.10, 127.88, 127.78, 63.09, 46.60, 35.61.

A5: (Pyridin-4-ylmethyl)-L-phenylalanine – HCl salt

^{1}H NMR (400 MHz, D$_2$O) δ 8.79 – 8.72 (m, 2H), 8.01 (dt, *J* = 5.8, 0.8 Hz, 2H), 7.33 – 7.26 (m, 3H), 7.24 (dd, *J* = 7.8, 1.8 Hz, 2H), 4.53 (s, 2H), 4.37 – 4.26 (m, 1H), 3.39 – 3.22 (m, 2H).

^{13}C NMR (100 MHz, D$_2$O) δ 170.43, 150.92, 141.86, 133.64, 129.38, 129.20, 128.09, 127.82, 62.16, 48.46, 35.12.

A6: (Pyridin-2-ylmethyl)-L-phenylalanine

^{1}H NMR (400 MHz, DMSO-d_6) δ 8.45 (dt, J = 4.9, 1.3 Hz, 1H), 7.68 (td, J = 7.7, 1.8 Hz, 1H), 7.27 – 7.18 (m, 7H), 3.89 (d, J = 14.7 Hz, 1H), 3.77 (d, J = 14.6 Hz, 1H), 3.44 (dd, J = 7.3, 6.0 Hz, 1H), 3.01 (dd, J = 13.8, 6.0 Hz, 1H), 2.90 (dd, J = 13.7, 7.3 Hz, 1H).

^{13}C NMR (100 MHz, DMSO-d_6) δ 174.55, 159.47, 149.07, 138.74, 136.80, 129.64, 128.42, 126.54, 122.37, 122.25, 62.65, 52.86, 38.79.

A7: (Thiophen-2-ylmethyl)-L-phenylalanine

^{1}H NMR (400 MHz, DMSO-d_6) δ 7.35 (dd, J = 5.1, 1.3 Hz, 1H), 7.27 – 7.22 (m, 2H), 7.21 – 7.17 (m, 3H), 6.91 (dd, J = 5.1, 3.4 Hz, 1H), 6.88 (d, J = 3.4 Hz, 1H), 3.95 (d, J = 14.2 Hz, 1H), 3.78 (d, J = 14.2 Hz, 1H), 3.38 (t, J = 6.7 Hz, 1H), 2.87 (qd, J = 13.6, 6.7 Hz, 2H).

^{13}C NMR (100 MHz, DMSO-d_6) δ 174.83, 143.76, 138.47, 129.72, 128.46, 127.01, 126.66, 125.53, 125.41, 61.64, 46.06, 38.75.

A8: (Furan-2-ylmethyl)-L-phenylalanine

^{1}H NMR (400 MHz, DMSO-d_6) δ 7.56 – 7.49 (m, 1H), 7.20 (ddt, J = 15.2, 12.3, 7.3 Hz, 7H), 6.37 – 6.31 (m, 1H), 6.16 (d, J = 3.2 Hz, 1H), 3.75 (d, J = 14.6 Hz, 1H), 3.63 (d, J = 14.3 Hz, 1H), 3.38 (t, J = 6.6 Hz, 1H), 2.91 (dd, J = 13.7, 6.2 Hz, 1H), 2.84 (dd, J = 13.7, 7.0 Hz, 1H).

^{13}C NMR (100 MHz, DMSO-d_6) δ 174.70, 153.96, 142.26, 138.66, 129.60, 128.40, 126.53, 110.64, 107.27, 61.95, 44.26, 38.84.

A10: (3-Hydroxybenzyl)-L-phenylalanine

^{1}H NMR (400 MHz, DMSO-d_6) δ 7.27 – 7.20 (m, 4H), 7.20 – 7.14 (m, 1H), 7.04 (t, J = 7.7 Hz, 1H), 6.72 (t, J = 2.1 Hz, 1H), 6.66 (dd, J = 8.1, 2.7 Hz, 2H), 3.69 (d, J = 13.4 Hz, 1H), 3.57 (d, J = 13.3 Hz, 1H), 3.40 (t, J = 6.6 Hz, 1H), 2.98 (dd, J = 13.8, 6.3 Hz, 1H), 2.90 (dd, J = 13.8, 6.9 Hz, 1H).

^{13}C NMR (100 MHz, DMSO-d_6) δ 172.74, 157.93, 138.35, 138.11, 129.75, 129.63, 128.52, 126.71, 119.62, 116.20, 115.09, 62.30, 50.58, 37.65.

A11: (Pyridin-4-ylmethyl)glycine

^{1}H NMR (400 MHz, DMSO-d_6) δ 8.52 (d, J = 5.0 Hz, 2H), 7.40 (d, J = 5.0 Hz, 2H), 3.90 (s, 2H), 3.16 (s, 2H).

^{13}C NMR (100 MHz, DMSO-d_6) δ 171.07, 150.01, 146.37, 123.97, 50.31, 50.13.

A12: (Pyridin-4-ylmethyl)-L-alanine

^{1}H NMR (400 MHz, D$_2$O) δ 8.50 (d, J = 5.4 Hz, 2H), 7.43 (d, J = 5.4 Hz, 2H), 4.19 (s, 2H), 3.65 (q, J = 7.2 Hz, 1H), 1.43 (d, J = 7.2 Hz, 3H).

^{13}C NMR (100 MHz, D$_2$O) δ 174.50, 149.36, 141.32, 124.64, 57.94, 48.28, 15.19.

A13: (Pyridin-4-ylmethyl)-D-alanine

^{1}H NMR (400 MHz, D$_2$O) δ 8.57 – 8.42 (m, 2H), 7.52 – 7.38 (m, 2H), 4.20 (s, 2H), 3.66 (q, J = 7.2 Hz, 1H), 1.44 (d, J = 7.2 Hz, 3H).

^{13}C NMR (100 MHz, D$_2$O) δ 174.50, 149.35, 141.34, 124.67, 57.95, 48.29, 15.22.

A14: ((1*H*-indol-5-yl)methyl)-L-alanine

^{1}H NMR (400 MHz, DMSO-d_6) δ 11.41 (s, NH), 7.64 (s, 1H), 7.46 – 7.29 (m, 2H), 7.20 (d, J = 8.3 Hz, 1H), 6.39 (s, 1H), 4.06 (t, J = 8.7 Hz, 2H), 3.19 (q, J = 7.2 Hz, 1H), 1.31 (d, J = 7.0 Hz, 3H).

^{13}C NMR (100 MHz, DMSO-d_6) δ 172.07, 136.21, 127.97, 126.56, 123.77, 123.11, 122.18, 111.90, 101.46, 56.58, 50.02, 16.30.

A15: ((1*H*-indol-5-yl)methyl)-D-alanine

^{1}H NMR (400 MHz, DMSO-d_6) δ 11.39 (s, NH), 7.63 (s, 1H), 7.44 – 7.30 (m, 2H), 7.20 (dd, J = 8.2, 1.7 Hz, 1H), 6.47 – 6.34 (m, 1H), 4.14 – 4.00 (m, 2H), 3.19 (q, J = 7.1 Hz, 1H), 1.31 (d, J = 7.0 Hz, 3H).

^{13}C NMR (100 MHz, DMSO-d_6) δ 171.55, 136.22, 127.96, 126.58, 123.46, 123.13, 122.26, 111.90, 101.47, 56.50, 49.91, 16.14.

15.9.2. Series B: *N*-substituted dipeptide esters

B1: Methyl 2-(2-(pyridin-3-ylmethylamino)acetamido)acetate

^{1}H NMR (400 MHz, CDCl$_3$) δ 8.54 (d, J = 2.3 Hz, 1H), 8.47 (d, J = 4.9 Hz, 1H), 7.67 (dt, J = 8.4, 1.8 Hz, 1H), 7.25 (dd, J = 7.8, 4.9 Hz, 1H), 4.04 (d, J = 5.3 Hz, 2H), 3.80 (s, 2H), 3.72 (s, 3H), 3.34 (s, 2H), 2.89 (s, NH).

^{13}C NMR (100 MHz, CDCl$_3$) δ 171.39, 170.32, 149.56, 148.72, 136.13, 134.44, 123.58, 52.37, 51.49, 50.99, 40.70.

IR (neat): 1745, 1659, 1207, 842, 712 cm^{-1}

MS (ESI): m/z = 238.1185 (MH$^+$), calculated for C$_{11}$H$_{16}$O$_3$N$_3$ = 238.1186

B3: (*S*)-methyl 2-(3-phenyl-2-(pyridin-3-ylmethylamino)propanamido)acetate

^{1}H NMR (400 MHz, CDCl$_3$) δ 8.47 (dd, *J* = 4.9, 1.7 Hz, 1H), 8.40 – 8.33 (m, 1H), 7.71 (t, *J* = 5.8 Hz, 1H), 7.45 – 7.39 (m, 1H), 7.35 – 7.23 (m, 2H + CDCl$_3$), 7.21 – 7.12 (m, 3H), 4.17 (dd, *J* = 18.3, 6.2 Hz, 1H), 4.01 (dd, *J* = 18.3, 5.1 Hz, 1H), 3.80 – 3.77 (m, 4H), 3.57 (d, *J* = 13.5 Hz, 1H), 3.41 (dd, *J* = 9.9, 4.1 Hz, 1H), 3.25 (dd, *J* = 14.0, 4.1 Hz, 1H), 2.73 (dd, *J* = 13.9, 9.9 Hz, 1H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 173.65, 170.31, 149.45, 148.61, 137.04, 135.76, 134.44, 128.97, 128.89, 127.11, 123.43, 63.19, 52.35, 49.88, 40.76, 39.19.

IR (neat): 1750, 1668, 1208, 702 cm^{-1}

MS (ESI): m/z = 328.1655 (MH$^+$), calculated for C$_{18}$H$_{22}$O$_3$N$_3$ = 328.1656

B5: (*S*)-methyl 3-(1*H*-indol-3-yl)-2-((*S*)-2-(pyridin-3-ylmethylamino)propanamido)propanoate

^{1}H NMR (400 MHz, CDCl$_3$) δ 8.40 (d, *J* = 4.8 Hz, 1H), 8.33 – 8.23 (m, 1H + NH), 7.63 (d, *J* = 8.7 Hz, 1H), 7.48 (d, *J* = 7.9 Hz, 1H), 7.28 (ddd, *J* = 7.4, 5.6, 1.6 Hz, 1H + CDCl$_3$), 7.16 – 6.97 (m, 3H), 6.95 – 6.88 (m, 1H), 4.89 (dt, *J* = 8.6, 5.5 Hz, 1H), 3.66 – 3.58 (m, 4H), 3.52 (d, *J* = 13.1 Hz, 1H), 3.35 – 3.21 (m, 2H), 3.14 (q, *J* = 6.9 Hz, 1H), 1.18 (d, *J* = 6.9 Hz, 3H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 174.31, 172.41, 149.48, 148.56, 136.12, 135.82, 134.84, 127.62, 123.38, 122.66, 122.26, 119.59, 118.69, 111.27, 110.06, 57.99, 52.32, 52.16, 49.68, 27.53, 19.46.

IR (neat): 2918, 1741, 1661, 743 cm^{-1}

MS (ESI): m/z = 381.1918 (MH$^+$), calculated for C$_{21}$H$_{25}$O$_3$N$_4$ = 381.1921

B6: (*R*)-methyl 3-(1*H*-indol-3-yl)-2-((*S*)-3-phenyl-2-(pyridin-3-ylmethylamino)propanamido)propanoate

^{1}H NMR (400 MHz, CDCl$_3$) δ 8.41 (dd, *J* = 4.7, 1.8 Hz, 1H), 8.19 (s, NH), 8.08 – 8.00 (m, 1H), 7.64 (d, *J* = 7.9 Hz, 1H), 7.54 (d, *J* = 7.6 Hz, 1H), 7.26 (qdd, *J* = 6.7, 4.9, 2.9 Hz, 3H+CDCl$_3$), 7.15 – 7.11 (m, 3H), 7.08 (ddd, *J* = 8.3, 7.0, 1.2 Hz, 1H), 7.02 (dd, *J* = 7.9, 4.7 Hz, 1H), 6.99 – 6.93 (m, 1H), 6.87 (d, *J* = 2.4 Hz, 1H), 4.96 (dt, *J* = 7.8, 6.1 Hz, 1H),

3.76 (d, J = 1.1 Hz, 3H), 3.40 (dd, J = 14.9, 5.5 Hz, 1H), 3.34 – 3.27 (m, 1H), 3.27 – 3.20 (m, 3H), 3.16 (d, J = 13.7 Hz, 1H), 2.70 (dd, J = 13.2, 9.0 Hz, 1H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 173.07, 172.41, 149.27, 148.35, 137.25, 136.10, 135.32, 134.44, 129.02, 128.80, 127.57, 126.96, 123.21, 122.58, 122.31, 119.69, 118.58, 111.25, 110.06, 63.08, 52.42, 52.29, 49.35, 38.86, 27.33.

IR (neat): 1742, 1661, 1512, 743 cm^{-1}

MS (ESI): m/z = 457.2237 (MH$^+$), calculated for C$_{27}$H$_{29}$O$_3$N$_4$ = 457.2234

B8: (*S*)-methyl 2-(3-phenyl-2-(pyridin-4-ylmethylamino)propanamido)acetate

^{1}H NMR (400 MHz, CDCl$_3$) δ 8.42 – 8.31 (m, 2H), 7.63 (t, J = 5.9 Hz, NH), 7.29 – 7.16 (m, 3H), 7.14 – 7.06 (m, 2H), 6.90 (d, J = 5.4 Hz, 2H), 4.07 (dd, J = 18.3, 6.1 Hz, 1H), 3.94 (dd, J = 18.3, 5.1 Hz, 1H), 3.76 – 3.65 (m, 4H), 3.51 (d, J = 14.5 Hz, 1H), 3.29 (dd, J = 9.9, 4.0 Hz, 1H), 3.16 (dd, J = 13.9, 4.0 Hz, 1H), 2.66 (dd, J = 13.9, 9.9 Hz, 1H), 1.86 (s, 1H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 172.61, 169.30, 148.78, 147.01, 136.07, 128.02, 127.88, 126.11, 121.75, 62.08, 51.32, 50.04, 39.74, 38.29.

IR (neat): 1750, 1670, 1206, 701 cm^{-1}

MS (ESI): m/z = 328.1657 (MH$^+$), calculated for C$_{18}$H$_{22}$O$_3$N$_3$ = 328.1656

B9: (*R*)-methyl 3-(1*H*-indol-3-yl)-2-((*S*)-3-phenyl-2-(thiophen-2-ylmethylamino)-propanamido)propanoate

^{1}H NMR (400 MHz, CDCl$_3$) δ 7.98 (s, NH), 7.68 (d, J = 8.1 Hz, 1H), 7.56 (d, J = 7.9 Hz, 1H), 7.30 (dd, J = 8.3, 7.2 Hz, 1H), 7.23 (dt, J = 8.7, 3.5 Hz, 2H), 7.19 – 7.14 (m, 1H), 7.13 (dt, J = 6.4, 1.5 Hz, 2H), 7.11 – 7.07 (m, 2H), 6.90 (d, J = 2.5 Hz, 1H), 6.82 (dd, J = 5.1, 3.4 Hz, 1H), 6.52 (dd, J = 3.5, 1.3 Hz, 1H), 5.00 – 4.87 (m, 1H), 3.72 (s, 3H), 3.48 (d, J = 1.7 Hz, 2H), 3.39 – 3.24 (m, 3H), 3.18 (dd, J = 14.0, 4.4 Hz, 1H), 2.73 (dd, J = 14.0, 9.6 Hz, 1H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 173.25, 172.42, 142.70, 137.26, 136.10, 129.08, 128.72, 127.48, 126.84, 126.58, 124.85, 124.41, 122.59, 122.25, 119.66, 118.63, 111.16, 110.22, 62.66, 52.35, 52.29, 46.63, 38.87, 27.61.

IR (neat): 3330, 1741, 1658, 1511, 1456, 743, 701 cm^{-1}

MS (ESI): m/z = 462.1848 (MH$^+$), calculated for C$_{26}$H$_{28}$O$_3$N$_3$S = 462.1846

B11: (*S*)-methyl 3-(4-hydroxyphenyl)-2-((*S*)-3-phenyl-2-(thiophen-2-ylmethylamino)-propanamido)propanoate

^{1}H NMR (400 MHz, CDCl$_3$) δ 7.85 (dd, *J* = 9.1, 2.5 Hz, 1H), 7.30 – 7.17 (m, 2H + CDCl$_3$), 7.13 (t, *J* = 3.6 Hz, 1H), 7.06 (d, *J* = 7.0 Hz, 2H), 6.98 – 6.90 (m, 2H), 6.86 (dt, *J* = 5.5, 2.8 Hz, 1H), 6.80 – 6.72 (m, 2H), 6.67 (d, *J* = 3.7 Hz, 1H), 4.93 (ddq, *J* = 8.0, 5.9, 3.4, 2.8 Hz, 1H), 3.83 (dd, *J* = 14.2, 2.4 Hz, 1H), 3.76 – 3.63 (m, 4H), 3.40 (dt, *J* = 10.5, 3.4 Hz, 1H), 3.12 (dt, *J* = 14.6, 3.3 Hz, 1H), 3.00 (ddt, *J* = 16.6, 14.1, 5.5 Hz, 2H), 2.50 (ddd, *J* = 12.2, 9.5, 2.4 Hz, 1H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 173.78, 171.97, 155.74, 142.26, 136.83, 130.26, 129.04, 128.76, 126.96, 126.88, 126.69, 125.32, 124.66, 115.62, 62.59, 52.80, 52.40, 46.68, 38.81, 37.43.

IR (neat): 3329, 2361, 1743, 1652, 1515, 1221, 700 cm^{-1}

MS (ESI): m/z = 439.1687 (MH$^+$), calculated for C$_{24}$H$_{27}$O$_4$N$_2$S = 439.1686

B12: (*S*)-methyl 2-((*S*)-2-(furan-2-ylmethylamino)-3-phenylpropanamido)-3-(1*H*-indol-3-yl)propanoate

^{1}H NMR (400 MHz, CDCl$_3$) δ 9.31 (d, *J* = 2.6 Hz, NH), 7.95 (d, *J* = 8.6 Hz, 1H), 7.55 (d, *J* = 7.9 Hz, 1H), 7.39 (d, *J* = 8.0 Hz, 1H), 7.31 – 7.18 (m, 4H + CDCl$_3$), 7.14 (t, *J* = 7.4 Hz, 1H), 7.07 (dd, *J* = 7.3, 2.1 Hz, 2H), 6.95 (d, *J* = 2.4 Hz, 1H), 6.22 (dd, *J* = 3.1, 1.8 Hz, 1H), 5.90 (d, *J* = 3.2 Hz, 1H), 5.03 (dt, *J* = 8.7, 5.9 Hz, 1H), 3.67 (d, *J* = 18.7 Hz, 4H), 3.51 (d, *J* = 14.9 Hz, 1H), 3.42 (dd, *J* = 9.5, 4.1 Hz, 1H), 3.36 (dd, *J* = 9.0, 5.8 Hz, 2H), 3.07 (dd, *J* = 13.9, 4.1 Hz, 1H), 2.58 (dd, *J* = 14.0, 9.5 Hz, 1H), 1.96 (s, NH).

^{13}C NMR (100 MHz, CDCl$_3$) δ 173.44, 172.37, 152.63, 141.82, 137.07, 136.37, 129.03, 128.67, 127.62, 126.85, 123.11, 121.86, 119.23, 118.55, 111.49, 110.03, 109.51, 107.22, 62.15, 52.64, 52.24, 44.42, 38.86, 27.81.

IR (neat): 3326, 1741, 1658, 1510, 741, 701 cm^{-1}

MS (ESI): m/z = 446.2066 (MH$^+$), calculated for C$_{26}$H$_{28}$O$_4$N$_3$ = 446.2074

B14: (*S*)-methyl 2-(2-(furan-2-ylmethylamino)-3-phenylpropanamido)acetate

^{1}H NMR (400 MHz, CDCl$_3$) δ 7.77 (s, NH), 7.37 – 7.18 (m, 4H), 7.18 – 7.03 (m, 2H), 6.24 (d, J = 5.9 Hz, 1H), 6.02 (d, J = 3.2 Hz, 1H), 4.06 (qt, J = 18.1, 5.6 Hz, 2H), 3.83 – 3.66 (m, 4H), 3.66 – 3.51 (m, 1H), 3.44 (dt, J = 9.7, 3.9 Hz, 1H), 3.19 (dd, J = 13.9, 4.1 Hz, 1H), 2.71 (qd, J = 9.7, 8.0, 3.6 Hz, 1H), 1.92 (s, NH).

^{13}C NMR (100 MHz, CDCl$_3$) δ 173.96, 170.28, 152.51, 141.98, 137.09, 128.97, 128.74, 126.89, 110.03, 107.42, 62.39, 52.25, 44.78, 40.76, 39.16.

IR (neat): 3340, 1750, 1668, 1207, 700 cm^{-1}

MS (ESI): m/z = 317.1498 (MH$^+$), calculated for C$_{17}$H$_{21}$O$_4$N$_2$ = 317.1496

B15: **(S)-methyl** **2-((S)-2-(furan-2-ylmethylamino)-3-phenylpropanamido)- propanoate**

^{1}H NMR (400 MHz, CDCl$_3$) δ 7.77 (d, J = 8.2 Hz, 1H), 7.30 – 7.27 (m, 1H), 7.25 – 7.22 (m, 2H), 7.16 – 7.08 (m, 2H), 6.24 (dd, J = 3.2, 1.9 Hz, 1H), 6.08 – 6.02 (m, 1H), 4.62 (dq, J = 8.3, 7.2 Hz, 1H), 3.82 – 3.75 (m, 1H), 3.74 (s, 3H), 3.59 (d, J = 14.8 Hz, 1H), 3.41 (dd, J = 9.6, 4.0 Hz, 1H), 3.15 (dd, J = 14.0, 4.0 Hz, 1H), 2.71 (dd, J = 13.9, 9.5 Hz, 1H), 1.71 (s, NH), 1.39 (d, J = 7.2 Hz, 3H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 173.27, 173.05, 152.61, 141.90, 137.02, 129.00, 128.65, 126.84, 110.00, 107.33, 62.16, 52.30, 47.41, 44.58, 39.16, 18.31.

IR (neat): 3336, 2925, 1743, 1668, 1507, 1149, 741, 701 cm^{-1}

MS (ESI): m/z = 331.1652 (MH$^+$), calculated for C$_{18}$H$_{23}$O$_4$N$_2$ = 331.1652

B16: (S)-methyl 2-(2-(3-hydroxybenzylamino)-3-phenylpropanamido)acetate

^{1}H NMR (400 MHz, Chloroform-d) δ 7.92 (t, J = 5.8 Hz, 1H), 7.34 – 7.25 (m, 3H), 7.22 – 7.14 (m, 2H), 7.11 (t, J = 7.8 Hz, 1H), 6.78 – 6.70 (m, 1H), 6.64 – 6.57 (m, 2H), 4.18 (dd, J = 18.3, 6.2 Hz, 1H), 4.01 (dd, J = 18.3, 5.1 Hz, 1H), 3.78 (s, 3H), 3.71 (d, J = 13.6 Hz, 1H), 3.55 (d, J = 13.6 Hz, 1H), 3.44 (dd, J = 9.7, 4.1 Hz, 1H), 3.24 (dd, J = 13.9, 4.1 Hz, 1H), 2.76 (dd, J = 13.9, 9.7 Hz, 1H).

^{13}C NMR (101 MHz, CDCl$_3$) δ 174.61, 170.47, 156.30, 140.67, 137.18, 129.68, 129.15, 128.82, 126.97, 119.78, 114.90, 114.29, 62.81, 52.44, 52.22, 40.86, 39.14.

IR (neat): 3332, 1749, 1654, 1213, 699 cm^{-1}

MS (ESI): m/z = 343.1655 (MH$^+$), calculated for C$_{19}$H$_{23}$O$_4$N$_2$ = 343.1652

B17: (*S*)-methyl 2-((*S*)-2-(3-hydroxybenzylamino)-3-phenylpropanamido)propanoate

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.89 (d, *J* = 8.3 Hz, 1H), 7.25 – 7.16 (m, 3H), 7.11 – 7.06 (m, 2H), 7.02 (t, *J* = 7.8 Hz, 1H), 6.69 – 6.60 (m, 1H), 6.60 – 6.56 (m, 1H), 6.53 (d, *J* = 7.5 Hz, 1H), 4.64 – 4.50 (m, 1H), 3.67 (s, 3H), 3.62 (d, *J* = 13.5 Hz, 1H), 3.47 (d, *J* = 13.5 Hz, 1H), 3.31 (dd, *J* = 9.3, 4.0 Hz, 1H), 3.11 (dd, *J* = 13.9, 4.0 Hz, 1H), 2.68 (dd, *J* = 13.9, 9.3 Hz, 1H), 1.33 (d, *J* = 7.2 Hz, 3H).

^{13}C NMR (101 MHz, CDCl$_3$) δ 173.70, 173.63, 156.29, 156.27, 140.75, 137.12, 129.68, 129.65, 129.21, 128.78, 126.96, 119.87, 114.95, 114.26, 62.65, 52.55, 52.17, 47.57, 39.13, 18.50.

IR (neat): 3325, 1743, 1651, 1453, 1217, 1156, 699 cm^{-1}

MS (ESI): m/z = 357.1809 (MH$^+$), calculated for C$_{20}$H$_{25}$O$_4$N$_2$ = 357.1809

B18: (*S*)-methyl 2-((*S*)-2-(3-hydroxybenzylamino)-3-phenylpropanamido)-3-(1*H*-indol-3-yl)propanoate

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 7.47 (d, *J* = 7.9 Hz, 1H), 7.31 (d, *J* = 8.1 Hz, 1H), 7.20 (d, *J* = 6.7 Hz, 3H), 7.11 – 6.95 (m, 6H), 6.61 (dd, *J* = 7.7, 2.4 Hz, 1H), 6.51 (s, 1H), 6.37 (d, *J* = 7.4 Hz, 1H), 4.79 (t, *J* = 6.5 Hz, 1H), 3.66 (s, 3H), 3.39 (d, *J* = 13.0 Hz, 1H), 3.29 – 3.21 (m, 4H), 2.85 (dd, *J* = 13.7, 5.8 Hz, 1H), 2.64 (dd, *J* = 13.7, 7.7 Hz, 1H).

^{13}C NMR (100 MHz, Methanol-*d*$_4$) δ 174.51, 172.17, 156.97, 140.57, 137.05, 136.64, 128.93, 128.89, 128.07, 127.28, 126.33, 123.16, 121.11, 119.03, 118.48, 117.85, 114.71, 113.50, 111.02, 108.88, 62.70, 52.64, 51.35, 51.31, 38.71, 27.06.

IR (neat): 3327, 2361, 1740, 1653, 1456, 740, 700 cm^{-1}

MS (ESI): m/z = 472.2238 (MH$^+$), calculated for C$_{28}$H$_{30}$O$_4$N$_3$ = 472.2231

B19: (*S*)-methyl 2-((*S*)-2-(3-hydroxybenzylamino)-3-phenylpropanamido)-3-(4-hydroxyphenyl)propanoate

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 7.27 – 7.16 (m, 3H), 7.09 (dd, *J* = 7.4, 1.7 Hz, 2H), 7.06 – 7.00 (m, 1H), 6.96 – 6.90 (m, 2H), 6.72 – 6.65 (m, 2H), 6.65 – 6.59 (m, 1H), 6.57 (d, *J* = 2.6 Hz, 1H), 6.52 – 6.45 (m, 1H), 4.71 – 4.62 (m, 1H), 3.67 (d, *J* = 1.1 Hz, 3H), 3.46

(d, *J* = 13.0 Hz, 1H), 3.32 (s, 2H), 3.00 (dd, *J* = 14.0, 5.3 Hz, 1H), 2.88 (dd, *J* = 13.8, 7.3 Hz, 2H), 2.71 (dd, *J* = 13.6, 7.7 Hz, 1H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 174.72, 171.98, 157.34, 156.47, 140.79, 137.29, 130.09, 129.15, 129.10, 128.28, 127.01, 126.52, 119.20, 115.20, 114.97, 113.79, 62.95, 53.41, 51.54, 51.49, 39.13, 36.48.

IR (neat): 3322, 1741, 1650, 1540, 1220,734 cm^{-1}

MS (ESI): m/z = 449.2079 (MH$^+$), calculated for $C_{26}H_{29}O_5N_2$ = 449.2071

B20:   (*R*)-methyl   3-(1*H*-indol-3-yl)-2-((*S*)-3-phenyl-2-(pyridin-4-ylmethylamino)-propanamido)propanoate

^{1}H NMR (400 MHz, CDCl$_3$) δ 8.84 (s, NH), 8.32 – 8.21 (m, 2H), 7.65 (d, *J* = 7.9 Hz, 1H), 7.52 (d, *J* = 7.8 Hz, 1H), 7.31 – 7.19 (m, 3H + CDCl$_3$), 7.15 – 7.09 (m, 3H), 7.06 (t, *J* = 7.6 Hz, 1H), 6.86 (d, *J* = 2.5 Hz, 1H), 6.59 (d, *J* = 5.2 Hz, 2H), 5.01 – 4.91 (m, 1H), 3.75 (s, 3H), 3.40 (dd, *J* = 15.0, 5.4 Hz, 1H), 3.29 (dd, *J* = 15.0, 6.6 Hz, 1H), 3.25 – 3.16 (m, 3H), 3.12 (d, *J* = 14.8 Hz, 1H), 2.68 (dd, *J* = 14.8, 10.7 Hz, 1H), 1.86 (s, NH).

^{13}C NMR (100 MHz, CDCl$_3$) δ 173.17, 172.48, 149.36, 148.27, 137.24, 136.20, 129.05, 128.82, 127.61, 127.01, 122.65, 122.53, 122.17, 119.57, 118.46, 111.36, 109.73, 63.08, 52.44, 52.42, 50.48, 39.03, 27.29.

IR (neat): 3309, 2361, 2337, 1741, 1659, 740, 701 cm^{-1}

MS (ESI): m/z = 457.2231 (MH$^+$), calculated for $C_{27}H_{29}O_3N_4$ = 457.2234

B21:   (*S*)-methyl   3-(1*H*-indol-3-yl)-2-(2-(pyridin-4-ylmethylamino)acetamido)-propanoate

^{1}H NMR (400 MHz, CDCl$_3$) δ 9.04 (s, NH), 8.44 – 8.35 (m, 2H), 7.62 (d, *J* = 8.3 Hz, 1H), 7.52 (d, *J* = 7.9 Hz, 1H), 7.34 – 7.27 (m, 1H), 7.18 – 7.10 (m, 1H), 7.10 – 7.02 (m, 1H), 6.93 (d, *J* = 2.7 Hz, 1H), 6.90 (d, *J* = 5.2 Hz, 2H), 4.96 (dt, *J* = 8.2, 5.3 Hz, 1H), 3.70 (d, *J* = 1.0 Hz, 3H), 3.60 – 3.43 (m, 2H), 3.40 – 3.27 (m, 2H), 3.25 – 3.18 (m, 2H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 172.39, 171.03, 149.62, 148.38, 136.24, 127.69, 122.85, 122.75, 122.13, 119.49, 118.40, 111.43, 109.61, 52.48, 52.32, 52.16, 51.79, 27.38.

IR (neat): 3323, 1740, 1662, 1518, 1213, 744 cm^{-1}

MS (ESI): m/z = 367.1768 (MH$^+$), calculated for $C_{20}H_{23}O_3N_4$ = 367.1765

B22: (*R*)-methyl 3-(1*H*-indol-3-yl)-2-(2-(pyridin-4-ylmethylamino)acetamido)-propanoate

^{1}H NMR (400 MHz, CDCl$_3$) δ 8.84 (s, NH), 8.45 – 8.37 (m, 2H), 7.60 (d, *J* = 8.2 Hz, 1H), 7.56 – 7.48 (m, 1H), 7.34 – 7.27 (m, 1H), 7.14 (ddd, *J* = 8.1, 6.9, 1.2 Hz, 1H), 7.06 (ddd, *J* = 8.0, 7.0, 1.1 Hz, 1H), 6.94 (d, *J* = 2.5 Hz, 1H), 6.93 – 6.89 (m, 2H), 4.97 (dt, *J* = 8.3, 5.4 Hz, 1H), 3.71 (s, 3H), 3.62 – 3.53 (m, 1H), 3.49 (dd, *J* = 15.2, 3.6 Hz, 1H), 3.40 – 3.28 (m, 2H), 3.23 (s, 2H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 172.37, 170.97, 149.67, 148.33, 136.20, 127.70, 122.77, 122.74, 122.18, 119.53, 118.43, 111.40, 109.71, 52.43, 52.38, 52.18, 51.81, 27.39.

IR (neat): 3315, 1741, 1661, 1519, 1216, 744 cm^{-1}

MS (ESI): m/z = 367.1775 (MH$^+$), calculated for C$_{20}$H$_{23}$O$_3$N$_4$ = 367.1765

B23: Methyl 2-(2-(pyridin-4-ylmethylamino)acetamido)acetate

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 8.47 (d, *J* = 5.1 Hz, 2H), 7.46 (d, *J* = 5.1 Hz, 2H), 4.00 (s, 2H), 3.85 (s, 2H), 3.73 (d, *J* = 1.4 Hz, 3H), 3.31 (s, 2H).

^{13}C NMR (100 MHz, Methanol-*d*$_4$) δ 173.21, 170.41, 150.25, 148.61, 123.48, 51.38, 50.66, 40.17, 39.02.

IR (neat): 3322, 1745, 1663, 1211 cm^{-1}

MS (ESI): m/z = 238.1189 (MH$^+$), calculated for C$_{11}$H$_{16}$O$_3$N$_3$ = 238.1186

B25: (*R*)-methyl 2-((*S*)-2-((1*H*-indol-5-yl)methylamino)propanamido)-3-(1*H*-indol-3-yl)propanoate

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 7.53 (d, *J* = 7.9 Hz, 1H), 7.31 (d, *J* = 7.6 Hz, 2H), 7.27 (d, *J* = 8.3 Hz, 1H), 7.19 (d, *J* = 3.1 Hz, 1H), 7.12 – 7.05 (m, 2H), 7.01 (ddd, *J* = 8.1, 6.8, 1.2 Hz, 1H), 6.93 (dd, *J* = 8.3, 1.7 Hz, 1H), 6.36 (dd, *J* = 3.3, 1.0 Hz, 1H), 4.78 (dd, *J* = 8.1, 5.4 Hz, 1H), 3.71 (s, 3H), 3.65 (d, *J* = 12.6 Hz, 1H), 3.50 (d, *J* = 12.6 Hz, 1H), 3.19 (dd, *J* = 14.7, 7.9 Hz, 2H), 1.11 (d, *J* = 6.9 Hz, 3H).

^{13}C NMR (100 MHz, Methanol-*d*$_4$) δ 176.33, 172.39, 136.66, 135.52, 129.43, 127.97, 127.31, 124.38, 123.05, 121.79, 121.08, 119.76, 118.44, 117.82, 111.00, 110.59, 109.08, 100.86, 56.50, 52.74, 51.96, 51.32, 27.10, 18.03.

IR (neat): 3402, 1737, 1672, 1369, 1166 cm^{-1}

MS (ESI): m/z = 419.2076 (MH$^+$), calculated for C$_{24}$H$_{27}$O$_3$N$_4$ = 419.2083

B26: (*R*)-methyl 2-((*R*)-2-((1*H*-indol-5-yl)methylamino)propanamido)-3-(1*H*-indol-3-yl)propanoate

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 7.58 – 7.51 (m, 1H), 7.35 – 7.27 (m, 2H), 7.24 (d, *J* = 8.3 Hz, 1H), 7.18 (d, *J* = 3.0 Hz, 1H), 7.12 – 7.05 (m, 2H), 7.05 – 6.97 (m, 1H), 6.87 (dd, *J* = 8.6, 1.7 Hz, 1H), 6.37 – 6.30 (m, 1H), 4.81 (d, *J* = 5.4 Hz, 1H), 3.68 (s, 3H), 3.53 (d, *J* = 12.4 Hz, 1H), 3.46 (d, *J* = 12.4 Hz, 1H), 3.28 – 3.22 (m, 1H), 3.18 (q, *J* = 6.8 Hz, 1H), 1.16 (d, *J* = 6.9 Hz, 3H).

^{13}C NMR (100 MHz, Methanol-*d*$_4$) δ 176.30, 172.36, 136.63, 135.49, 129.40, 127.94, 127.28, 124.35, 123.02, 121.77, 121.05, 119.73, 118.41, 117.80, 110.97, 110.56, 109.05, 100.83, 56.47, 52.71, 51.93, 51.29, 27.08, 18.00.

IR (neat): 3257, 2929, 1594, 1354, 1023,742 cm^{-1}

MS (ESI): m/z = 419.2077 (MH$^+$), calculated for C$_{24}$H$_{27}$O$_3$N$_4$ = 419.2083

B27: (*R*)-methyl 2-((*S*)-2-(3-hydroxybenzylamino)-3-phenylpropanamido)-3-(1*H*-indol-3-yl)propanoate

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 7.48 (dt, *J* = 7.9, 1.0 Hz, 1H), 7.31 (dt, *J* = 8.2, 1.0 Hz, 1H), 7.16 (dd, *J* = 5.0, 1.9 Hz, 3H), 7.08 (ddd, *J* = 8.1, 7.0, 1.2 Hz, 1H), 7.05 – 7.02 (m, 2H), 7.02 – 6.96 (m, 2H), 6.90 (s, 1H), 6.62 (ddd, *J* = 8.2, 2.5, 1.0 Hz, 1H), 6.54 – 6.49 (m, 1H), 6.43 (dt, *J* = 7.5, 1.4 Hz, 1H), 4.72 (dd, *J* = 7.8, 5.5 Hz, 1H), 3.68 (s, 3H), 3.43 (d, *J* = 13.1 Hz, 1H), 3.32 (d, *J* = 5.7 Hz, 1H), 3.27 – 3.17 (m, 2H), 3.10 (dd, *J* = 14.7, 7.8 Hz, 1H), 2.88 (dd, *J* = 13.5, 6.3 Hz, 1H), 2.71 (dd, *J* = 13.5, 7.6 Hz, 1H).

^{13}C NMR (100 MHz, Methanol-*d*$_4$) δ 174.76, 172.41, 157.03, 140.58, 137.11, 136.67, 128.95, 128.83, 128.06, 127.18, 126.32, 123.02, 121.13, 119.01, 118.49, 117.71, 114.70, 113.56, 111.00, 108.90, 62.37, 52.87, 51.36, 51.18, 38.86, 26.89.

IR (neat): 3331, 1740, 1652, 1456, 1218, 744 cm^{-1}

MS (ESI): m/z = 472.2232 (MH$^+$), calculated for C$_{28}$H$_{30}$O$_4$N$_3$ = 472.2231

B28: (*S*)-ethyl 2-((*S*)-2-(3-hydroxybenzylamino)-3-phenylpropanamido)-3-phenyl-propanoate

^{1}H NMR (400 MHz, Methanol-d_4) δ 7.30 – 7.17 (m, 6H), 7.17 – 7.12 (m, 2H), 7.12 – 7.06 (m, 2H), 7.02 (t, J = 7.8 Hz, 1H), 6.62 (ddd, J = 8.1, 2.5, 1.0 Hz, 1H), 6.58 – 6.52 (m, 1H), 6.48 (dt, J = 7.5, 1.5 Hz, 1H), 4.73 (dd, J = 8.5, 5.5 Hz, 1H), 4.13 (qd, J = 7.2, 3.2 Hz, 2H), 3.43 (d, J = 13.0 Hz, 1H), 3.30 – 3.25 (m, 2H), 3.11 (dd, J = 13.8, 5.5 Hz, 1H), 2.98 (dd, J = 13.8, 8.4 Hz, 1H), 2.88 (dd, J = 13.6, 5.9 Hz, 1H), 2.68 (dd, J = 13.6, 7.8 Hz, 1H), 1.20 (t, J = 7.1 Hz, 3H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 174.46, 171.16, 157.05, 140.54, 137.09, 136.43, 128.96, 128.93, 128.89, 128.16, 128.10, 126.62, 126.37, 118.99, 114.76, 113.56, 62.70, 53.37, 53.08, 51.31, 37.47, 37.14, 13.04.

IR (neat): 3327, 1736, 1652, 1454, 1280, 1198, 782, 699 cm–1

MS (ESI): m/z = 447.2282 (MH$^+$), calculated for $C_{27}H_{31}O_4N_2$ = 447.2278

B29: (*R*)-methyl 2-((*S*)-2-(3-hydroxybenzylamino)-3-phenylpropanamido)-3-phenyl-propanoate

^{1}H NMR (400 MHz, Methanol-d_4) δ 7.24 – 7.16 (m, 6H), 7.09 (tt, J = 6.1, 1.7 Hz, 4H), 7.03 (t, J = 7.8 Hz, 1H), 6.64 (ddd, J = 8.1, 2.6, 1.0 Hz, 1H), 6.56 (t, J = 2.0 Hz, 1H), 6.51 (dt, J = 7.6, 1.3 Hz, 1H), 4.69 (dd, J = 8.5, 5.5 Hz, 1H), 3.70 (s, 3H), 3.49 (d, J = 13.3 Hz, 1H), 3.37 – 3.32 (m, 1H), 3.26 (d, J = 13.2 Hz, 1H), 3.09 (dd, J = 13.9, 5.6 Hz, 1H), 2.90 (ddd, J = 13.6, 7.4, 4.0 Hz, 2H), 2.72 (dd, J = 13.6, 7.9 Hz, 1H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 174.74, 171.83, 157.11, 140.57, 137.14, 136.44, 128.98, 128.86, 128.78, 128.21, 128.13, 126.58, 126.38, 118.93, 114.72, 113.61, 62.42, 53.27, 51.38, 51.27, 37.48, 36.86.

IR (neat): 3326, 1740, 1650, 1454, 1203 cm^{-1}

MS (ESI): m/z = 433.2125 (MH$^+$), calculated for $C_{26}H_{29}O_4N_2$ = 433.2122

B30: Methyl (2*S*)-3-(1*H*-indol-3-yl)-2-[(2*S*)-3-phenyl-2-{[(thiophen-2-yl)methyl]-amino}propanamido]propanoate

^{1}H NMR (400 MHz, Chloroform-d) δ 8.19 (s, 1H), 7.85 (d, J = 8.7 Hz, 1H), 7.54 (d, J = 7.9 Hz, 1H), 7.37 (d, J = 8.1 Hz, 1H), 7.33 – 7.18 (m, 4H), 7.18 – 7.07 (m, 4H), 6.92 (d, J = 2.4 Hz, 1H), 6.86 (dd, J = 5.1, 3.5 Hz, 1H), 6.62 (d, J = 3.4 Hz, 1H), 5.01 (dt, J = 8.7, 5.7 Hz, 1H), 3.83 (d, J = 14.1 Hz, 1H), 3.72 (d, J = 18.1 Hz, 4H), 3.48 – 3.25 (m, 3H), 3.10 (dd, J = 13.9, 4.0 Hz, 1H), 2.59 (dd, J = 13.9, 9.3 Hz, 1H).

^{13}C NMR (101 MHz, CDCl$_3$) δ 173.09, 172.30, 142.62, 137.15, 136.14, 129.15, 128.75, 127.67, 126.94, 126.62, 125.12, 124.53, 122.71, 122.24, 119.60, 118.78, 111.24, 110.21, 62.57, 52.38, 52.32, 46.70, 38.74, 27.80.

IR (neat): 3327, 1741, 1659, 1577, 1439 cm^{-1}

MS (ESI): m/z = 462.1841 (MH$^+$), calculated for C$_{26}$H$_{28}$O$_3$N$_3$S = 462.1846

B31: Methyl (2*S*)-3-(1*H*-indol-3-yl)-2-[(2*S*)-3-phenyl-2-{[(pyridin-3-yl)methyl]-amino}propanamido]propanoate

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.38 – 8.29 (m, 1H), 8.15 (s, 1H), 7.46 (d, *J* = 7.9 Hz, 1H), 7.32 (d, *J* = 8.1 Hz, 1H), 7.29 – 7.14 (m, 6H), 7.14 – 7.05 (m, 3H), 7.04 – 6.95 (m, 2H), 3.69 (s, 3H), 3.40 – 3.34 (m, 1H), 3.30 – 3.27 (m, 1H), 3.25 (d, *J* = 6.5 Hz, 2H), 2.92 (dd, *J* = 13.6, 5.7 Hz, 1H), 2.73 (dd, *J* = 13.6, 7.5 Hz, 1H).

^{13}C NMR (101 MHz, MeOD) δ 174.35, 172.20, 148.37, 147.08, 137.16, 136.74, 136.62, 135.69, 129.04, 128.10, 127.32, 126.41, 123.63, 123.29, 121.20, 118.60, 117.95, 111.08, 108.88, 63.15, 52.49, 51.43, 48.37, 38.73, 26.91.

IR (neat): 3320, 2929, 1745, 1666, 1521, 1462, 747 cm^{-1}

MS (ESI): m/z = 457.2230 (MH$^+$), calculated for C$_{27}$H$_{29}$O$_3$N$_4$ = 457.2234

B32: Methyl (2*S*)-3-hydroxy-2-[(2*S*)-2-{[(3-hydroxyphenyl)methyl]amino}-3-phenyl-propanamido]butanoate

^{1}H NMR (400 MHz, Methanol-d_4) δ 7.33 – 7.13 (m, 5H), 7.04 (dd, *J* = 9.1, 7.2 Hz, 1H), 6.66 – 6.57 (m, 3H), 4.45 (d, *J* = 2.6 Hz, 1H), 4.33 (qd, *J* = 6.4, 2.6 Hz, 1H), 3.73 (s, 3H), 3.49 (d, *J* = 12.9 Hz, 1H), 3.43 (dd, *J* = 8.6, 4.9 Hz, 1H), 3.08 (dd, *J* = 13.7, 4.9 Hz, 1H), 2.85 – 2.75 (m, 1H), 1.12 (d, *J* = 6.4 Hz, 3H).

^{13}C NMR (101 MHz, MeOD) δ 175.49, 170.95, 157.21, 140.66, 137.16, 129.01, 128.88, 128.26, 126.50, 119.04, 114.90, 113.69, 66.94, 62.78, 57.38, 51.64, 51.49, 39.02, 19.04.

IR (neat): 3339, 2929, 1752, 1655, 1525, 1462, 1283 cm^{-1}

MS (ESI): m/z = 387.1919 (MH$^+$), calculated for C$_{21}$H$_{27}$O$_5$N$_2$ = 387.1914

15.9.3.Series C: Piperazine-2,5-diones formed as cyclic dimers

C1: 1,4-bis(pyridin-3-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.54 (d, J = 2.2 Hz, 2H), 8.49 (dd, J = 5.0, 1.6 Hz, 2H), 7.86 – 7.79 (m, 2H), 7.43 (dd, J = 7.9, 4.9 Hz, 2H), 4.66 (s, 4H), 4.07 (s, 4H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 164.70, 148.68, 148.21, 136.86, 132.23, 124.04, 49.12, 46.36.

IR (neat): 1653, 1422, 715, 640 cm^{-1}

MS (ESI): m/z = 297.1342 (MH+), calculated for $C_{16}H_{17}O_2N_4$ = 297.1346

C2: (3*S*,6*S*)-3,6-dimethyl-1,4-bis(pyridin-3-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.51 (d, J = 2.5 Hz, 2H), 8.47 (dd, J = 5.0, 1.7 Hz, 2H), 7.79 (dt, J = 8.0, 2.1 Hz, 2H), 7.42 (dd, J = 7.9, 4.9 Hz, 2H), 4.91 (d, J = 15.4 Hz, 2H), 4.47 (d, J = 15.4 Hz, 2H), 4.12 (q, J = 7.0 Hz, 2H), 1.55 (d, J = 7.0 Hz, 6H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 167.82, 148.35, 148.09, 136.47, 133.19, 124.07, 56.47, 45.25, 18.43.

IR (neat): 1654, 1481, 1429, 828 cm^{-1}

MS (ESI): m/z = 325.1662 (MH$^+$), calculated for $C_{18}H_{21}O_2N_4$ = 325.1659

C3: (3*S*,6*S*)-3,6-dibenzyl-1,4-bis(pyridin-3-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.52 – 8.36 (m, 2H), 8.26 (s, 2H), 7.59 (d, J = 7.7 Hz, 2H), 7.34 (td, J = 19.5, 17.5, 7.1 Hz, 8H), 7.08 (d, J = 7.4 Hz, 4H), 5.05 (d, J = 15.2 Hz, 2H), 4.24 (t, J = 5.0 Hz, 2H), 4.02 (d, J = 15.3 Hz, 2H), 2.93 – 2.78 (m, 2H), 2.37 (dd, J = 14.1, 6.0 Hz, 2H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 166.37, 148.56, 148.15, 136.61, 136.59, 132.51, 129.42, 128.67, 127.14, 123.99, 61.56, 45.39, 38.28.

IR (neat): 1658, 1425, 843 cm^{-1}

MS (ESI): m/z = 477.2293 (MH$^+$), calculated for $C_{30}H_{29}O_2N_4$ = 477.2285

C4: (3*S*,6*S*)-3,6-dibenzyl-1,4-bis(pyridin-4-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.57 – 8.35 (m, 4H), 7.38 – 7.32 (m, 4H), 7.32 – 7.25 (m, 2H), 7.17 – 7.12 (m, 4H), 7.12 – 7.07 (m, 4H), 5.07 (d, J = 15.9 Hz, 2H), 4.28 (dd, J =

6.2, 4.6 Hz, 2H), 4.02 (d, J = 15.9 Hz, 2H), 2.90 (dd, J = 14.4, 4.6 Hz, 2H), 2.39 (dd, J = 14.4, 6.3 Hz, 2H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 166.41, 149.02, 146.33, 136.51, 129.43, 128.69, 127.16, 122.89, 61.88, 47.02, 38.35.

IR (neat): 1660, 1601, 1459, 1414, 702 cm^{-1}

MS (ESI): m/z = 477.2286 (MH$^+$), calculated for $C_{30}H_{29}O_2N_4$ = 477.2285

C5: (3S,6S)-3,6-dibenzyl-1,4-bis(thiophen-2-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, CDCl$_3$) δ 7.35 (dd, J = 8.2, 6.5 Hz, 4H), 7.29 (d, J = 7.1 Hz, 2H), 7.22 (dd, J = 5.3, 1.3 Hz, 2H), 7.12 – 7.06 (m, 4H), 6.88 (dd, J = 5.1, 3.4 Hz, 2H), 6.70 (d, J = 3.4 Hz, 2H), 5.31 (d, J = 15.2 Hz, 2H), 4.23 (dd, J = 6.7, 4.3 Hz, 2H), 3.77 (d, J = 15.3 Hz, 2H), 2.92 (dd, J = 14.2, 4.3 Hz, 2H), 2.39 (dd, J = 14.2, 6.6 Hz, 2H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 165.49, 137.64, 136.80, 129.68, 128.94, 127.60, 127.41, 126.60, 126.32, 60.45, 42.49, 38.93.

IR (neat): 1654, 1454, 699 cm^{-1}

MS (ESI): m/z = 487.1545 (MH$^+$), calculated for $C_{26}H_{27}O_2N_2S_2$ = 487.1508

C8: 1,4-bis(pyridin-4-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.56 – 8.44 (m, 2H), 7.48 – 7.31 (m, 2H), 4.68 (s, 2H), 4.11 (d, J = 2.2 Hz, 2H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 164.83, 149.09, 146.13, 123.00, 49.47, 47.97.

IR (neat): 1664, 1604, 1479, 1417, 1334 cm^{-1}

MS (ESI): m/z = 297.1350 (MH$^+$), calculated for $C_{16}H_{17}N_4O_2$ = 297.1351

15.9.4.Series D: Piperazine-2,5-diones

D1: 1-(pyridin-3-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.55 (s, 1H), 8.49 (d, J = 4.8 Hz, 1H), 7.83 (d, J = 7.9 Hz, 1H), 7.45 (dd, J = 7.9, 5.0 Hz, 1H), 4.67 (s, 2H), 4.03 (s, 2H), 3.97 (s, 2H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 166.55, 165.41, 148.59, 148.15, 136.81, 132.38, 124.06, 48.86, 46.57, 44.22.

IR (neat): 1662, 1476, 1327 cm^{-1}

MS (ESI): m/z = 206.0921 (MH$^+$), calculated for $C_{10}H_{12}O_2N_3$ = 206.0924

D2: (*S*)-6-methyl-1-(pyridin-3-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 8.53 (d, *J* = 2.6 Hz, 1H), 8.47 (dd, *J* = 4.9, 1.6 Hz, 1H), 7.81 (dt, *J* = 8.0, 2.0 Hz, 1H), 7.42 (dd, *J* = 7.9, 4.9 Hz, 1H), 4.99 (d, *J* = 15.4 Hz, 1H), 4.39 (d, *J* = 15.4 Hz, 1H), 4.17 (d, *J* = 17.7 Hz, 1H), 4.01 – 3.83 (m, 2H), 1.46 (d, *J* = 7.0 Hz, 3H).

^{13}C NMR (100 MHz, Methanol-*d*$_4$) δ 169.97, 165.79, 148.38, 148.06, 136.50, 133.26, 124.01, 56.48, 44.98, 43.84, 16.32.

IR (neat): 1649, 1466, 1322, 711 cm^{-1}

MS (ESI): m/z = 220.1086 (MH$^+$), calculated for $C_{11}H_{14}O_2N_3$ = 220.1086

D3: (*S*)-6-benzyl-1-(pyridin-3-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 8.55 (dd, *J* = 2.1, 1.0 Hz, 1H), 8.48 (dd, *J* = 4.9, 1.6 Hz, 1H), 7.87 – 7.79 (m, 1H), 7.48 – 7.39 (m, 1H), 7.35 – 7.27 (m, 3H), 7.20 – 7.13 (m, 2H), 5.27 (d, *J* = 15.2 Hz, 1H), 4.30 – 4.16 (m, 2H), 3.43 (dd, *J* = 17.7, 0.7 Hz, 1H), 3.36 – 3.28 (m, 2H), 3.20 (dd, *J* = 14.1, 3.7 Hz, 1H), 2.46 – 2.34 (m, 1H).

^{13}C NMR (100 MHz, Methanol-*d*$_4$) δ 168.43, 166.04, 148.64, 148.19, 136.71, 134.71, 132.87, 129.86, 128.39, 127.46, 124.07, 61.18, 44.74, 43.22, 36.27.

IR (neat): 1656, 1455, 1324, 702 cm^{-1}

MS (ESI): m/z = 296.1398 (MH$^+$), calculated for $C_{17}H_{18}O_2N_3$ = 296.1399

D4: (3*S*,6*S*)-6-benzyl-3-methyl-1-(pyridin-3-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 8.57 (d, *J* = 2.5 Hz, 1H), 8.48 (dd, *J* = 4.9, 1.6 Hz, 1H), 7.85 (dt, *J* = 8.0, 1.9 Hz, 1H), 7.43 (dd, *J* = 7.9, 4.9 Hz, 1H), 7.34 – 7.25 (m, 3H), 7.14 (dt, *J* = 6.0, 1.6 Hz, 2H), 5.33 (d, *J* = 15.1 Hz, 1H), 4.34 (d, *J* = 15.2 Hz, 1H), 4.26 (dd, *J* = 4.7, 3.4 Hz, 1H), 3.86 (q, *J* = 7.1 Hz, 1H), 3.38 (dd, *J* = 14.3, 4.7 Hz, 1H), 3.25 (dd, *J* = 14.2, 3.6 Hz, 1H), 0.42 (d, *J* = 7.1 Hz, 3H).

^{13}C NMR (100 MHz, Methanol-*d*$_4$) δ 168.11, 166.39, 148.76, 148.14, 136.80, 134.96, 132.77, 130.16, 128.46, 127.27, 124.05, 60.67, 50.64, 44.63, 35.96, 19.09.

IR (neat): 1655, 1454, 1425, 1321, 704 cm^{-1}

MS (ESI): m/z = 310.1552 (MH$^+$), calculated for $C_{18}H_{20}O_2N_3$ = 310.1550

D5: (3*S*,6*S*)-3-((1*H*-indol-3-yl)methyl)-6-methyl-1-(pyridin-3-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.39 (s, 2H), 7.57 (d, J = 7.9 Hz, 1H), 7.50 (d, J = 7.9 Hz, 1H), 7.33 (d, J = 8.1 Hz, 1H), 7.27 (d, J = 6.3 Hz, 1H), 7.08 (d, J = 9.2 Hz, 2H), 7.02 (t, J = 7.6 Hz, 1H), 4.71 (d, J = 15.4 Hz, 1H), 4.40 (d, J = 4.6 Hz, 1H), 4.16 (d, J = 15.5 Hz, 1H), 3.67 (q, J = 7.7 Hz, 1H), 3.20 – 3.14 (m, 1H), 1.32 – 1.27 (m, 1H), 0.33 (d, J = 7.0 Hz, 3H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 168.62, 167.00, 148.19, 147.70, 136.43, 136.39, 133.03, 127.66, 124.47, 123.84, 121.23, 118.82, 118.56, 110.94, 107.98, 56.38, 55.60, 45.15, 29.68, 16.86.

IR (neat): 3287, 1656, 1112, 611 cm^{-1}

MS (ESI): m/z = 349.1657 (MH$^+$), calculated for $C_{20}H_{21}O_2N_4$ = 349.1659

D6: (3*R*,6*S*)-3-((1*H*-indol-3-yl)methyl)-6-benzyl-1-(pyridin-3-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.34 – 8.27 (m, 1H), 8.23 (d, J = 2.3 Hz, 1H), 7.50 (d, J = 8.0 Hz, 1H), 7.33 (dt, J = 16.4, 7.4 Hz, 4H), 7.14 (dd, J = 7.4, 1.9 Hz, 2H), 7.07 (t, J = 7.6 Hz, 1H), 6.99 – 6.87 (m, 3H), 6.60 – 6.49 (m, 1H), 5.26 (d, J = 15.3 Hz, 1H), 4.06 (d, J = 15.3 Hz, 1H), 3.81 (t, J = 3.9 Hz, 1H), 3.28 – 3.20 (m, 2H), 3.09 (dd, J = 14.1, 3.2 Hz, 1H), 2.97 – 2.85 (m, 2H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 167.65, 167.45, 148.25, 147.71, 136.27, 135.84, 134.45, 132.17, 129.97, 128.28, 127.47, 127.42, 124.14, 123.89, 121.03, 118.77, 118.55, 110.77, 107.78, 60.16, 53.98, 43.61, 36.01, 27.70.

IR (neat): 1668, 1454, 1425, 743, 703 cm^{-1}

MS (ESI): m/z = 425.1970 (MH$^+$), calculated for $C_{26}H_{25}O_2N_4$ = 425.1972

D7: (3*S*,6*S*)-3-(4-hydroxybenzyl)-6-methyl-1-(pyridin-3-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.51 (s, 1H), 8.46 (d, J = 4.8 Hz, 1H), 7.74 (dt, J = 8.0, 1.8 Hz, 1H), 7.40 (dd, J = 7.9, 4.9 Hz, 1H), 6.98 (d, J = 8.1 Hz, 2H), 6.71 (d, J = 8.1 Hz, 2H), 4.77 (d, J = 15.3 Hz, 1H), 4.43 – 4.31 (m, 2H), 3.78 (q, J = 7.0 Hz, 1H), 3.22 (dd, J = 14.0, 4.1 Hz, 1H), 2.90 (dd, J = 13.9, 4.5 Hz, 1H), 0.64 (d, J = 7.0 Hz, 3H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 168.68, 166.43, 156.91, 148.55, 147.87, 136.72, 133.19, 131.29, 125.60, 123.82, 115.13, 56.50, 55.89, 45.47, 38.67, 17.41.

IR (neat): 1652, 1514, 1463 cm^{-1}

MS (ESI): m/z = 326.1497 (MH$^+$), calculated for C$_{18}$H$_{20}$O$_3$N$_3$ = 326.1499

D8: (*S*)-6-benzyl-1-(pyridin-4-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.51 (s, 2H), 7.33 (tdd, J = 6.7, 4.5, 1.7 Hz, 5H), 7.24 – 7.13 (m, 2H), 5.28 (d, J = 15.9 Hz, 1H), 4.30 – 4.11 (m, 2H), 3.49 (d, J = 17.7 Hz, 1H), 3.30 – 3.26 (m, 1H), 3.21 (dd, J = 14.1, 3.9 Hz, 1H), 2.52 (d, J = 17.5 Hz, 1H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 168.43, 166.17, 149.14, 146.54, 134.77, 129.82, 128.40, 127.43, 122.85, 61.65, 46.33, 43.33, 36.39.

IR (neat): 1665, 1456, 1415, 1325 cm^{-1}

MS (ESI): m/z = 296.1396 (MH$^+$), calculated for C$_{17}$H$_{18}$O$_2$N$_3$ = 296.1394

D9: (3*R*,6*S*)-3-((1*H*-indol-3-yl)methyl)-6-benzyl-1-(thiophen-2-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 7.42 (d, J = 7.9 Hz, 1H), 7.30 (dd, J = 7.6, 4.0 Hz, 2H), 7.25 (td, J = 6.7, 5.7, 4.1 Hz, 3H), 7.10 – 7.02 (m, 3H), 7.00 – 6.94 (m, 1H), 6.92 (s, 1H), 6.85 (dd, J = 5.2, 3.5 Hz, 1H), 6.69 (d, J = 3.4 Hz, 1H), 5.31 (d, J = 15.2 Hz, 1H), 4.35 (d, J = 15.2 Hz, 1H), 4.00 (t, J = 4.0 Hz, 1H), 3.25 (dd, J = 14.1, 4.7 Hz, 1H), 3.13 – 2.98 (m, 3H), 2.88 (t, J = 5.0 Hz, 1H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 168.04, 167.03, 137.65, 136.54, 134.58, 129.89, 128.24, 127.36, 127.33, 126.35, 125.72, 123.85, 121.02, 118.44, 118.40, 110.74, 107.80, 104.99, 60.04, 53.51, 41.52, 36.03, 27.60.

IR (neat): 3346, 1678, 1454, 738, 701 cm^{-1}

MS (ESI): m/z = 430.1587 (MH$^+$), calculated for C$_{25}$H$_{24}$O$_2$N$_3$S = 430.1589

D10: (3*S*,6*S*)-6-benzyl-3-methyl-1-(thiophen-2-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, CDCl$_3$) δ 7.41 (d, *J* = 3.0 Hz, 1H), 7.33 – 7.23 (m, 3H + CDCl$_3$), 7.10 – 7.05 (m, 2H), 7.01 (dd, *J* = 3.5, 1.3 Hz, 1H), 6.95 (dd, *J* = 5.1, 3.5 Hz, 1H), 5.53 (dd, *J* = 15.0, 0.9 Hz, 1H), 4.25 (t, *J* = 4.3 Hz, 1H), 4.16 (d, *J* = 15.1 Hz, 1H), 3.95 – 3.83 (m, 1H), 3.33 – 3.18 (m, 2H), 0.58 (d, *J* = 7.1 Hz, 3H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 166.76, 166.56, 137.41, 134.84, 130.21, 128.83, 127.95, 127.59, 126.81, 126.46, 59.58, 51.13, 41.90, 36.72, 20.66.

IR (neat): 3241, 2361, 1682, 1655, 1454, 1321, 701 cm^{-1}

MS (ESI): m/z = 315.1170 (MH$^+$), calculated for C$_{17}$H$_{19}$O$_7$N$_2$S = 315.1167

D11: (3*S*,6*S*)-3-((1*H*-indol-3-yl)methyl)-6-benzyl-1-(furan-2-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 10.45 (s, NH), 7.55 (dd, *J* = 7.9, 1.2 Hz, 1H), 7.41 (dd, *J* = 1.9, 0.8 Hz, 1H), 7.36 – 7.31 (m, 1H), 7.27 – 7.17 (m, 3H), 7.13 (ddd, *J* = 8.2, 7.0, 1.2 Hz, 1H), 7.05 (ddd, *J* = 8.0, 6.9, 1.1 Hz, 1H), 6.96 (d, *J* = 1.8 Hz, 1H), 6.79 – 6.70 (m, 2H), 6.32 (dd, *J* = 3.2, 1.9 Hz, 1H), 6.15 (d, *J* = 3.2 Hz, 1H), 5.06 (d, *J* = 15.5 Hz, 1H), 4.16 (dd, *J* = 7.1, 3.9 Hz, 1H), 3.99 (dd, *J* = 6.1, 4.6 Hz, 1H), 3.70 (d, *J* = 15.5 Hz, 1H), 3.03 – 2.91 (m, 1H), 2.70 (dd, *J* = 14.2, 4.6 Hz, 1H), 2.31 (dd, *J* = 14.5, 7.0 Hz, 1H), 2.14 – 2.02 (m, 1H).

^{13}C NMR (100 MHz, Methanol-*d*$_4$) δ 167.27, 166.90, 148.83, 142.67, 136.71, 136.47, 129.47, 128.34, 127.29, 126.82, 124.19, 121.30, 118.77, 118.31, 111.16, 110.04, 109.53, 108.55, 60.45, 56.13, 40.22, 38.08, 30.14.

IR (neat): 3305, 1670, 1455, 744 cm^{-1}

MS (ESI): m/z = 414.1811 (MH$^+$), calculated for C$_{25}$H$_{24}$O$_3$N$_3$ = 414.1812

D12:   (3*S*,6*S*)-6-benzyl-1-(furan-2-ylmethyl)-3-(4-hydroxybenzyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 7.49 – 7.43 (m, 1H), 7.35 (dd, *J* = 8.3, 6.7 Hz, 2H), 7.30 – 7.23 (m, 1H), 7.13 – 7.06 (m, 2H), 6.89 – 6.81 (m, 2H), 6.77 – 6.69 (m, 2H), 6.38 (dd, *J* = 3.2, 1.9 Hz, 1H), 6.32 (d, *J* = 3.2 Hz, 1H), 5.14 (d, *J* = 15.5 Hz, 1H), 4.17 – 4.07

(m, 2H), 3.97 (dd, J = 8.6, 3.8 Hz, 1H), 3.01 (dd, J = 14.3, 4.9 Hz, 1H), 2.71 (dd, J = 14.2, 5.0 Hz, 1H), 2.61 (dd, J = 13.8, 3.8 Hz, 1H), 1.55 (dd, J = 13.8, 8.6 Hz, 1H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 167.06, 166.48, 156.32, 148.94, 142.72, 136.06, 130.43, 129.82, 128.55, 127.06, 126.67, 115.08, 110.16, 109.73, 60.42, 56.96, 40.22, 39.41, 37.07.

IR (neat): 3272, 2361, 1669, 1516, 1456, 702 cm^{-1}

MS (ESI): m/z = 391.1654 (MH$^+$), calculated for $C_{23}H_{23}O_4N_2$ =391.1652

D13: (*S*)-6-benzyl-1-(furan-2-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 7.52 (d, J = 5.6 Hz, 1H), 7.36 – 7.18 (m, 3H), 7.17 – 6.99 (m, 2H), 6.51 – 6.34 (m, 2H), 5.06 (d, J = 15.6 Hz, 1H), 4.38 (d, J = 15.5 Hz, 1H), 4.27 – 4.13 (m, 1H), 3.37 (d, J = 17.7 Hz, 1H), 3.19 (tdd, J = 17.9, 11.2, 4.4 Hz, 2H), 2.34 (d, J = 17.6 Hz, 1H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 168.61, 165.50, 149.10, 142.84, 134.77, 129.86, 128.31, 127.36, 110.31, 109.78, 60.86, 43.27, 40.12, 36.21.

IR (neat): 3247, 1660, 1455, 1323, 702 cm^{-1}

MS (ESI): m/z = 285.1236 (MH$^+$), calculated for $C_{16}H_{17}O_3N_2$ = 285.1234

D14: (3*S*,6*S*)-6-benzyl-1-(furan-2-ylmethyl)-3-methylpiperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 7.51 (dd, J = 2.1, 1.1 Hz, 1H), 7.35 – 7.21 (m, 3H), 7.12 – 7.03 (m, 2H), 6.47 (d, J = 3.3 Hz, 1H), 6.42 (dd, J = 3.3, 1.8 Hz, 1H), 5.21 (d, J = 15.5 Hz, 1H), 4.39 (d, J = 15.4 Hz, 1H), 4.21 (t, J = 4.1 Hz, 1H), 3.80 (q, J = 7.1 Hz, 1H), 3.38 – 3.34 (m, 1H), 3.22 (dd, J = 14.2, 3.6 Hz, 1H), 0.39 (dd, J = 7.1, 1.2 Hz, 3H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 167.62, 166.55, 149.06, 142.76, 134.99, 130.13, 128.36, 127.16, 110.24, 109.82, 60.31, 50.67, 39.93, 35.73, 19.12.

IR (neat): 3255, 1655, 1453, 1322, 702 cm^{-1}

MS (ESI): m/z = 299.1393 (MH$^+$), calculated for $C_{17}H_{19}O_3N_2$ = 299.1390

D15: (*S*)-6-benzyl-1-(3-hydroxybenzyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 7.30 (qd, J = 4.8, 2.2 Hz, 3H), 7.20 – 7.11 (m, 3H), 6.73 (dq, J = 6.6, 1.5 Hz, 3H), 5.23 (dd, J = 14.9, 1.6 Hz, 1H), 4.14 – 4.06 (m, 1H), 3.99

(dd, J = 14.9, 1.6 Hz, 1H), 3.45 (dd, J = 17.5, 1.7 Hz, 1H), 3.25 (ddd, J = 14.0, 4.9, 1.7 Hz, 1H), 3.13 (ddd, J = 14.0, 4.1, 1.6 Hz, 1H), 2.53 (d, J = 17.6 Hz, 1H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 168.78, 165.79, 158.18, 137.02, 134.88, 129.85, 129.58, 128.33, 127.37, 118.67, 114.81, 114.69, 60.26, 46.61, 43.31, 35.98.

IR (neat): 3248, 2361, 1652, 1455, 702 cm^{-1}

MS (ESI): m/z = 311.1396 (MH$^+$), calculated for $C_{18}H_{19}O_3N_2$ = 311.1390

D16: (3S,6S)-6-benzyl-1-(3-hydroxybenzyl)-3-methylpiperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 7.30 (dt, J = 13.5, 6.9 Hz, 3H), 7.23 – 7.05 (m, 3H), 6.77 (q, J = 7.6 Hz, 3H), 5.38 (d, J = 14.8 Hz, 1H), 4.12 (dd, J = 12.1, 8.2 Hz, 2H), 3.87 (q, J = 7.0 Hz, 1H), 3.40 – 3.33 (m, 1H), 3.19 (dd, J = 14.3, 3.9 Hz, 1H), 0.45 (d, J = 7.0 Hz, 3H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 167.91, 166.73, 157.75, 136.95, 135.10, 130.14, 129.62, 128.42, 127.20, 118.98, 114.64, 114.55, 59.69, 50.75, 46.37, 35.64, 19.22.

IR (neat): 3203, 2361, 1671, 1454, 705 cm^{-1}

MS (ESI): m/z = 325.1553 (MH$^+$), calculated for $C_{19}H_{21}O_3N_2$ = 325.1547

D17: (3S,6S)-3-((1H-indol-3-yl)methyl)-6-benzyl-1-(3-hydroxybenzyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 7.58 (dt, J = 7.8, 1.0 Hz, 1H), 7.35 (dt, J = 8.1, 0.9 Hz, 1H), 7.28 – 7.18 (m, 3H), 7.14 (ddd, J = 8.1, 7.0, 1.2 Hz, 1H), 7.09 – 7.04 (m, 2H), 7.00 (s, 1H), 6.78 – 6.72 (m, 2H), 6.67 (ddd, J = 8.2, 2.5, 0.9 Hz, 1H), 6.51 (t, J = 2.0 Hz, 1H), 6.46 (dt, J = 7.6, 1.2 Hz, 1H), 5.21 (d, J = 15.0 Hz, 1H), 4.25 (dd, J = 6.9, 3.9 Hz, 1H), 3.93 – 3.84 (m, 1H), 3.42 (d, J = 15.0 Hz, 1H), 3.01 (ddd, J = 14.6, 4.0, 0.8 Hz, 1H), 2.69 (dd, J = 14.2, 4.4 Hz, 1H), 2.40 (dd, J = 14.5, 6.9 Hz, 1H), 1.99 (dd, J = 14.2, 6.3 Hz, 1H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 167.52, 167.18, 157.72, 136.76, 136.73, 136.67, 129.44, 129.41, 128.36, 127.35, 126.81, 124.23, 121.33, 118.81, 118.68, 118.37, 114.55, 114.34, 111.19, 108.58, 59.96, 56.22, 46.73, 38.12, 30.14.

IR (neat): 3317, 2361, 1669, 1456 745, 701 cm^{-1}

MS (ESI): m/z = 462.1789 (MNa$^+$), calculated for $C_{27}H_{25}O_3N_3Na$ = 462.1788

D18: (3*S*,6*S*)-6-benzyl-1-(3-hydroxybenzyl)-3-(4-hydroxybenzyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 7.36 (td, *J* = 8.5, 7.8, 2.0 Hz, 2H), 7.32 – 7.24 (m, 1H), 7.13 (ddd, *J* = 8.0, 4.7, 3.2 Hz, 3H), 6.91 – 6.85 (m, 2H), 6.78 – 6.68 (m, 3H), 6.66 – 6.60 (m, 2H), 5.30 (dd, *J* = 15.0, 2.3 Hz, 1H), 4.09 – 4.05 (m, 1H), 4.03 (t, *J* = 5.3 Hz, 1H), 3.82 (d, *J* = 14.9 Hz, 1H), 3.01 (dd, *J* = 14.3, 4.6 Hz, 1H), 2.70 – 2.57 (m, 2H), 1.65 (dd, *J* = 13.7, 8.5 Hz, 1H).

^{13}C NMR (100 MHz, Methanol-*d*$_4$) δ 167.28, 166.76, 157.70, 156.37, 136.83, 136.22, 130.49, 129.77, 129.53, 128.58, 127.07, 126.69, 118.87, 115.14, 114.60, 114.46, 59.84, 57.04, 46.69, 39.40, 37.09.

IR (neat): 3267, 2361, 1669, 1131, 702 cm^{-1}

MS (ESI): m/z = 439.1627 (MNa$^+$), calculated for C$_{25}$H$_{24}$O$_4$N$_2$Na = 439.1628

D19: (*S*)-3-((1*H*-indol-3-yl)methyl)-1-(pyridin-4-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 8.30 – 8.19 (m, 2H), 7.56 (d, *J* = 7.9 Hz, 1H), 7.39 (d, *J* = 8.1 Hz, 1H), 7.19 – 7.09 (m, 1H), 7.07 – 6.97 (m, 2H), 6.74 (d, *J* = 5.4 Hz, 2H), 4.41 (t, *J* = 4.0 Hz, 1H), 4.32 (d, *J* = 15.7 Hz, 1H), 4.23 (d, *J* = 15.7 Hz, 1H), 3.55 (dd, *J* = 14.7, 3.9 Hz, 1H), 3.39 (d, *J* = 17.4 Hz, 1H), 3.14 (dd, *J* = 14.7, 4.2 Hz, 1H), 2.70 (d, *J* = 17.4 Hz, 1H).

^{13}C NMR (100 MHz, Methanol-*d*$_4$) δ 167.30, 166.18, 148.89, 145.45, 136.55, 127.38, 124.64, 122.61, 121.43, 118.95, 118.43, 111.03, 107.70, 56.24, 48.59, 48.06, 29.84.

IR (neat): 3256, 1659, 1417, 1324, 744 cm^{-1}

MS (ESI): m/z = 335.1503 (MH$^+$), calculated for C$_{19}$H$_{18}$O$_2$N$_4$ = 335.1503

D20: (*R*)-3-((1*H*-indol-3-yl)methyl)-1-(pyridin-4-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 8.24 (dd, *J* = 4.8, 1.6 Hz, 2H), 7.56 (d, *J* = 8.0 Hz, 1H), 7.39 (d, *J* = 8.2 Hz, 1H), 7.13 (t, *J* = 7.6 Hz, 1H), 7.07 – 6.97 (m, 2H), 6.78 – 6.68 (m, 2H), 4.40 (t, *J* = 4.1 Hz, 1H), 4.32 (d, *J* = 15.7 Hz, 1H), 4.22 (d, *J* = 15.7 Hz, 1H), 3.55 (dd, *J* = 14.7, 3.9 Hz, 1H), 3.39 (d, *J* = 17.4 Hz, 1H), 3.13 (dd, *J* = 14.7, 4.2 Hz, 1H), 2.77 – 2.63 (m, 1H).

^{13}C NMR (100 MHz, Methanol-*d*$_4$) δ 167.28, 166.18, 148.80, 145.54, 136.47, 127.33, 124.62, 122.60, 121.42, 118.94, 118.43, 111.02, 107.54, 56.16, 48.46, 47.90, 29.80.

IR (neat): 1658, 1417, 1326, 745 cm^{-1}

MS (ESI): m/z = 335.1506 (MH$^+$), calculated for $C_{19}H_{18}O_2N_4$ = 335.1503

D21: 1-(pyridin-4-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.50 (d, J = 4.5 Hz, 2H), 7.34 (dd, J = 4.4, 2.3 Hz, 2H), 4.67 – 4.63 (m, 2H), 4.06 – 4.02 (m, 2H), 3.99 – 3.95 (m, 2H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 166.54, 165.58, 149.13, 146.15, 122.86, 49.32, 48.13, 44.25.

IR (neat): 1664, 1473, 1326 cm^{-1}

MS (ESI): m/z = 206.0920 (MH$^+$), calculated for $C_{10}H_{12}O_2N_3$ = 206.0924

D22: (*S*)-3-(4-hydroxybenzyl)-1-(pyridin-4-ylmethyl)piperazine-2,5-dione

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.47 (d, J = 5.7 Hz, 2H), 7.12 (dd, J = 5.1, 2.2 Hz, 2H), 6.93 (d, J = 8.0 Hz, 2H), 6.62 (d, J = 8.0 Hz, 2H), 4.63 (s, 1H), 4.32 (d, J = 15.6 Hz, 2H), 3.55 (d, J = 17.4 Hz, 1H), 3.23 – 3.10 (m, 1H), 2.98 – 2.81 (m, 2H).

^{13}C NMR (100 MHz, Acetone-d_6) δ 165.99, 164.57, 156.68, 149.88, 144.82, 131.19, 126.24, 122.92, 115.18, 56.62, 48.98, 48.17, 39.22.

IR (neat): 1664, 1514, 1418 cm^{-1}

MS (ESI): m/z = 312.1348 (MH$^+$), calculated for $C_{17}H_{18}O_3N_3$ = 312.1343

15.9.5. Tartaric acid anhydride (E)

E1: (3*R*,4*R*)-2,5-dioxotetrahydrofuran-3,4-diyl diacetate

^{1}H NMR (400 MHz, CDCl$_3$) δ 5.68 (s, 2H), 2.23 (s, 6H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 169.66, 163.25, 72.04, 20.05.

15.9.6. Tartaric acid monoamide (F)

F1: (2*R*,3*R*)-2,3-diacetoxy-4-oxo-4-(pyridin-3-ylamino)butanoic acid

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.76 (d, J = 2.5 Hz, 1H), 8.32 (dd, J = 4.9, 1.5 Hz, 1H), 8.11 (dt, J = 8.5, 1.9 Hz, 1H), 7.46 (dd, J = 8.4, 4.9 Hz, 1H), 5.72 (d, J = 2.8 Hz, 1H), 5.69 (d, J = 2.9 Hz, 1H), 2.20 (s, 3H), 2.10 (s, 3H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 170.01, 169.95, 168.40, 166.14, 144.11, 140.75, 135.18, 129.08, 124.14, 72.63, 71.15, 19.00, 18.90.

F2: (2*R*,3*R*)-2,3-diacetoxy-4-oxo-4-((pyridin-4-ylmethyl)amino)butanoic acid

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.63 – 8.40 (m, 2H), 7.59 – 7.35 (m, 2H), 5.69 (d, *J* = 2.6 Hz, 1H), 5.57 (d, *J* = 2.6 Hz, 1H), 4.55 (d, *J* = 16.4 Hz, 1H), 4.46 (d, *J* = 16.4 Hz, 1H), 2.17 (s, 3H), 2.07 (s, 3H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 170.08, 169.89, 168.86, 168.04, 151.59, 146.97, 122.98, 72.64, 71.66, 41.57, 19.08, 19.03.

F3: (2*R*,3*R*)-2,3-diacetoxy-4-morpholino-4-oxobutanoic acid

^{1}H NMR (400 MHz, Methanol-d_4) δ 5.88 – 5.83 (m, 1H), 5.65 – 5.60 (m, 1H), 3.79 – 3.53 (m, 8H), 2.13 (s, 3H), 2.13 (s, 3H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 170.13, 169.96, 168.08, 165.15, 69.74, 69.61, 66.18, 66.02, 45.93, 42.65, 18.97, 18.84.

F4: (2*R*,3*R*)-2,3-diacetoxy-4-(diethylamino)-4-oxobutanoic acid

^{1}H NMR (400 MHz, CDCl$_3$) δ 12.14 (s, 1H), 5.76 (d, *J* = 4.0 Hz, 1H), 5.52 (d, *J* = 3.9 Hz, 1H), 3.49 (td, *J* = 15.2, 14.4, 7.4 Hz, 1H), 3.36 (dt, *J* = 14.9, 7.2 Hz, 1H), 3.13 (dq, *J* = 14.0, 7.0 Hz, 1H), 3.01 (h, *J* = 6.9 Hz, 1H), 2.10 (s, 3H), 2.09 (s, 3H), 1.25 (q, *J* = 4.0 Hz, 3H), 1.04 (t, *J* = 7.1 Hz, 3H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 170.17, 170.01, 169.04, 165.24, 70.26, 69.36, 42.06, 41.04, 20.39, 20.38, 13.71, 12.33.

F5: (2*R*,3*R*)-2,3-diacetoxy-4-oxo-4-(piperidin-1-yl)butanoic acid

^{1}H NMR (400 MHz, Methanol-d_4) δ 5.87 (dd, *J* = 3.6, 0.8 Hz, 1H), 5.58 (dd, *J* = 3.6, 0.8 Hz, 1H), 3.63 – 3.43 (m, 4H), 2.12 (s, 3H), 2.12 (s, 3H), 1.72 – 1.49 (m, 6H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 170.04, 169.92, 168.83, 164.14, 69.97, 68.92, 46.87, 44.01, 26.06, 25.31, 24.28, 20.44, 20.41.

15.9.7. Tartaric acid bisamides (G)

G1: (2*R*,3*R*)-2,3-dihydroxy-4-morpholino-4-oxo-*N*-(pyridin-3-yl)butanamide

^{1}H NMR (400 MHz, Acetic Acid-d_4) δ 9.16 (d, J = 13.3 Hz, 1H), 8.76 – 8.60 (m, 1H), 8.60 – 8.48 (m, 1H), 7.76 (dd, J = 8.6, 4.8 Hz, 1H), 5.17 (d, J = 2.2 Hz, 1H), 4.66 (d, J = 2.2 Hz, 1H), 3.80 – 3.60 (m, 8H).

^{13}C NMR (100 MHz, Acetic Acid-d_4) δ 171.74, 170.23, 140.10, 136.41, 136.31, 132.73, 125.89, 72.44, 69.38, 66.15, 66.08, 45.82, 42.85.

IR (neat): 3315, 1634, 1527, 1424, 1111 cm^{-1}

MS (ESI): m/z = 296.1245 (MH$^+$), calculated for C$_{13}$H$_{18}$O$_5$N$_3$ = 296.1246

G2: (2*R*,3*R*)-2,3-dihydroxy-4-morpholino-4-oxo-*N*-(pyridin-2-yl)butanamide

^{1}H NMR (400 MHz, Acetic Acid-d_4) δ 8.33 (d, J = 9.0 Hz, 2H), 8.01 (d, J = 8.5 Hz, 1H), 7.29 (s, 1H), 5.15 (s, 1H), 4.66 (s, 1H), 3.87 – 3.65 (m, 8H).

^{13}C NMR (100 MHz, Acetic Acid-d_4) δ 171.30, 169.92, 149.92, 145.10, 141.02, 120.57, 115.38, 72.21, 69.70, 69.46, 66.16, 45.71, 42.92.

IR (neat): 3347, 3195, 1656, 1436, 1389, 1113 cm^{-1}

MS (ESI): m/z = 296.1246 (MH$^+$), calculated for C$_{13}$H$_{18}$O$_5$N$_3$ = 296.1246

G3: (2*R*,3*R*)-*N*-(2-fluoropyridin-4-yl)-2,3-dihydroxy-4-morpholino-4-oxobutanamide

^{1}H NMR (400 MHz, Acetic Acid-d_4) δ 8.18 (dt, J = 6.2, 3.1 Hz, 1H), 7.64 – 7.47 (m, 2H), 5.21 (s, 1H), 4.63 (d, J = 7.7 Hz, 1H), 3.79 – 3.69 (m, 8H).

^{13}C NMR (100 MHz, Acetic Acid-d_4) δ 171.48, 170.49, 165.53, 163.16, 149.52, 149.37, 147.20, 147.08, 112.27, 99.29, 98.88, 72.48, 69.30, 66.08, 52.45, 45.90, 42.94.

IR (neat): 3194, 1609, 1509, 1264, 1110 cm^{-1}

MS (ESI): m/z = 314.1150 (MH$^+$), calculated for C$_{13}$H$_{17}$O$_5$N$_3$F = 314.1152

15.9.8. Cyclic tartrimides (H)

H1: (3*R*,4*R*)-2,5-dioxo-1-(pyridin-4-ylmethyl)pyrrolidine-3,4-diyl diacetate

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.90 – 8.80 (m, 2H), 8.08 (d, J = 6.4 Hz, 2H), 5.76 (d, J = 1.1 Hz, 2H), 5.09 (d, J = 3.0 Hz, 2H), 2.18 (s, 6H).

^{13}C NMR (100 MHz, Methanol-d_4) δ 170.49, 169.54, 156.67, 141.49, 125.67, 73.01, 41.10, 18.77.

IR (neat): 3327, 2361, 1740, 1653, 1456, 740, 700 cm^{-1}

MS (ESI): m/z = 307.0925 (MH$^+$), calculated for C$_{14}$H$_{15}$O$_6$N$_2$ = 307.0925

H2: (3*R*,4*R*)-4-(acetyloxy)-1-(4-methylphenyl)-2,5-dioxopyrrolidin-3-yl acetate

^{1}H NMR (400 MHz, CDCl$_3$) δ 7.32 (d, *J* = 8.3 Hz, 2H), 7.25 (d, *J* = 8.4 Hz, 2H), 5.69 (s, 2H), 2.43 (s, 3H), 2.27 (s, 6H).

^{13}C NMR (101 MHz, CDCl$_3$) δ 170.12, 168.20, 136.07, 129.99, 128.69, 73.17, 20.46, 17.72.

IR (neat): 1737, 1518, 1380, 1227, 1201 cm^{-1}

MS (ESI): m/z = 328.0799 (MNa$^+$), calculated for C$_{15}$H$_{15}$O$_6$NNa = 328.0792

H3: (3*R*,4*R*)-4-(acetyloxy)-1-(2,6-dimethylphenyl)-2,5-dioxopyrrolidin-3-yl acetate

^{1}H NMR (400 MHz, CDCl$_3$) δ 7.32 – 7.26 (m, 2H), 7.18 (d, *J* = 7.6 Hz, 2H), 5.69 (s, 2H), 2.27 (s, 6H), 2.22 (s, 6H).

^{13}C NMR (101 MHz, CDCl$_3$) δ 170.12, 168.20, 136.07, 129.99, 128.69, 73.17, 20.46, 17.72.

IR (neat): 1734, 1380, 1223, 1201, 1100 cm^{-1}

MS (ESI): m/z = 342.0954 (MNa$^+$), calculated for C$_{16}$H$_{17}$O$_6$NNa = 342.0948.

H4: (3*R*,4*R*)-1-(4-methoxyphenyl)-2,5-dioxopyrrolidine-3,4-diyl diacetate

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 7.28 (d, *J* = 8.9 Hz, 2H), 7.05 (d, *J* = 8.9 Hz, 2H), 5.88 (s, 2H), 3.86 (s, 3H), 2.21 (s, 6H).

^{13}C NMR (101 MHz, Methanol-*d*$_4$) δ 170.26, 169.55, 160.03, 127.54, 123.83, 113.92, 72.78, 54.59, 18.85.

IR (neat): 1734, 1518, 1257, 1231 cm^{-1}

MS (ESI): m/z = 344. 0745 (MNa$^+$), calculated for C$_{15}$H$_{15}$O$_7$NNa = 344.0741

15.10.NMR spectra

A1

Proton_cd3od_T25_001
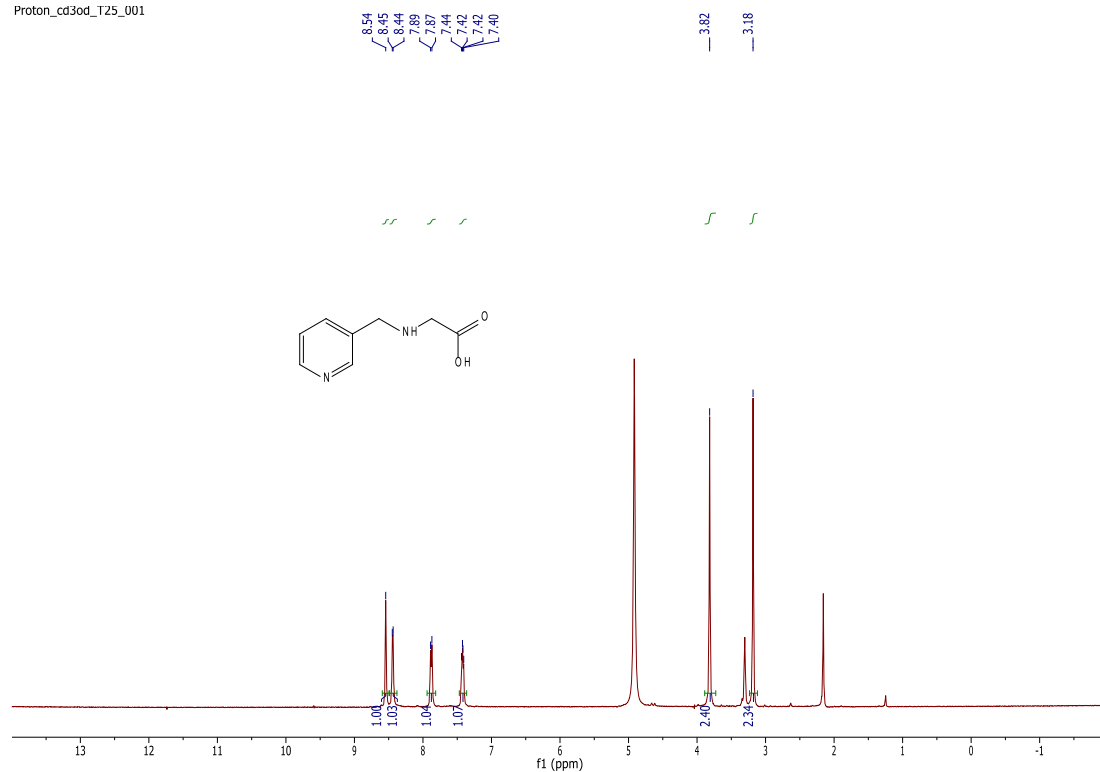
Carbon_dmso_T25_001
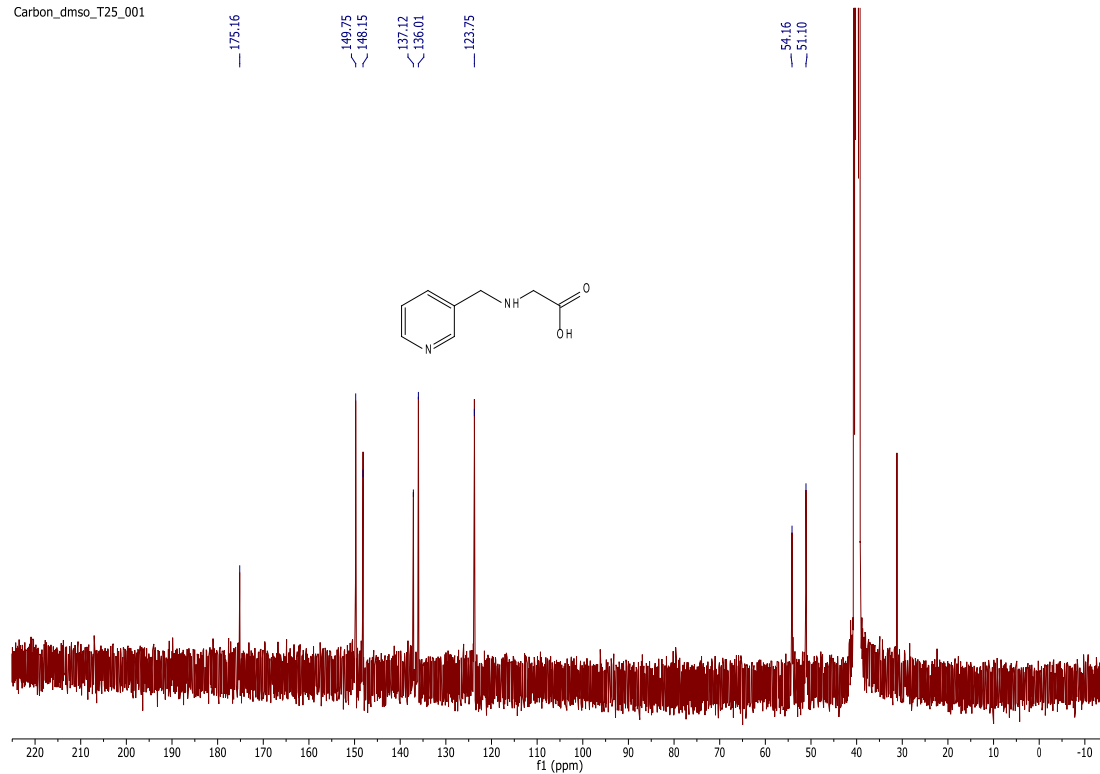
A2

Proton_cd3od_T25_001
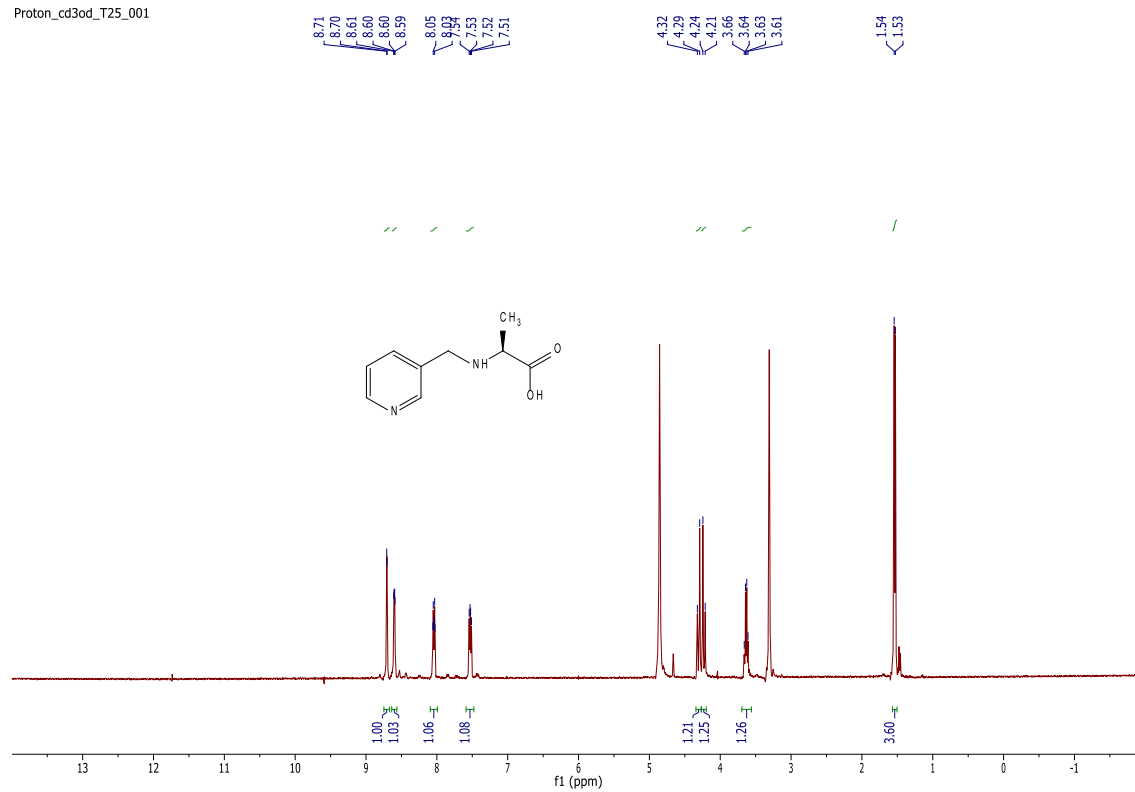
Carbon_cd3od_T25_001

A3 – HCl salt

pad=10 run with findz0 before acquisition

8.38 8.38 8.36 8.33 8.32
7.68 7.24 7.24 7.23 7.22 7.21 7.20 7.19 7.14 7.13 7.12
3.91 3.88 3.77 3.74
3.46 3.44 3.43
2.92 2.91

1.00 1.17 1.12 1.20 3.82 2.58
1.38 1.34 1.32
2.76

f1 (ppm)

Carbon_dmso_T25_001

177.17
149.67 148.12 140.97 136.73 135.87 129.80 128.10 125.88 123.61
64.47
49.46

f1 (ppm)

155

A5 – HCl salt

8.77, 8.76, 8.76, 8.75, 8.75, 8.74, 8.74
8.02, 8.02, 8.01
7.30, 7.29, 7.29, 7.28, 7.28, 7.25, 7.24
4.55, 4.53, 4.52, 4.34, 4.34, 4.32, 4.31, 4.31
3.35, 3.34, 3.31, 3.30, 3.29, 3.28, 3.26, 3.24

1.89, 2.00, 3.19, 2.24, 2.38, 1.18, 2.37

f1 (ppm)

170.43
150.92
141.86, 133.64, 129.38, 129.20, 128.09, 127.82
62.16
48.46
35.12

f1 (ppm)

156

A6

Carbon_dmso_T80_001

157

A7

Proton_dmso_T25_001

7.36 7.35 7.35 7.34 7.26 7.26 7.25 7.24 7.23 7.20 7.19 7.18 7.17 7.16 7.16 7.15 6.92 6.91 6.90 6.88 6.87 3.97 3.94 3.80 3.77 3.40 3.38 3.37 2.92 2.91 2.89 2.88 2.86 2.84 2.83 2.81

1.00 2.23 3.28 1.10 1.16   1.19 1.24 1.22   2.45

13   12   11   10   9   8   7   6   5   4   3   2   1   0   -1
f1 (ppm)

Carbon_dmso_T25_001

174.83   143.76 138.47 129.72 128.46 127.01 126.66 125.53 125.41   61.64   46.06 39.75

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

A8

7.53 7.52 7.25 7.24 7.22 7.20 7.20 7.19 7.18 7.17 7.15 7.15 7.14 6.35 6.34 6.33 6.16 6.16 3.77 3.73 3.65 3.61 3.39 3.38 3.36 2.93 2.90 2.92 2.88 2.86 2.84 2.83 2.81

1.00 5.32 1.08 1.09 1.19 1.24 1.19 2.42

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

174.70 153.96 142.26 138.66 129.60 128.40 126.53 110.64 107.27 61.95 44.26

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

159

A10

160

A11

8.52
8.51
7.41
7.40
3.90
3.16

2.04   1.98   2.04   2.00

f1 (ppm)

171.07
150.01
146.37
123.97
50.31
50.13

f1 (ppm)

A12

8.51
8.49
7.44
7.43
4.19
3.68
3.66
3.64
3.62
1.44
1.42

1.70
1.76
1.94
0.95
3.00

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

174.50
149.36
141.32
124.64
57.94
48.28
15.19

230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

162

A13

8.51 8.50 8.50 8.49   7.45 7.45 7.44   4.20 3.69 3.67 3.65 3.64   1.45 1.43

1.83   1.88   2.08 1.05   3.00

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

174.50   149.35 141.34   124.67   57.95 48.29   15.22

230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

163

A14

PROTON_dmso_T25_001
Gradient Shimming

11.41

7.64 7.40 7.38 7.35 7.34 7.33 7.22 7.19 6.39

4.11 4.08 4.07 4.04 3.22 3.20 3.19 3.17

1.32 1.30

0.79

1.00 1.90 1.01

0.89

1.90

1.03

2.82

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

CARBON_dmso_T25_001
Gradient Shimming

172.07

136.21 127.97 126.56 123.77 123.11 122.18 111.90

101.46

56.58 50.02

16.30

230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

164

A15

11.39

7.63
7.40
7.38
7.35
7.34
7.34
7.21
7.21
7.19
7.19
6.40
6.40
6.39

4.11
4.08
4.07
4.04
3.21
3.20
3.18
3.16

1.31
1.30

0.82

1.00
1.90
0.93

0.90

2.04

1.10

2.91

13   12   11   10   9   8   7   6   5   4   3   2   1   0   -1
f1 (ppm)

171.55

136.22
127.96
126.58
123.46
123.13
122.26

111.90

101.47

56.50
49.91

16.14

230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

B1

8.54
8.54
8.54
8.48
8.46
7.70
7.69
7.68
7.68
7.67
7.66
7.66
7.26
7.25
7.24
7.23
4.04
4.03
3.80
3.72
3.34
2.89

1.00
1.11
1.29
1.13
2.12
2.13
3.06
2.13
1.31

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

171.39
170.32
149.56
148.72
136.13
134.44
123.58
52.37
51.49
50.99
40.70

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

B3

8.48
8.48
8.47
8.37
8.36
7.31
7.29
7.29
7.26
7.18
7.17
7.17
7.16
7.15
4.16
4.14
4.04
4.02
3.80
3.77
3.77
3.59
3.55
3.41
2.76
2.74
2.73
2.70

1.00
1.02
1.09
1.10
4.32
3.46
1.20
1.31
4.61
1.31
1.26
1.34
1.28

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

173.65
170.31
149.45
148.61
137.04
135.76
134.44
128.97
128.89
127.11
123.43
63.19
52.35
49.88
40.76
39.19

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

B5

8.41 8.40 8.30 8.29 8.28
7.47 7.27 7.27 7.09 7.08 7.02 7.02 6.93 6.92
4.90 4.89 4.88 4.87
3.63 3.62 3.60 3.53 3.50 3.30 3.29 3.28 3.27 3.15 3.13
1.19 1.17

1.04 1.97
1.01 1.10 1.98 3.25 1.03
1.02
4.02 1.29 2.36 1.20
3.12

f1 (ppm)
13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2

174.31 172.41
149.48 148.56
136.12 135.82 134.84
122.66 122.26 119.59 118.87 111.20 110.06
57.99 52.32 52.16 49.68
27.53
19.46

f1 (ppm)
220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10

B6

Proton_cdcl3_T25_002
pad=10 run with findz0 before acquisition

Carbon_cdcl3_T25_001

pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition

B8

8.37
8.37
8.36
8.36
7.24
7.22
7.21
7.11
7.11
7.09
7.09
6.90
6.90
6.89
6.87
4.09
4.06
4.04
3.96
3.95
3.92
3.91
3.74
3.73
3.70
3.69
3.53
3.49
3.30
3.29
3.28
3.27
3.19
3.18
3.15
3.14
2.69
2.67
2.66
2.63

2.00
0.92
2.98
1.97
1.80
1.00
1.01
3.82
1.04
0.98
1.02
0.98
1.02

f1 (ppm)

172.61
169.30
148.78
147.01
136.07
128.02
127.88
126.11
121.75
62.08
51.32
50.04
39.74
38.29

f1 (ppm)

B9

B11

Proton_cdcl3_T25_001
pad=10 run with findz0 before acquisition

7.86 7.86 7.86 7.84 7.84 7.24 7.23 7.22 7.22 7.21 7.20 7.19 7.14 7.14 7.13 7.12 7.07 7.05 6.95 6.94 6.93 6.92 6.87 6.86 6.85 6.77 6.76 6.75 6.74 6.68 6.67 4.94 4.93 4.92 3.85 3.84 3.81 3.72 3.71 3.71 3.68 3.67 3.67 3.42 3.41 3.40 3.39 3.13 3.11 3.10 3.09 3.04 3.03 3.01 3.00 3.00 3.00 2.98 2.97 2.53 2.51 2.50

0.89 3.01 1.04 1.90 1.97 0.96 1.88 0.91 0.92 1.04 3.77 1.06 1.19 2.12 1.01

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

Carbon_cdcl3_T25_001

pad=10 run with findz0 before acquisition

173.78 171.97 155.74 142.26 136.83 130.26 129.04 128.76 126.96 126.88 126.69 125.32 124.66 115.62 62.59 52.80 52.40 46.68 38.81 37.43

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

172

B12 (as a mixture with tetramethylurea)

Proton_cdcl3_T25_001
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition

BST-L1-45-FB-13C
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition

B14

BST-L1-50-FB_H
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition

Carbon_cdcl3_T25_001

pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition

B15 (as a mixture with tetramethylurea)

Proton_cdcl3_T25_002

pad=10 run with findz0 before acquisition

BST-L1-51-FB-C13

pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition

B16

176

B17

177

B18

PROTON_cd3od_T25_001
Gradient Shimming

CARBON_cd3od_T25_001
Gradient Shimming

B20
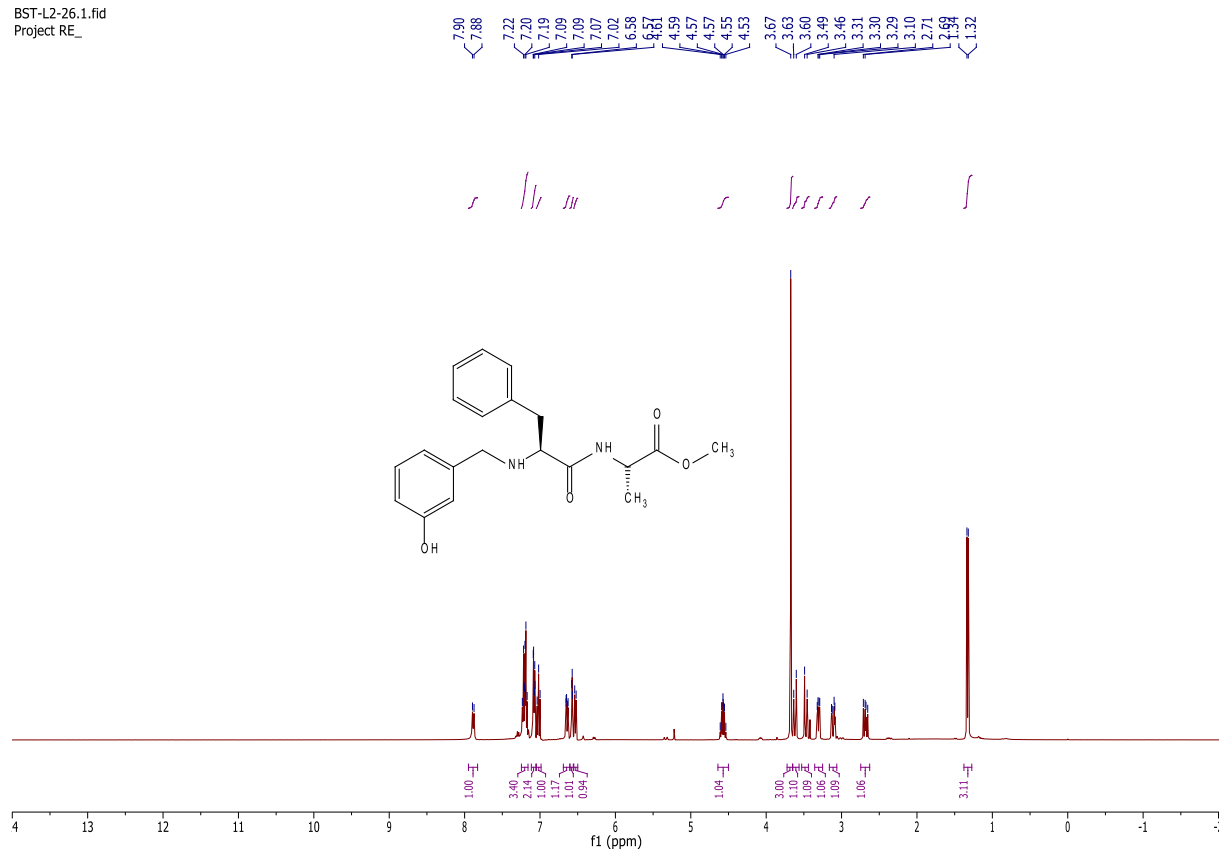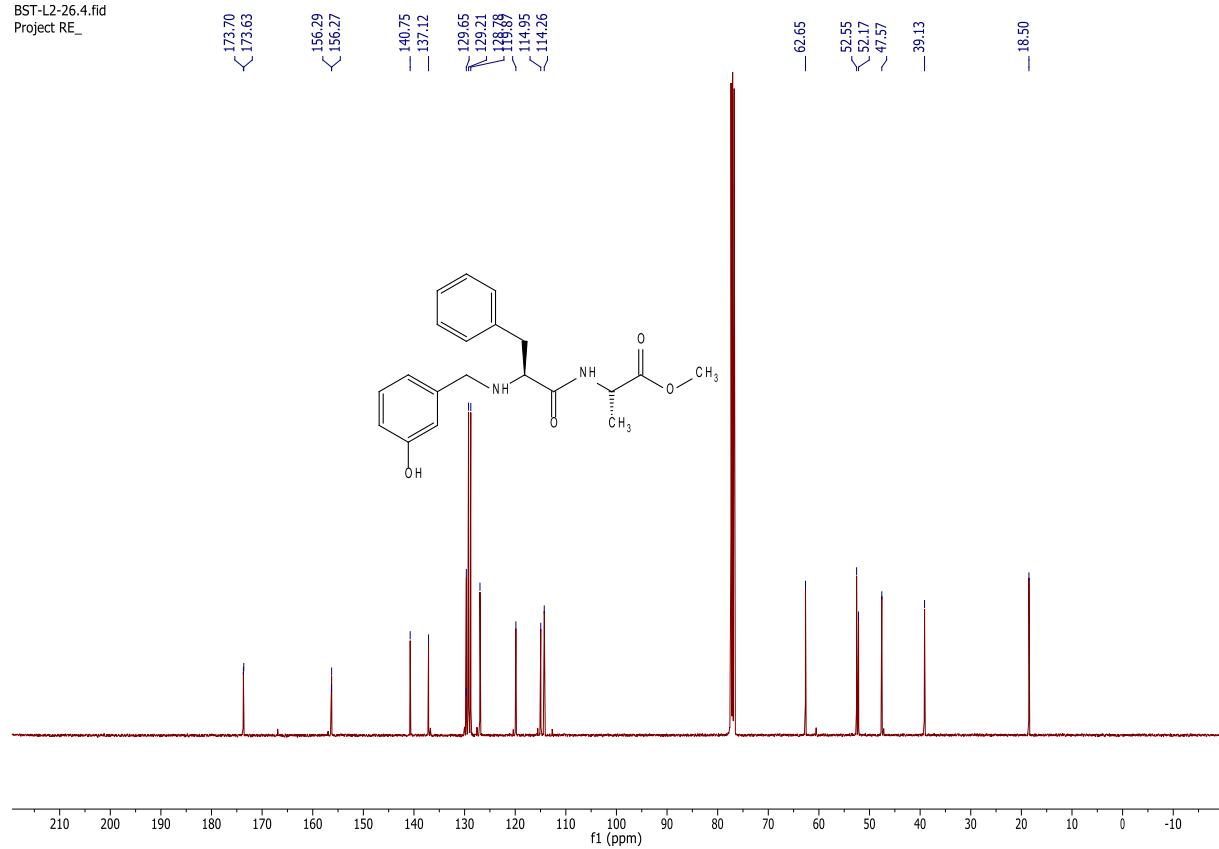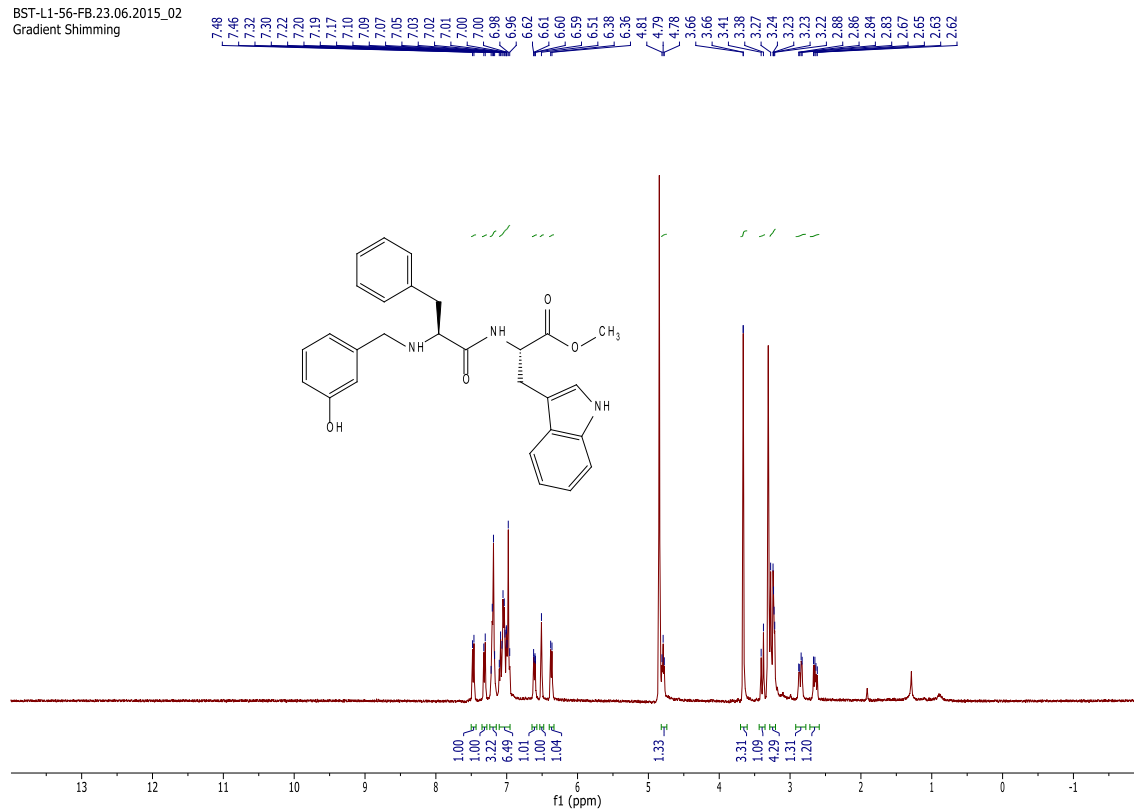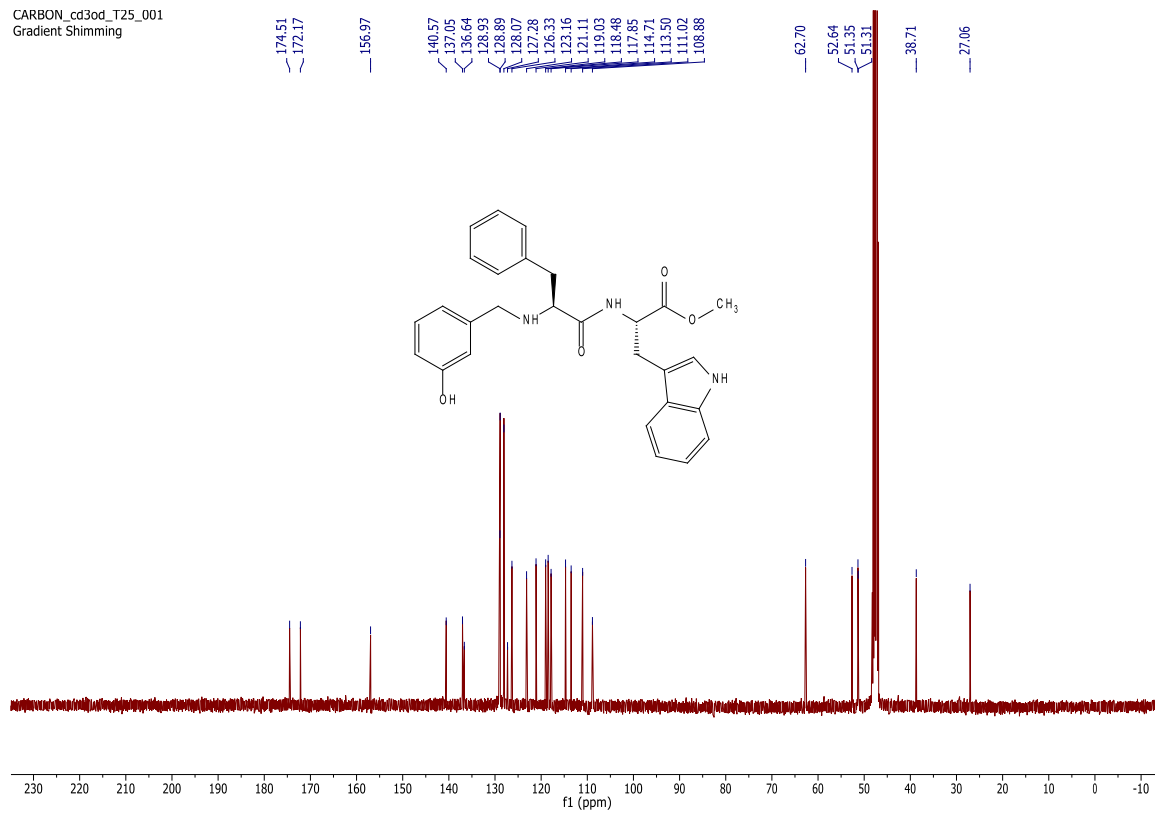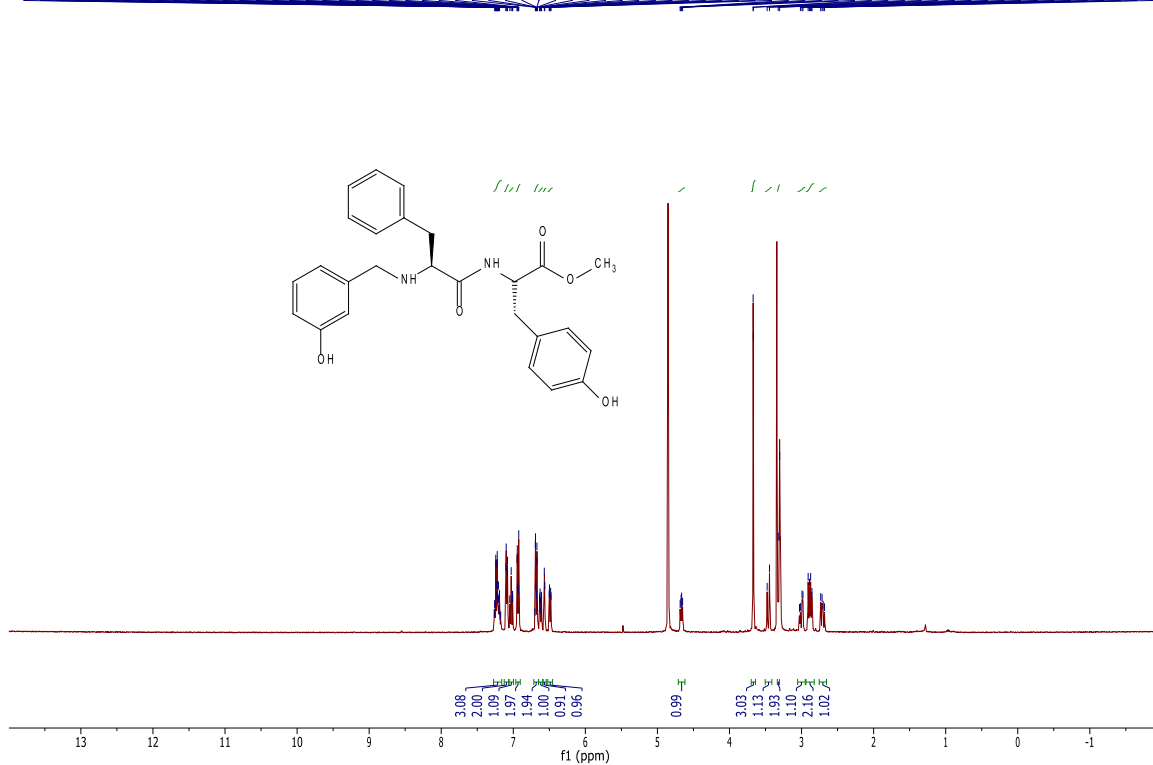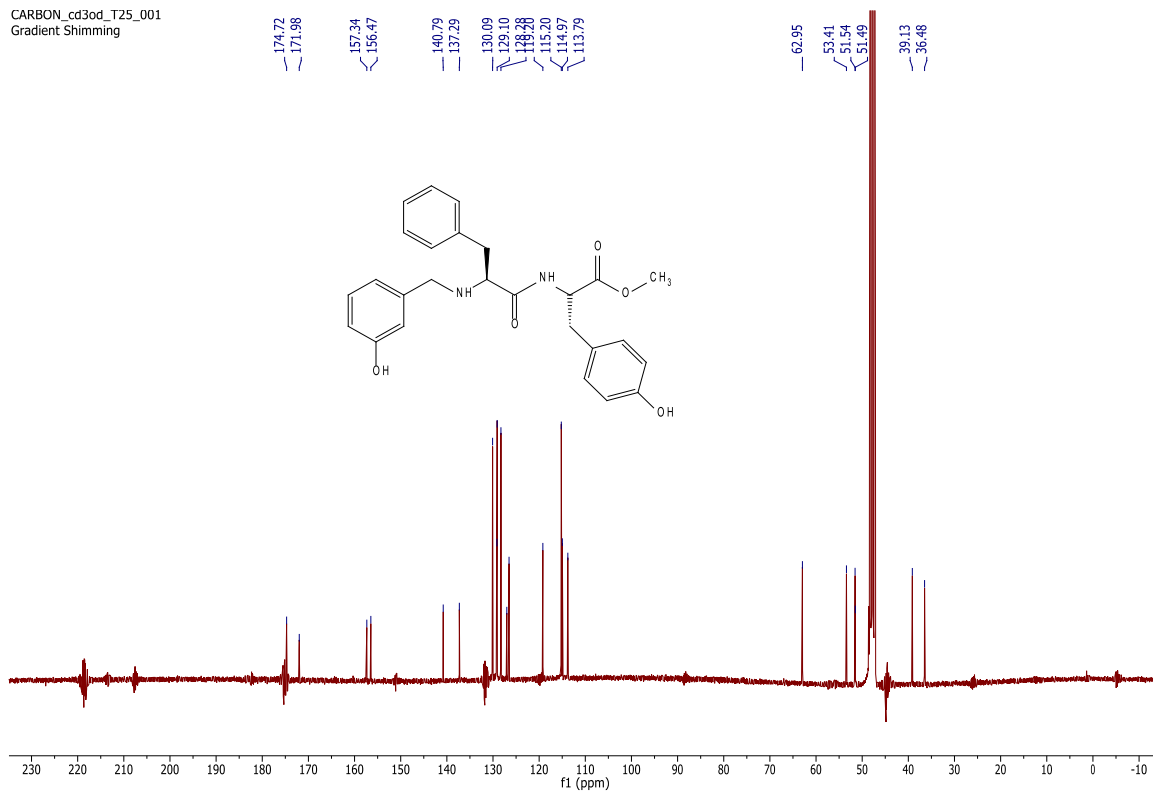
8.84
8.28 8.27 8.26
7.26 7.25 7.24 7.12 7.12 7.10 6.86 6.59 6.58
4.97 4.97 4.96 4.95 4.94 3.41 3.38 3.37 3.31 3.30 3.28 3.26 3.25 3.22 3.21 3.20 3.19 3.18 3.17 3.14 3.10 2.72 2.70 2.69 2.68 2.65
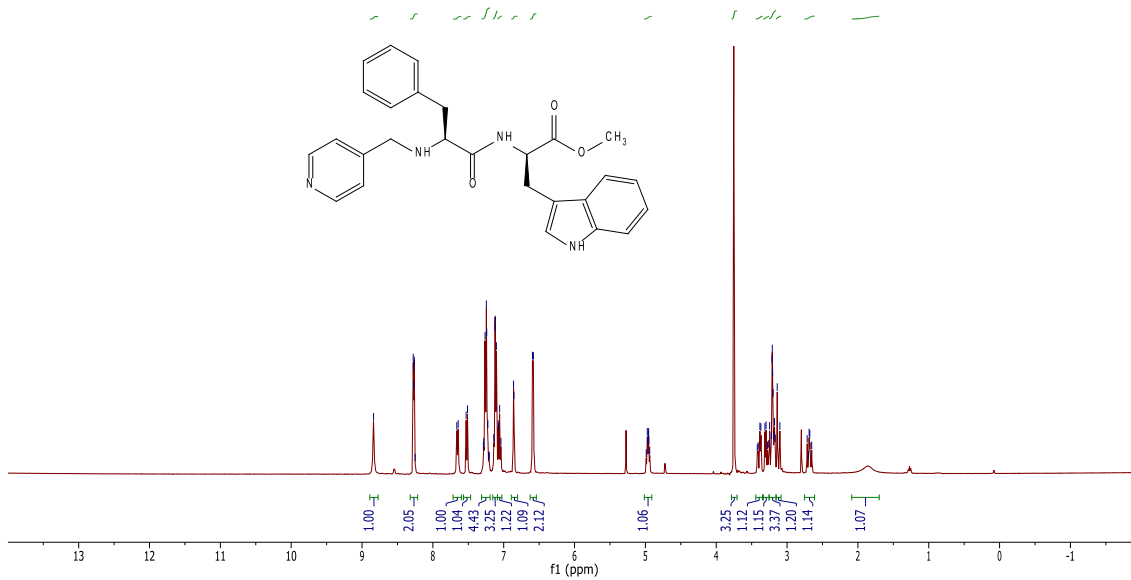
1.00 2.05 1.00 1.04 4.43 3.25 1.22 1.09 2.12 1.06 3.25 1.12 1.15 3.37 1.20 1.14 1.07

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

173.17 172.48
149.36 148.27
137.24 136.20
129.05 128.82 122.53 122.17 119.56 119.32 109.73
63.08
52.44 52.42 50.48
39.03
27.29

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

B21

Proton_cdcl3_T25_001

9.04
8.41 8.41 8.39
7.51 7.28 7.28 7.14 7.05 6.94 6.93 6.91
4.97 4.96 4.95 4.94
3.70 3.58 3.57 3.55 3.54 3.53 3.52 3.50 3.49 3.46 3.45 3.39 3.38 3.35 3.34 3.31 3.29 3.22 3.21

0.93 | 1.93 | 0.98 1.02 1.04 1.04 1.05 1.03 1.94 | 0.99 | 3.00 2.30 2.26 2.11

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

Carbon_cdcl3_T25_001

172.39 171.03
149.62 148.38
136.24
127.69 122.75 122.13 119.49 118.49 118.40 109.61
52.48 52.32 52.16 51.79
27.38

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

B22

8.84 8.42 8.42 8.41 7.51 7.29 7.29 7.06 6.94 6.93 6.92 6.91 6.99 4.98 4.97 4.96 4.95 4.94 3.71 3.59 3.56 3.56 3.55 3.54 3.52 3.51 3.50 3.48 3.46 3.40 3.38 3.36 3.35 3.33 3.31 3.30 3.23

0.91 1.92 0.93 1.03 1.04 1.02 1.03 0.99 1.94 0.97 3.00 1.07 1.35 2.23 2.11

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

172.37 170.97 149.67 148.33 136.20 127.70 122.74 119.53 118.40 118.40 109.71 52.43 52.38 52.18 51.81 27.39

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

B23

8.48
8.46
7.47
7.46
4.00
3.85
3.73
3.73
3.31

2.00
1.82
1.90
1.79
2.71
2.49

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

BST-L1-81-FB-C13

173.21
170.41
150.25
148.61
123.48
51.38
50.66
40.17
39.02

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

B25

7.37 7.31 7.28 7.20 7.19 7.10 7.09 7.08 7.07 7.06 7.03 7.01 7.01 7.00 7.00 6.99 6.94 6.93 6.92 6.91 6.37 6.37 6.36 6.36

4.80 4.78 4.76

3.71 3.66 3.63 3.52 3.51 3.50 3.48 3.24 3.22 3.20 3.18 3.16

1.12 1.11

0.99 1.94 1.00 0.93 1.99 1.05 0.95 0.85

1.08

3.00 1.23 1.09 2.17

3.10

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

176.33 172.39

136.66 135.52 129.43 127.97 127.31 124.38 123.05 121.79 121.08 119.76 118.44 117.82 111.00 110.59 109.08 100.86

56.50 52.74 51.96 51.32

27.10

18.03

230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

184

B26

B27

186

B28

PROTON_cd3od_T25_001
Gradient Shimming

CARBON_cd3od_T25_001
Gradient Shimming

174.46
171.16
157.05
140.54
137.09
136.43
128.96
128.89
118.99
118.95
114.76
113.56
62.70
53.37
53.08
51.31
37.47
37.14
13.04

B29

PROTON_cd3od_T25_001
Gradient Shimming

7.21 7.21 7.20 7.20 7.19 7.19 7.18 7.18 7.18 7.17 7.16 7.11 7.10 7.09 7.08 7.08 7.07 7.07 7.07 7.05 7.03 7.01 6.65 6.65 6.64 6.63 6.63 6.62 6.57 6.56 6.53 6.52 6.52 6.51 6.50 4.71 4.70 4.69 4.68 3.70 3.51 3.47 3.34 3.33 3.32 3.28 3.25 3.08 3.07 2.94 2.92 2.91 2.90 2.89 2.88 2.74 2.72

f1 (ppm)

6.20 4.06 1.01 0.89 0.90 0.92 1.04 3.00 1.02 0.87 0.92 1.12 2.29 0.77

CARBON_cd3od_T25_001
Gradient Shimming

174.74 171.83 157.11 140.57 137.14 136.44 128.86 128.78 128.13 128.93 114.72 113.61 62.42 53.27 51.38 51.27 37.48 36.86

f1 (ppm)

B30

7.55 7.53 7.38 7.36 7.30 7.29 7.28 7.27 7.26 7.26 7.24 7.24 7.22 7.20 7.20 7.15 7.13 7.12 7.11 7.11 7.10 6.92 6.92 6.87 6.86 6.85 6.85 6.63 6.62 5.04 5.02 5.02 5.01 5.00 4.99 3.85 3.81 3.74 3.70 3.69 3.45 3.44 3.42 3.41 3.40 3.39 3.37 3.35 3.33 3.32 3.29 3.28 3.12 3.11 3.09 3.08 2.62 2.59 2.58 2.56

0.87 0.90 1.00 1.05 4.35 3.84 0.98 0.95 0.92 0.94 1.00 3.78 2.92 0.94 0.92

14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2
f1 (ppm)

173.09 172.30 142.62 137.15 136.14 129.15 128.75 126.62 122.24 119.50 111.50 110.21 62.57 52.38 52.32 46.70 38.74 27.80

210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

189

B31

8.34 8.33 8.33 8.15
7.31 7.25 7.24 7.23 7.23 7.20 7.12 7.11 7.10 6.98 6.98 3.39 3.37 3.35 3.29 3.28 3.27 3.26 3.24 2.95 2.93 2.91 2.90 2.76 2.74 2.72 2.71

1.00 1.01   1.06 1.08 5.53 3.40 2.23   3.04 1.32 1.14 2.28 1.05 1.05

14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2
f1 (ppm)

174.35 172.20   148.37 147.08 137.16 136.74 136.62 135.69 129.04 128.10 127.32 126.41 123.63 123.29 121.20 118.60 117.95 111.08 108.88   63.15   52.49 51.43 48.37   38.73   26.91

210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

190

B32

C1

8.54 8.54 8.50 8.49 8.48 8.48 7.83 7.81 7.45 7.44 7.43 7.42

4.66 4.07

1.00 0.94 1.00 1.04

2.34 2.34

f1 (ppm)

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1

164.70

148.68 148.21

136.86 132.23

124.04

49.12

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10

f1 (ppm)

C2

8.52
8.51
8.50
8.48
8.47
8.46
7.80
7.79
7.78
7.78
7.77
7.44
7.43
7.42
7.41
4.93
4.89
4.49
4.45
4.15
4.13
4.11
4.09
1.56
1.54

1.00
0.91
0.94
1.05
0.85
0.84
0.81
2.63

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

167.82
148.35
148.09
136.47
133.19
124.07
56.47
45.25
18.43

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

193

C3

8.45
8.44
8.43
8.26
7.60
7.58
7.39
7.37
7.35
7.33
7.30
7.28
7.26
7.09
7.07
5.07
5.04
4.25
4.24
4.22
4.04
4.00
2.88
2.85
2.84
2.40
2.38
2.36
2.35

0.92
0.87
0.94
4.24
2.00
1.16
1.06
1.09
1.17
1.09

13   12   11   10   9   8   7   6   5   4   3   2   1   0   -1
f1 (ppm)

166.37
148.56
148.15
136.61
136.59
132.51
129.42
128.67
127.14
123.99
61.56
45.39
38.28

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

194

C4

8.47
8.46
8.45
7.35
7.34
7.33
7.15
7.15
7.14
7.13
7.11
7.09
7.09
5.09
5.08
5.05
4.30
4.28
4.28
4.04
4.02
3.99
2.91
2.88
2.42
2.40
2.38
2.37

3.83
3.80
2.04
3.68
3.82
1.96
1.94
1.98
1.99
1.95

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

166.41
149.02
146.33
136.51
129.43
128.69
127.16
122.89
61.88
47.02
38.35

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

195

C5

7.37 7.35 7.35 7.34 7.33 7.30 7.29 7.29 7.28 7.27 7.26 7.22 7.21 7.21 7.10 7.08 6.89 6.88 6.87 6.70 6.69 5.33 5.29 4.25 4.24 4.23 4.22 3.79 3.75 2.91 2.90 2.41 2.40 2.39 2.38 2.36

4.21 1.98 2.00 4.25 2.00 1.94 2.09 2.07 2.08 2.09 2.15

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)
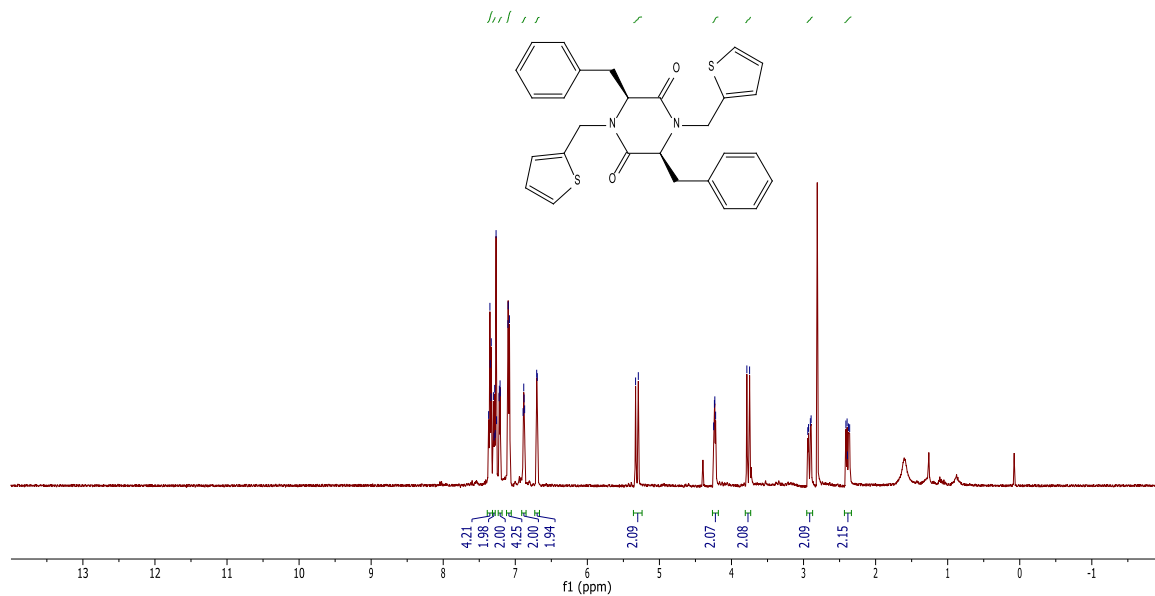
BST-L1-40-C1-FB

165.49 137.64 136.80 129.68 128.94 127.60 127.41 126.60 126.32 60.45 42.49 38.93

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

196

C8

8.52
8.51
8.50
7.39
7.38
7.37
4.68
4.11
4.11

2.00
1.94
2.17
1.92

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2
f1 (ppm)

164.83
149.09
146.13
123.00
49.47

230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

197

D1

Proton_cd3od_T25_001

8.55 8.50 8.49 7.84 7.82 7.46 7.45 7.44 7.43   4.67   4.03 3.97

1.00 1.05   1.08   1.22   2.27   2.31 2.19

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

Carbon_cd3od_T25_001

166.55 165.41   148.59 148.15   136.81 132.38   124.06   48.86 44.22

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

198

D2

8.54 8.53 8.53 8.48 8.48 8.47 8.47
7.83 7.82 7.82 7.81 7.80 7.80 7.44 7.43 7.42 7.41
5.01 4.97 4.41 4.37 4.20 4.15 3.94 3.92 3.89
1.47 1.46

1.00 0.93
0.98
1.02
1.05
1.10 1.10 2.25
3.31

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

169.97 165.79
148.38 148.06
136.50 133.26
124.01
56.48
47.30 43.84
16.32

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

199

D3

f1 (ppm)

1.00 0.99 1.01 1.04 3.09 2.10 1.08 2.21 1.12 2.31 1.16 1.12

Carbon_cd3od_T25_001

168.43 166.04 148.64 148.19 136.71 134.71 132.87 129.86 128.39 127.46 124.07 61.18 44.74 43.22 36.27

f1 (ppm)

200

D4

8.58
8.57
8.57
8.49
8.48
8.48
7.32
7.31
7.30
7.30
7.28
7.27
7.15
7.13
7.13
5.35
5.31
4.36
4.32
4.27
4.26
4.26
4.25
3.88
3.87
3.85
3.83
3.41
3.40
3.37
3.36
3.28
3.27
3.24
3.23
0.43
0.42

1.00
0.98
0.98
1.04
3.20
2.07
1.04
1.06
1.16
1.00
1.05
1.22
2.97

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

168.11
166.39
148.76
148.14
136.80
134.96
132.77
130.16
128.46
127.27
124.05
60.67
50.64
44.63
35.96
19.09

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

201

D5

8.39
7.58 7.56 7.51 7.49 7.34 7.32 7.26 7.09 7.07
4.73 4.69 4.42 4.41 4.40 4.17 4.17 4.14
3.70 3.68 3.66 3.64 3.30 3.20 3.18 3.17 3.15 3.13
1.32 1.30 1.28
0.34 0.32

2.06
1.09 1.05 1.05 1.06 2.01 1.16
1.31 1.13 1.12
1.21
1.46
1.49
3.00

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

168.62 167.00
148.19 147.70 136.43 136.39 133.03 127.66 124.47 123.84 121.23 118.82 118.56 110.94 107.98
56.38 55.60
45.15
29.68
16.86

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

D6

Proton_cd3od_T25_001

8.31
8.30
8.29
8.23
8.23
7.35
7.33
7.32
7.30
7.15
7.15
7.13
6.95
6.94
5.93
5.91
5.24
4.08
4.04
3.82
3.81
3.80
3.27
3.27
3.26
3.25
3.23
3.22
3.21
3.11
3.10
3.07
3.06
2.94
2.93
2.92
2.89
2.88

1.00
1.07
1.05
4.46
2.34
1.31
3.30
1.04
1.12
1.19
1.14
1.93
1.26
2.35

13　12　11　10　9　8　7　6　5　4　3　2　1　0　-1
f1 (ppm)

Carbon_cd3od_T25_001

167.65
167.45
148.25
147.71
135.84
134.45
129.97
128.28
127.42
124.14
121.03
118.77
118.57
107.78
60.16
53.98
43.61
36.01
27.70

220　210　200　190　180　170　160　150　140　130　120　110　100　90　80　70　60　50　40　30　20　10　0　-10
f1 (ppm)

203

D7

8.51
8.51
8.46
8.45
7.75
7.74
7.73
7.42
7.41
6.99
6.97
6.72
6.71
6.70
5.48
4.79
4.75
4.40
4.36
4.35
3.79
3.74
3.73
3.23
3.21
3.20
2.92
2.91
2.88
2.87
0.65
0.63

1.04
0.94
1.01
1.01
2.09
2.00
1.19
2.12
1.07
1.13
1.09
2.93

f1 (ppm)

168.68
166.43
156.91
148.55
147.87
136.72
133.19
131.29
125.60
123.82
115.13
56.50
55.89
45.47
38.67
17.41

f1 (ppm)

D8

8.51
7.35
7.34
7.34
7.33
7.33
7.32
7.32
7.31
7.20
7.18
5.30
5.10
5.26
4.22
4.19
4.18
4.17
3.51
3.47
3.23
2.54
2.51
2.50

2.21
5.06
2.28
0.94
2.08
1.14
1.24
1.11
1.00

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

Carbon_cd3od_T50_002

168.43
166.17
149.14
146.54
134.77
129.82
128.40
127.43
122.85
61.65
43.33
36.39

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

205

D9

7.43, 7.41, 7.34, 7.32, 7.30, 7.29, 7.28, 7.26, 7.26, 7.25, 7.25, 7.24, 7.24, 7.23, 7.22, 7.09, 7.08, 7.06, 7.05, 7.04, 6.99, 6.98, 6.97, 6.95, 6.95, 6.92, 6.86, 6.85, 6.84, 6.84, 6.69, 6.69, 5.33, 5.29, 4.37, 4.33, 4.01, 4.00, 3.99, 3.28, 3.26, 3.24, 3.23, 3.12, 3.10, 3.10, 3.09, 3.08, 3.07, 3.06, 3.05, 3.04, 3.02, 3.01, 2.89, 2.88, 2.87

1.00, 2.03, 3.16, 3.22, 1.13, 0.98, 1.00, 1.07, 1.11, 1.10, 1.26, 3.45, 1.11

f1 (ppm)

Carbon_cd3od_T25_001
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition

168.04, 167.03, 137.65, 136.54, 134.58, 129.89, 128.24, 127.36, 127.33, 126.35, 125.72, 123.85, 121.02, 118.44, 118.40, 110.74, 107.80, 104.99, 60.04, 53.51, 41.52, 36.03, 27.60

f1 (ppm)

206

Proton_cdcl3_T25_001
pad=10 run with findz0 before acquisition

BST-L1-42-FB

pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition

D11

208
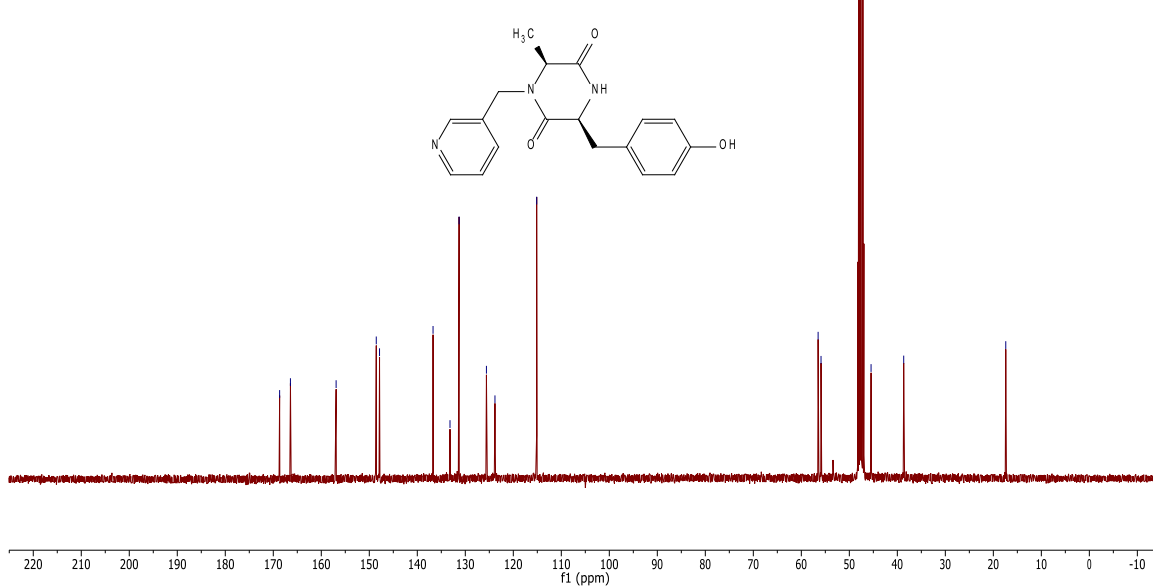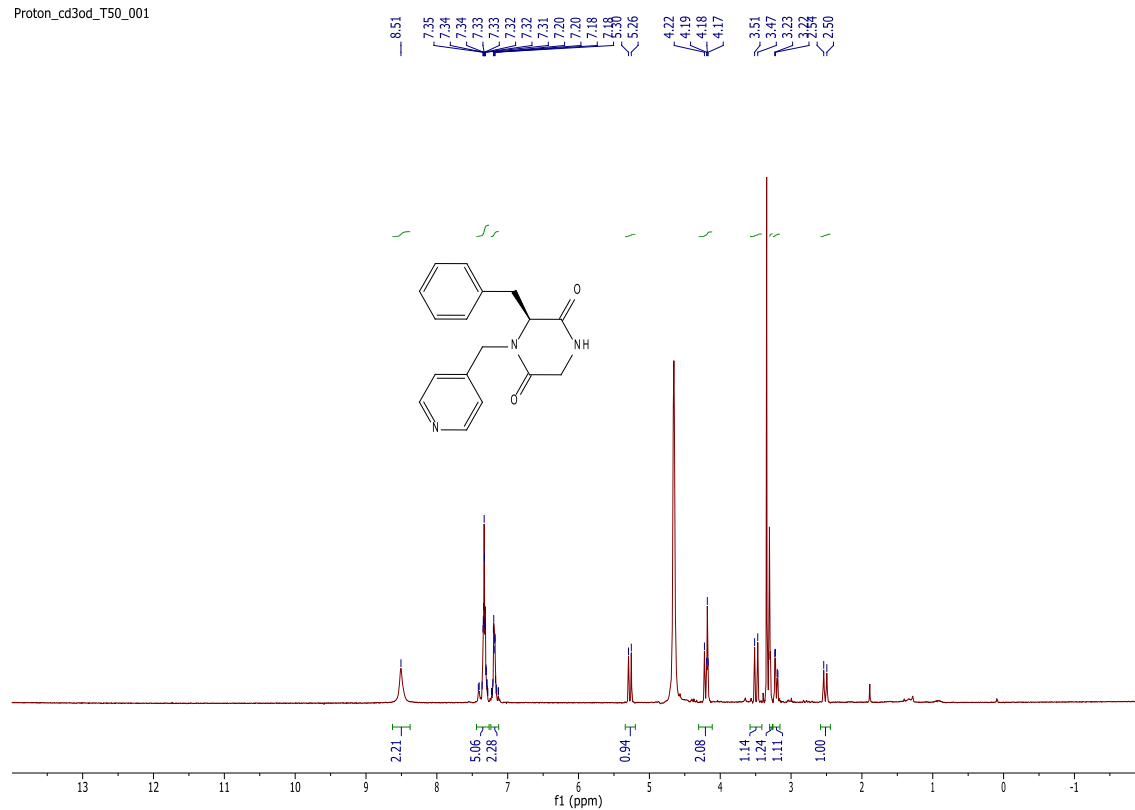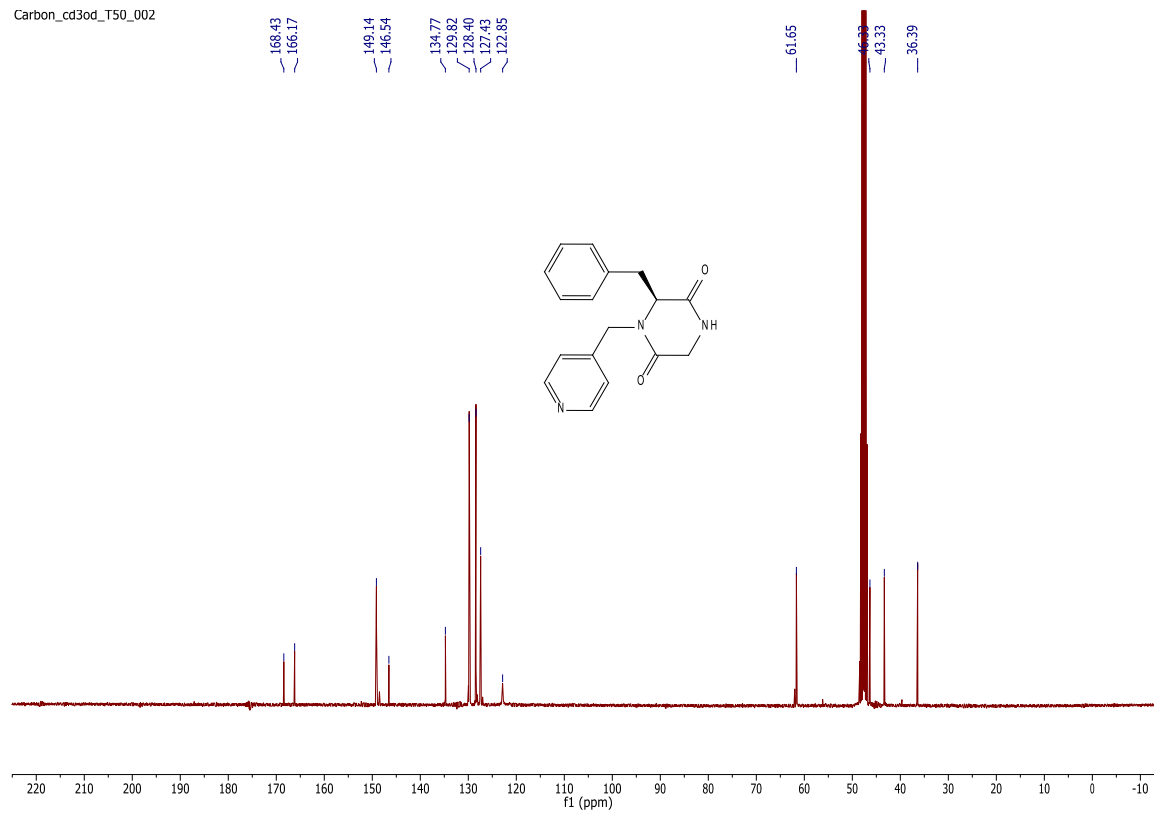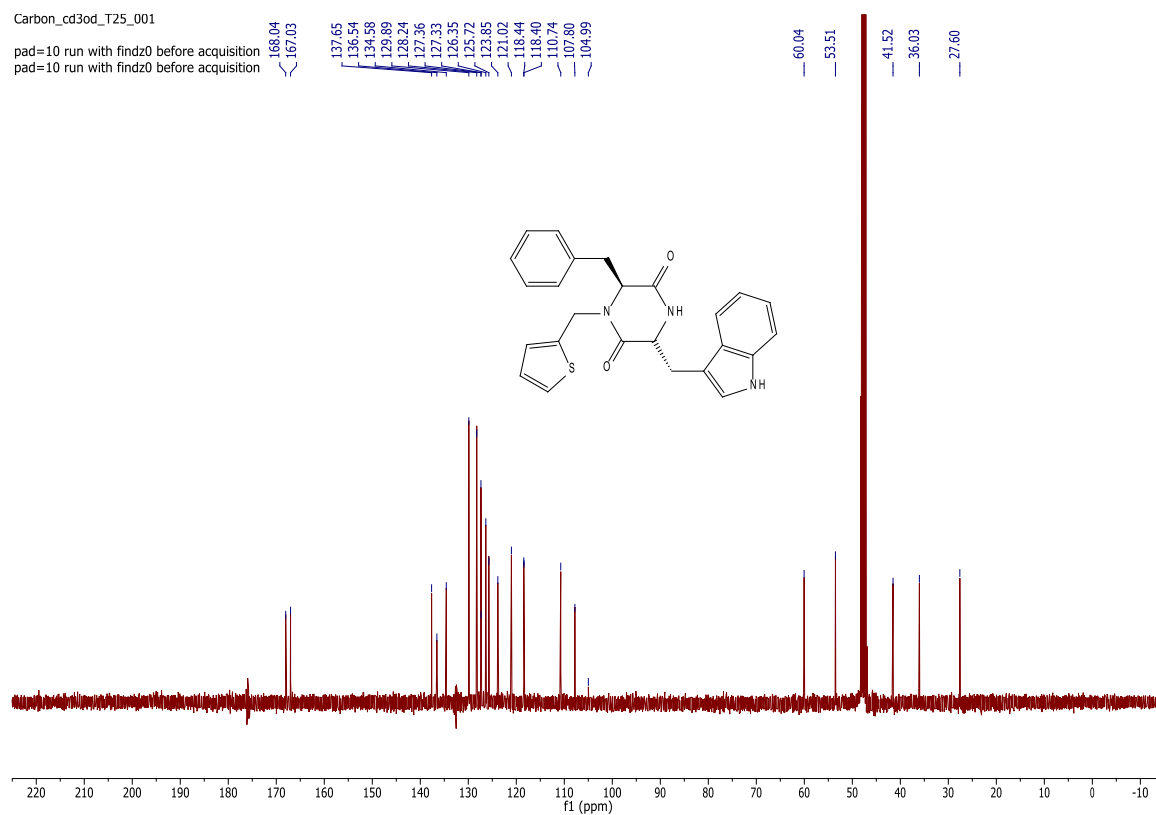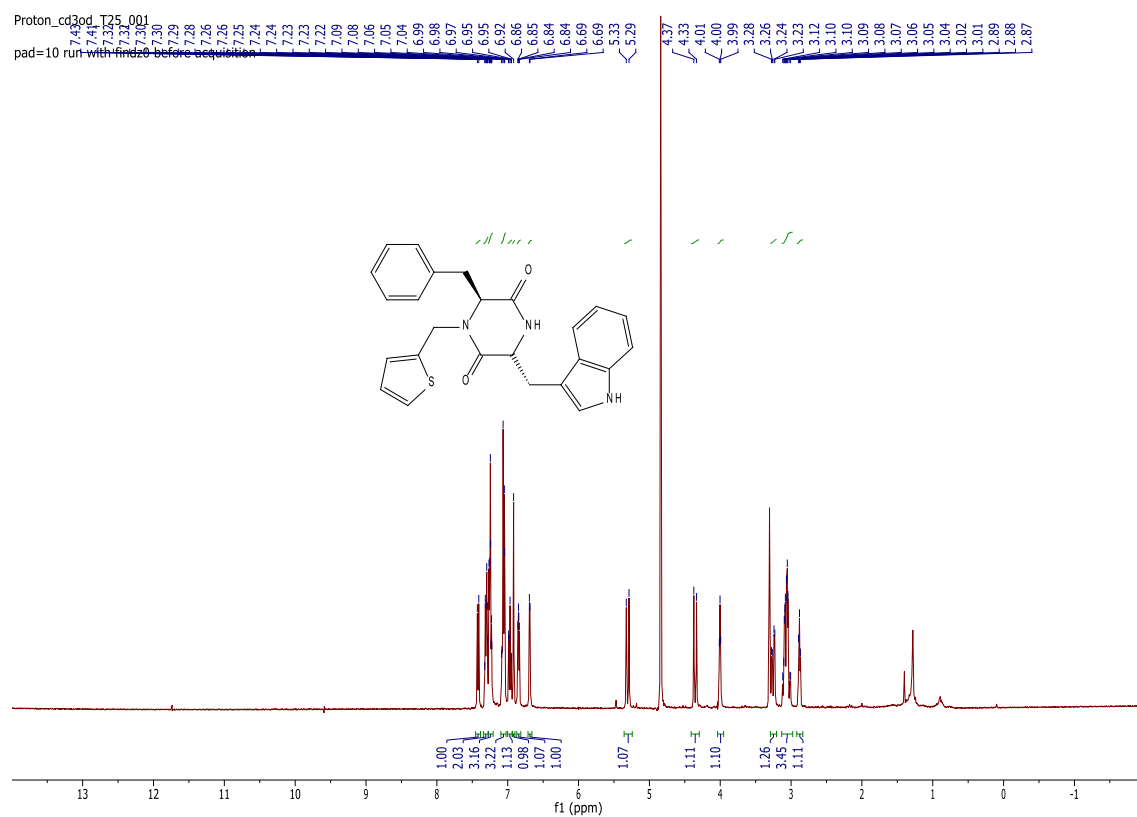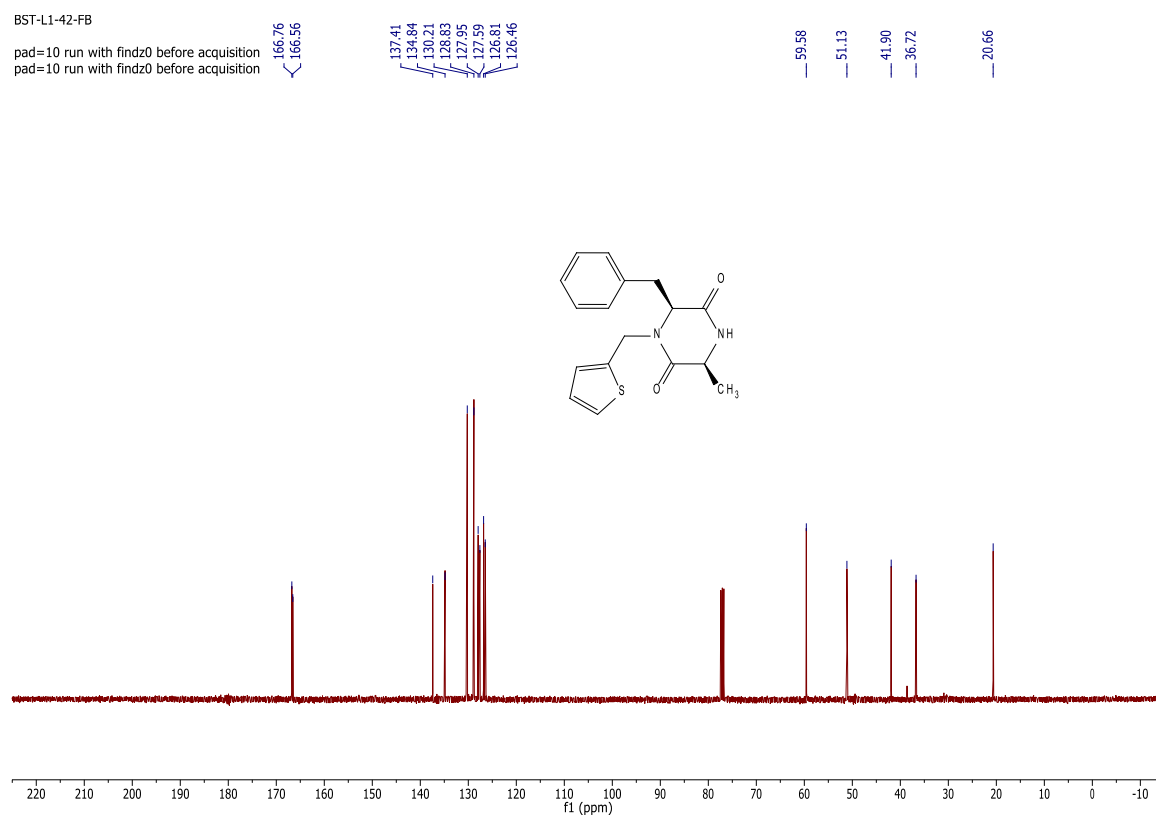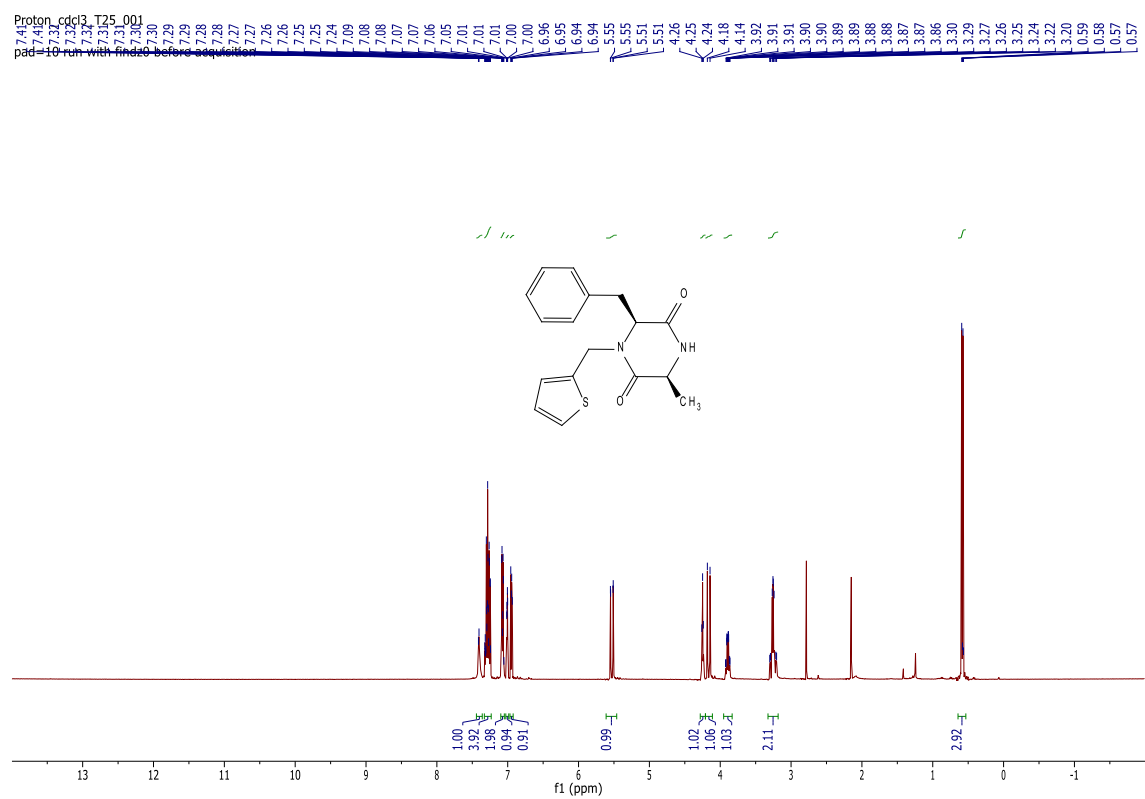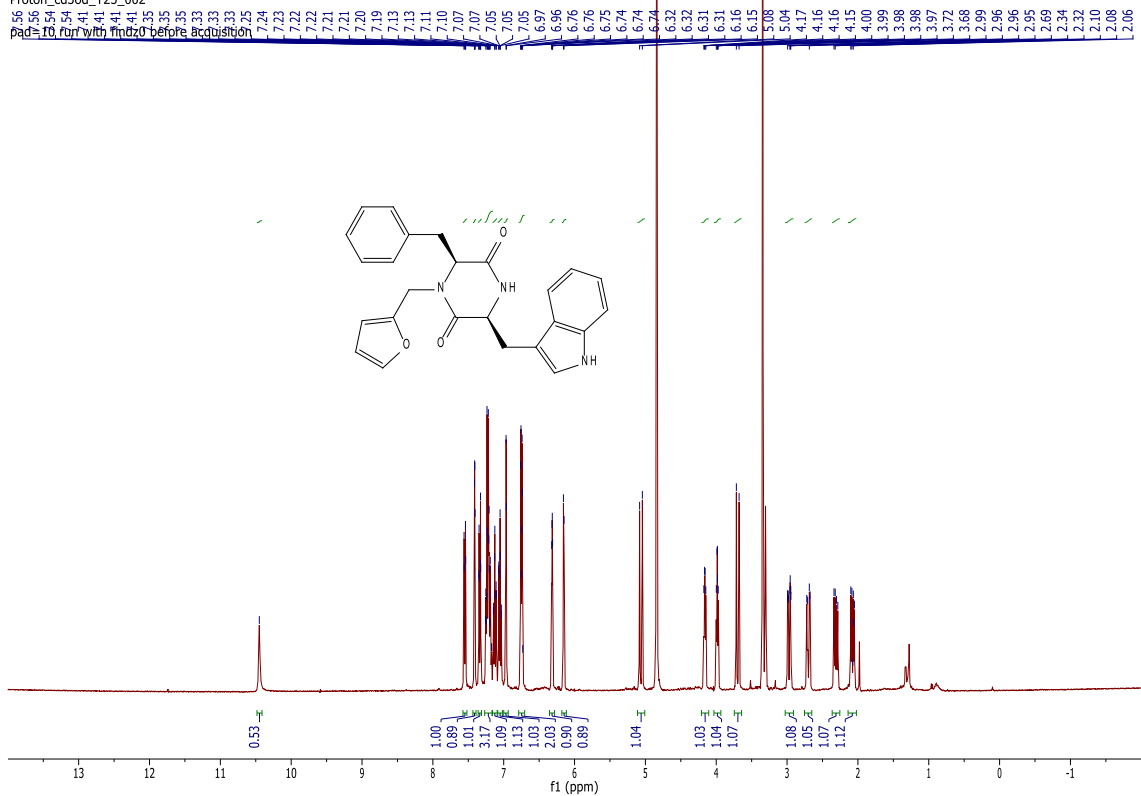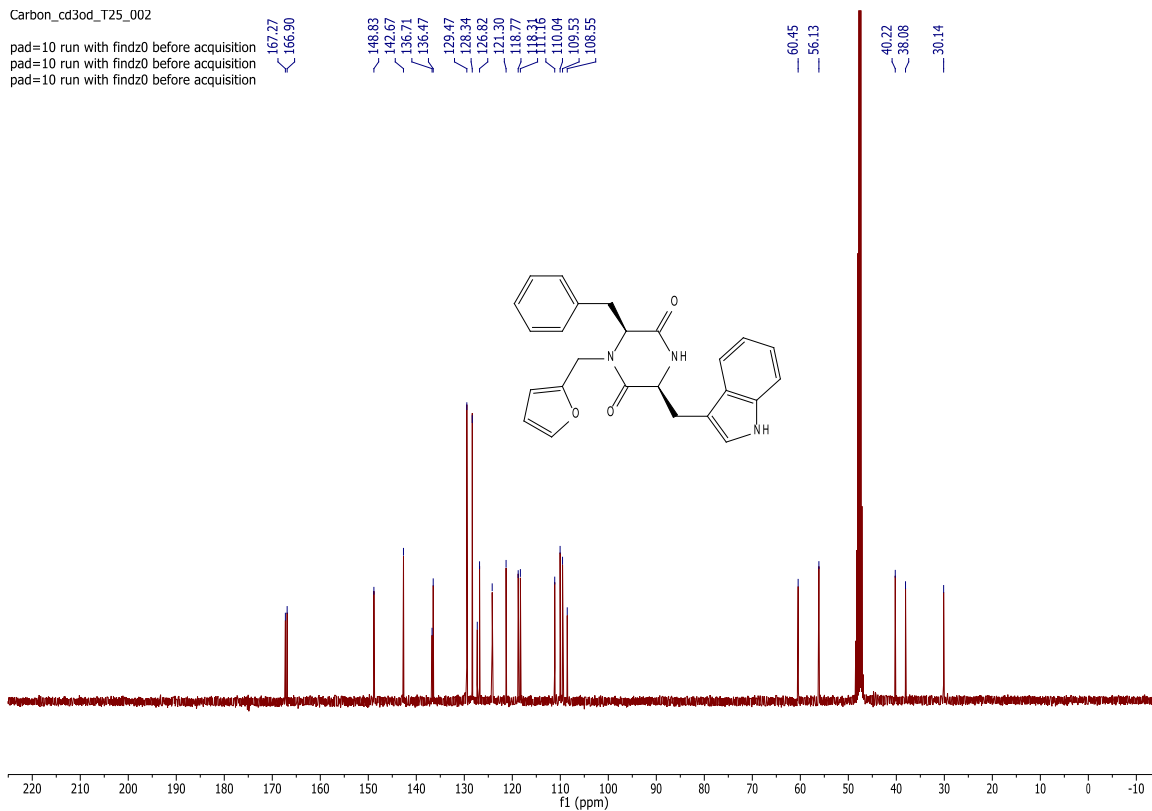
D12

Proton_cd3od_T25_003

pad=10 run with findz0 before acquisition

7.47 7.46 7.45 7.37 7.36 7.35 7.34 7.33 7.29 7.28 7.27 7.26 7.25 7.11 7.09 6.87 6.86 6.86 6.85 6.84 6.83 6.74 6.73 6.72 6.71 6.38 6.38 6.38 6.37 6.33 6.32 5.16 5.12 4.14 4.13 4.12 4.10 3.98 3.97 3.00 2.99 2.73 2.72 2.63 2.58 1.56 1.55 1.52

0.65 1.74 1.27 2.01 2.00 1.74 0.74 0.76 0.88 1.81 0.90 1.08 1.17 1.00 0.91

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

Carbon_cd3od_T25_002

pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition

167.06 166.48 156.32 148.94 142.72 136.06 130.43 129.82 128.55 127.06 126.67 115.08 110.16 109.73 60.42 56.96 40.22 39.41 37.07

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

D13

7.53 7.51 7.51 7.51 7.33 7.32 7.31 7.30 7.30 7.29 7.28 7.27 7.27 7.26 7.13 7.11 7.11 7.10 7.09 6.46 6.46 6.43 6.42 6.42 5.04 4.40 4.36 4.22 4.21 3.39 3.35 3.22 3.17 2.35 2.32

0.82 3.10 1.98 1.71 0.95 1.01 0.99 1.00 2.19 1.08

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

168.61 165.50 149.10 142.84 134.77 129.86 128.31 127.36 110.31 109.78 60.86 43.27 40.12 36.21

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

210

D14

7.51
7.51
7.51
7.50
7.32
7.31
7.29
7.29
7.28
7.27
7.26
7.26
7.25
7.23
7.09
7.08
7.07
7.06
7.06

6.47
6.46
6.43
6.42
6.42
5.23
5.19
4.41
4.37
4.22
4.21
4.20
3.82
3.81
3.79
3.79
3.77
3.37
3.36
3.34
3.34
3.33
3.24
3.23
3.20
3.20

| 0.87 | 3.12 | 1.98 | 0.88 | 0.87 | | 0.99 | | 1.06 | 1.00 | 0.93 | 1.12 | 1.25 | | | 2.92 |

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

167.62
166.55

149.06
142.76

134.99
130.13
128.36
127.16

110.24
109.82

60.31

50.67

39.93
35.73

19.12

230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

D15

7.32, 7.32, 7.31, 7.31, 7.30, 7.29, 7.29, 7.28, 7.28, 7.27, 7.18, 7.17, 7.17, 7.16, 7.15, 7.15, 7.15, 7.14, 7.14, 7.13, 7.13, 6.75, 6.75, 6.74, 6.74, 6.73, 6.73, 6.72, 6.72, 5.25, 5.25, 5.21, 5.21, 4.09, 4.01, 4.00, 3.97, 3.47, 3.43, 3.43, 2.55, 2.55, 2.51

2.93, 3.00, 2.85, 1.01, 1.02, 1.08, 1.07, 1.08, 1.06, 1.05

f1 (ppm)

BST-L1-58-FB-C13

168.78, 165.79, 158.18, 137.02, 134.88, 129.85, 129.58, 128.33, 127.37, 118.67, 114.81, 114.69, 60.26, 46.61, 43.31, 35.98

f1 (ppm)

D16

7.33 7.31 7.30 7.28 7.26 7.25 7.21 7.20 7.18 7.16 7.15 7.14 7.14 7.13 6.80 6.78 6.76 6.74 6.73 5.40 5.36
4.14 4.13 4.12 4.09 3.88 3.87 3.36 3.35 3.33 3.31 3.21 3.20 3.17 3.16
0.45 0.44

3.00 3.07 2.77
0.97
2.06 0.97
0.92 1.16
2.86

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

167.91 166.73 157.75 136.95 135.10 130.14 129.62 128.42 127.20 118.98 114.64 114.55 59.69 50.75 46.37 35.64 19.22

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

213

D17

D18

Proton_cd3od_T25_002
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition

BST-L1-61-FB-C13

pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition
pad=10 run with findz0 before acquisition

D19

D20

8.25
8.25
8.24
8.24
7.57
7.55
7.38
7.13
7.03
7.02
6.74
6.74
6.73
4.40
4.39
4.34
4.30
4.24
4.20
3.57
3.56
3.53
3.41
3.36
3.16
3.15
3.12
3.11
2.72
2.72
2.68

1.81

0.92
0.88
0.94
1.81
1.83

1.00
1.05
1.09

1.03
0.90
1.07
0.95

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

CARBON_cd3od_T25_001
Gradient Shimming

167.28
166.18
148.80
145.54
136.47
127.33
124.62
122.60
121.42
118.94
118.43
111.02
107.54
56.16
48.46
77.90
29.80

230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

217

D21

8.51
8.51
8.50
7.36
7.35
7.34
7.33
4.66
4.65
4.65
4.05
4.04
4.04
3.97
3.97
3.97
3.96

2.09
2.00
1.93
2.01
1.97

13　12　11　10　9　8　7　6　5　4　3　2　1　0　-1
f1 (ppm)

166.54
165.58
149.13
146.15
122.86
49.32
44.25

220　210　200　190　180　170　160　150　140　130　120　110　100　90　80　70　60　50　40　30　20　10　0　-10
f1 (ppm)

218

D22

219

E1

5.68

2.23

1.63

6.00

13   12   11   10   9   8   7   6   5   4   3   2   1   0   -1
f1 (ppm)

169.66
163.25

72.04

20.05

230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

F1

PROTON_cd3od_T25_001
Gradient Shimming

8.76
8.75
8.33
8.32
8.13
8.12
8.12
8.11
8.10
8.10
7.48
7.47
7.46
7.45
5.73
5.72
5.69
5.68
2.20
2.10

0.85
0.91
0.92
0.96
0.96
0.95
3.34
3.02

f1 (ppm)

CARBON_cd3od_T25_001
Gradient Shimming

170.01
169.95
168.40
166.14
144.11
140.75
135.18
129.08
124.14
72.63
71.15
19.00
18.90

f1 (ppm)

221

F2

8.54
8.54
8.53
8.53
7.47
7.47
7.46
7.46
5.70
5.69
5.58
5.57
4.57
4.53
4.48
4.44
2.17
2.07

2.06   2.00   0.95 0.95   1.03 1.04   2.99 3.04

f1 (ppm)
13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1

170.08
169.89
168.86
168.04
151.59
146.97
122.98
72.64
71.66
41.57
19.08
19.03

f1 (ppm)
230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10

222

F3

F4

224

F5

5.88 5.88 5.87 5.87 5.59 5.58 5.58 5.57 3.59 3.59 3.58 3.57 3.57 3.56 3.52 3.50 3.52 2.12 1.71 1.68 1.55 1.53

1.00 0.98   4.31   2.90 2.95   6.42

f1 (ppm)

170.04 169.92 168.83 164.14   69.97 68.92   46.87 44.01   26.06 25.31 24.28 20.44 20.41

f1 (ppm)

225

G1

9.18 9.15 9.14 8.65 8.63 8.62 8.6 7.78 7.77 7.76 7.75 7.73

5.18 5.17 4.66 4.65 3.80 3.78 3.78 3.76 3.75 3.74 3.73 3.72 3.71 3.70 3.68 3.66 3.65 3.64 3.62 3.61 3.60 3.59 3.58

1.00 0.92 1.16 0.98 0.68 0.72 8.39

f1 (ppm)

171.74 170.23 140.10 136.41 136.31 132.73 125.89 72.44 69.38 66.15 66.08 45.82 42.85

f1 (ppm)

226

G2

PROTON_acetic_acid_T25_001
Gradient Shimming

8.34
8.32
8.02
8.00
7.98
7.29
5.15
4.66
3.81
3.79
3.78
3.76
3.73
3.69
3.63

1.68
0.93
1.00
0.52
0.60
8.31

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

CARBON_acetic_acid_T25_001
Gradient Shimming

171.30
169.92
149.92
145.10
141.02
120.57
115.38
72.21
69.70
69.46
66.16
45.71
42.92

230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

227

G3

8.19
8.19
8.18
8.17
8.16
7.58
7.57
7.56
7.55
7.53
5.21
4.64
4.63
4.61
3.79
3.77
3.75
3.69

1.00
2.03
0.74
0.97
8.41

13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1
f1 (ppm)

171.48
170.49
165.53
163.16
149.52
149.37
147.20
147.08
112.27
99.29
98.88
72.48
69.30
66.08
52.45
45.90
42.94

230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

228

H1

PROTON_cd3od_T25_001
Gradient Shimming

8.86
8.86
8.85
8.84
8.84
8.08
8.07

5.76
5.75
5.10
5.09

2.18

2.01
2.00
1.82
2.11
6.04

f1 (ppm)

CARBON_cd3od_T25_001
Gradient Shimming

170.49
169.54
156.67
141.49
125.67
73.01
41.10
18.77

f1 (ppm)

H2

7.33
7.31
7.26
7.24

5.69

2.43
2.27

2.13
1.95

2.13

3.31
6.00

14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2
f1 (ppm)

170.00
168.59

139.48

129.98
128.08
125.97

72.85

21.27
20.44

210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

230

H3

7.31
7.29
7.29
7.27
7.19
7.17

5.69

2.27
2.22

14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2
f1 (ppm)

1.77
2.16
2.10
6.00
6.07

170.12
168.20

136.07
129.99
128.96
128.69

73.17

20.46
17.72

210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

H4

7.29
7.26
7.06
7.04
5.88
3.86
2.21

2.00
2.03
1.90
3.28
5.92

14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2
f1 (ppm)

170.26
169.55
160.03
127.54
123.83
113.92
72.78
54.59
18.85

210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

232

16. Bioactivity assays

Kinase profiling studies were carried out at International centre for kinase profiling.[210] Cell viability (MTS) assays, anti-oxidant activity (CLPAA) assays, antibacterial activity assays and biofilm assays were conducted at MarBio, Tromsø.

16.1. 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).

16.2. 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.

16.3. 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

16.4. 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.

16.5. SILAC study

16.5.1. SILAC labelling

The SILAC labeling was performed using the components from the Pierce SILAC Quantitation Kit (Fisher Scientific, 89983).

A2058 cells (ATCC CRL-11147) were grown in one flask with ^{12}C and one flask with ^{13}C SILAC Dulbecco's Modified Eagle Medium, according to the protocol from Thermo Scientific. In addition to the supplements provided in the kit, 50 mg/ml proline and 10 µg/ml gentamicin (Biochrom, A2712) were added to the growth media before sterile filtration. The cells were grown in these two media for 2 weeks. Then the cells were seeded in 3 cell cultivation plates (VWR, 734-2043) per growth media, each with 2.2 X 10^6 cells. The cells were incubated overnight at 5 % CO$_2$, 37 ºC. Then ^{13}C labelled (test) cells were incubated for 4 hours with compound B18 (final concentration 5, 2.5 or 0.5 µM). The ^{12}C labelled (control) cells were incubated with the same volume of dilution buffer for 4 hours. After 4 hours of incubation the cells were washed twice with ice cold PBS. The cells were then lysed with 250 µl 2.5 X SDS lysis buffer. By use of a cell scraper, the cell lysate was transferred to a LoBind Eppendorf tube. The cell lysates were stored at -20 ºC.

After thawing the samples on ice, 100 mM DTT (final concentration) was added to the samples before the samples were incubated at 95 degree C for 15 min. The lysis buffer makes the samples very viscous, and the high temperature makes them liquid again. 10 % of the lysate was removed for analysis of ^{13}C incorporation. The rest of the lysate was mixed 1:1 for ^{12}C and ^{13}C (i.e. ^{13}C 5 µM B18 with ^{12}C treated with the same amount of dilution buffer, and so on).

The samples were separated on a polyacrylamide gel for SDS-PAGE (Thermo Scientific, 25204). The samples were run approximately 10 mm into the resolving part of the gel. The gel was stained using Simply Blue safe stain (Thermo Scientific, LC 6065) before the stained area was cut into 3 slices that were placed into high quality Eppendorf tubes. The cutting was performed under a keratin fan to avoid contamination of the samples.

16.5.2. LC-MS/MS

Gel pieces were subjected to in gel reduction, alkylation, and tryptic digestion using 6 ng/µl trypsin (V511A, Promega, Wisconsin, USA).[211] OMIX C18 tips (Varian, Inc., Palo Alto, CA, USA) was used for sample cleanup and concentration. Peptide mixtures containing 0.1% formic acid were loaded onto a Thermo Fisher Scientific EASY-nLC1000 system and EASY-Spray column (C18, 2µm, 100 Å, 75µm, 50 cm). Peptides were fractionated using a 2-100% acetonitrile gradient in 0.1 % formic acid over 200 min at a flow rate of 250 nl/min. The separated peptides was analysed using a Thermo Scientific Q-Exactive mass spectrometer. Data was collected in data dependent mode using a Top10 method.

16.5.3. SILAC Quantitation

Raw files from the Q-Exactive MS were analysed using the quantitative proteomics software MaxQuant[212] (version 1.5.6.0). SILAC pairs were quantitated in MaxQuant and proteins were identified using the built in Andromeda search engine using the Uniprot Homo sapiens (Human) database (november.2016). Main search peptide tolerance was set to 4.5 ppm and MS/MS mass tolerance was set to 20 ppm. A FDR ration of 0.01 were needed to give å protein identification. At least 2 peptides had to be quantitated to give a quantitation value.

Statistical validation of protein regulation was done with the Perseus 1.5.6.0 software. To determine significant outliers in the experiments a significance B test was performed. This was done with the Benjamini–Hochberg procedure with a FDR of 0.05 on normalized ratios.

References

.(1)   Hughes, J.; Rees, S.; Kalindjian, S.; Philpott, K. Principles of Early Drug Discovery. *Br. J. Pharmacol.* **2011**, *162* (6), 1239–1249.

(2)   Rick Mullin. Cost to Develop New Pharmaceutical Drug Now Exceeds $2.5B http://www.scientificamerican.com/article/cost-to-develop-new-pharmaceutical-drug-now-exceeds-2-5b/ (accessed Dec 22, 2015).

(3)   PR Tufts CSDD 2014 Cost Study | Tufts Center for the Study of Drug Development http://csdd.tufts.edu/news/complete_story/pr_tufts_csdd_2014_cost_study (accessed Dec 22, 2015).

(4)   DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. The Cost of Drug Development. *N. Engl. J. Med.* **2015**, *372* (20), 1972–1972.

(5)   Nicolaou, K. C. Advancing the Drug Discovery and Development Process. *Angew. Chem. Int. Ed.* **2014**, *53* (35), 9128–9140.

(6)   Pammolli, F.; Magazzini, L.; Riccaboni, M. The Productivity Crisis in Pharmaceutical R&D. *Nat. Rev. Drug Discov.* **2011**, *10* (6), 428–438.

(7)   Liu, H.; Schmid, M. B. Maturation of the Biotechnology Industry Changes Job Opportunities for Scientists. *J. Commer. Biotechnol.* **2008**, *15* (3), 199–214.

(8)   Hartung, T. Food for Thought Look Back in Anger – What Clinical Studies Tell Us About Preclinical Work. *ALTEX* **2013**, *30* (3), 275–291.

(9)   Begley, C. G.; Ellis, L. M. Drug Development: Raise Standards for Preclinical Cancer Research. *Nature* **2012**, *483* (7391), 531–533.

(10)   Harrison, C. Phenotypic Screening: A More Rapid Route to Target Deconvolution. *Nat. Rev. Drug Discov.* **2014**, *13* (2), 102–103.

(11)   Moffat, J. G.; Rudolph, J.; Bailey, D. Phenotypic Screening in Cancer Drug Discovery — Past, Present and Future. *Nat. Rev. Drug Discov.* **2014**, *13* (8), 588–602.

(12)   Priest, B. T.; Erdemli, G. Phenotypic Screening in the 21st Century. *Exp. Pharmacol. Drug Discov.* **2014**, *5*, 264.

(13)   Zheng, W.; Thorne, N.; McKew, J. C. Phenotypic Screens as a Renewed Approach for Drug Discovery. *Drug Discov. Today* **2013**, *18* (21–22), 1067–1073.

(14)   Zanella, F.; Lorens, J. B.; Link, W. High Content Screening: Seeing Is Believing. *Trends Biotechnol.* **2010**, *28* (5), 237–245.

(15)   Martin, H. L.; Adams, M.; Higgins, J.; Bond, J.; Morrison, E. E.; Bell, S. M.; Warriner, S.; Nelson, A.; Tomlinson, D. C. High-Content, High-Throughput Screening for the Identification of Cytotoxic Compounds Based on Cell Morphology and Cell Proliferation Markers. *PLoS ONE* **2014**, *9* (2), e88338.

(16)   Acker, M. G.; Auld, D. S. Considerations for the Design and Reporting of Enzyme Assays in High-Throughput Screening Applications. *Perspect. Sci.* **2014**, *1* (1–6), 56–73.

(17)   Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R. P.; Foloppe, N. CURRENT PROGRESS IN STRUCTURE-BASED RATIONAL DRUG DESIGN MARKS A NEW MINDSET IN DRUG DISCOVERY. *Comput. Struct. Biotechnol. J.* **2013**, *5* (6), 1–14.

(18)   Verma, S.; Prabhakar, Y. S. Target Based Drug Design - a Reality in Virtual Sphere. *Curr. Med. Chem.* **2015**, *22* (13), 1603–1630.

(19)   Anderson, A. C. The Process of Structure-Based Drug Design. *Chem. Biol.* **2003**, *10* (9), 787–797.

(20)   Wang, T.; Wu, M.-B.; Chen, Z.-J.; Chen, H.; Lin, J.-P.; Yang, L.-R. Fragment-Based Drug Discovery and Molecular Docking in Drug Design. *Curr. Pharm. Biotechnol.* **2015**, *16* (1), 11–25.

(21)   Scott, D. E.; Coyne, A. G.; Hudson, S. A.; Abell, C. Fragment-Based Approaches in Drug Discovery and Chemical Biology. *Biochemistry (Mosc.)* **2012**, *51* (25), 4990–5003.

(22)   Joseph-McCarthy, D.; Campbell, A. J.; Kern, G.; Moustakas, D. Fragment-Based Lead Discovery and Design. *J. Chem. Inf. Model.* **2014**, *54* (3), 693–704.

(23)   Harner, M. J.; Frank, A. O.; Fesik, S. W. Fragment-Based Drug Discovery Using NMR Spectroscopy. *J. Biomol. NMR* **2013**, *56* (2), 65–75.

(24)   Zhang, K. Y. J.; Milburn, M. V.; Artis, D. R. Scaffold-Based Drug Discovery. In *Structure-Based Drug Discovery*; Springer Netherlands, 2007; pp 129–153.

(25)   Kirkpatrick, P. Small Is Beautiful. *Nat. Rev. Drug Discov.* **2005**, *4* (3), 190–190.

(26)   Welsch, M. E.; Snyder, S. A.; Stockwell, B. R. Privileged Scaffolds for Library Design and Drug Discovery. *Curr. Opin. Chem. Biol.* **2010**, *14* (3), 347–361.

(27)   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. A Family of Phosphodiesterase Inhibitors Discovered by

Cocrystallography and Scaffold-Based Drug Design. *Nat. Biotechnol.* **2005**, *23* (2), 201–207.

(28)   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. Phylogenies Reveal Predictive Power of Traditional Medicine in Bioprospecting. *Proc. Natl. Acad. Sci.* **2012**, *109* (39), 15835–15840.

(29)   Harvey, A. L.; Gericke, N. Bioprospecting: Creating a Value for Biodiversity. In *Research in Biodiversity - Models and Applications*; Pavlinov, I., Ed.; InTech: Rijeka, Croatia, 2011.

(30)   Bailey, F.; Dundas, I. Bioprospecting: Discoveries Changing the Future. *Aust. TPOTCO Ed Canberra House Represent. Standing Comm. Prim. Ind. Reg. Serv.* **2001**.

(31)   Galloway, W. R. J. D.; Isidro-Llobet, A.; Spring, D. R. Diversity-Oriented Synthesis as a Tool for the Discovery of Novel Biologically Active Small Molecules. *Nat. Commun.* **2010**, *1*, 80.

(32)   Wassermann, A. M.; Camargo, L. M.; Auld, D. S. Composition and Applications of Focus Libraries to Phenotypic Assays. *Front. Pharmacol.* **2014**, *5*, 164.

(33)   Petrone, P. M.; Wassermann, A. M.; Lounkine, E.; Kutchukian, P.; Simms, B.; Jenkins, J.; Selzer, P.; Glick, M. Biodiversity of Small Molecules – a New Perspective in Screening Set Selection. *Drug Discov. Today* **2013**, *18* (13–14), 674–680.

(34)   Hert, J.; Irwin, J. J.; Laggner, C.; Keiser, M. J.; Shoichet, B. K. Quantifying Biogenic Bias in Screening Libraries. *Nat. Chem. Biol.* **2009**, *5* (7), 479–483.

(35)   Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How Many Drug Targets Are There? *Nat. Rev. Drug Discov.* **2006**, *5* (12), 993–996.

(36)   Frearson, J. A.; Collie, I. T. HTS and Hit Finding in Academia – from Chemical Genomics to Drug Discovery. *Drug Discov. Today* **2009**, *14* (23–24), 1150–1158.

(37)   Tan, D. S. Diversity-Oriented Synthesis: Exploring the Intersections between Chemistry and Biology. *Nat. Chem. Biol.* **2005**, *1* (2), 74–84.

(38)   Brüstle, M.; Beck, B.; Schindler, T.; King, W.; Mitchell, T.; Clark, T. Descriptors, Physical Properties, and Drug-Likeness. *J. Med. Chem.* **2002**, *45* (16), 3345–3355.

(39)   Walters, W. P.; Murcko, M. A. Prediction of "drug-Likeness." *Adv. Drug Deliv. Rev.* **2002**, *54* (3), 255–271.

(40)   Ursu, O.; Rayan, A.; Goldblum, A.; Oprea, T. I. Understanding Drug-Likeness. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1* (5), 760–781.

(41)   Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **1997**, *23* (1–3), 3–25.

(42)   Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A "Rule of Three" for Fragment-Based Lead Discovery? *Drug Discov. Today* **2003**, *8* (19), 876–877.

(43)   Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A. L. Quantifying the Chemical Beauty of Drugs. *Nat. Chem.* **2012**, *4* (2), 90–98.

(44)   Cherkasov, A. Can "Bacterial-Metabolite-Likeness" Model Improve Odds of "in Silico" Antibiotic Discovery? *J. Chem. Inf. Model.* **2006**, *46* (3), 1214–1222.

(45)   Vistoli, G.; Pedretti, A.; Testa, B. Assessing Drug-Likeness--What Are We Missing? *Drug Discov. Today* **2008**, *13* (7–8), 285–294.

(46)   Macarrón, R.; Luengo, J. I. Yin and Yang in Medicinal Chemistry: What Does Drug-Likeness Mean? *Future Med. Chem.* **2011**, *3* (5), 505–507.

(47)   Vagner, J.; Qu, H.; Hruby, V. J. Peptidomimetics, a Synthetic Tool of Drug Discovery. *Curr. Opin. Chem. Biol.* **2008**, *12* (3), 292–296.

(48)   Trabocchi, A.; Guarna, A. The Basics of Peptidomimetics. In *Peptidomimetics in Organic and Medicinal Chemistry*; John Wiley & Sons, Ltd, 2014; pp 1–17.

(49)   Avan, I.; Hall, C. D.; Katritzky, A. R. Peptidomimetics via Modifications of Amino Acids and Peptide Bonds. *Chem. Soc. Rev.* **2014**, *43* (10), 3575–3594.

(50)   Ruzza, P. Peptides and Peptidomimetics in Medicinal Chemistry. In *Medicinal Chemistry and Drug Design*; Ekinci, D., Ed.; InTech, 2012.

(51)   Alberto López-Cobeñas, P. C. Microwave-Assisted Synthesis of 2,5-Piperazinediones under Solvent-Free Conditions. *Synthesis* **2005**, *2005* (19), 3412.

(52)   Pérez-Picaso, L.; Escalante, J.; Olivo, H. F.; Rios, M. Y. Efficient Microwave Assisted Syntheses of 2,5-Diketopiperazines in Aqueous Media. *Molecules* **2009**, *14* (8), 2836–2849.

(53)   Pachaly, P.; Pelzer, H.-J. 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)* **1983**, *316* (7), 653–655.

(54)   Tadesse, M.; Strøm, M. B.; Svenson, J.; Jaspars, M.; Milne, B. F.; Tørfoss, V.; Andersen, J. H.; Hansen, E.; Stensvåg, K.; Haug, T. Synoxazolidinones A and B: Novel Bioactive Alkaloids from the Ascidian Synoicum Pulmonaria. *Org. Lett.* **2010**, *12* (21), 4752–4755.

(55)   Trepos, R.; Cervin, G.; Hellio, C.; Pavia, H.; Stensen, W.; Stensvåg, K.; Svendsen, J.-S.; Haug, T.; Svenson, J. Antifouling Compounds from the Sub-Arctic Ascidian Synoicum Pulmonaria: Synoxazolidinones A and C, Pulmonarins A and B, and Synthetic Analogues. *J. Nat. Prod.* **2014**, *77* (9), 2105–2113.

(56)   Tadesse, M.; Svenson, J.; Jaspars, M.; Strøm, M. B.; Abdelrahman, M. H.; Andersen, J. H.; Hansen, E.; Kristiansen, P. E.; Stensvag, K.; Haug, T. Synoxazolidinone C; a Bicyclic Member of the Synoxazolidinone Family with Antibacterial and Anticancer Activities. *Tetrahedron Lett.* **2011**, *52* (15), 1804–1806.

(57)   Hanssen, K. Ø.; Andersen, J. H.; Stiberg, T.; Engh, R. A.; Svenson, J.; Genevière, A.-M.; Hansen, E. Antitumoral and Mechanistic Studies of Ianthelline Isolated from the Arctic Sponge Stryphnus Fortis. *Anticancer Res.* **2012**, *32* (10), 4287–4297.

(58)   Lind, K. F.; Hansen, E.; Østerud, B.; Eilertsen, K.-E.; Bayer, A.; Engqvist, M.; Leszczak, K.; Jørgensen, T. Ø.; Andersen, J. H. Antioxidant and Anti-Inflammatory Activities of Barettin. *Mar. Drugs* **2013**, *11* (7), 2655–2666.

(59)   Borthwick, A. D. 2,5-Diketopiperazines: Synthesis, Reactions, Medicinal Chemistry, and Bioactive Natural Products. *Chem. Rev.* **2012**, *112* (7), 3641–3716.

(60)   Martins, M. B.; Carvalho, I. Diketopiperazines: Biological Activity and Synthesis. *Tetrahedron* **2007**, *63* (40), 9923–9932.

(61)   Daugan, A.; Grondin, P.; Ruault, C.; Le Monnier de Gouville, A.-C.; Coste, H.; Linget, J. M.; Kirilovsky, J.; Hyafil, F.; Labaudinière, R. 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.* **2003**, *46* (21), 4533–4542.

(62)   Fintan, K.; Richard, T. L.; Angus, M. M.; Kevin, J. M. Piperazine Derivatives. GB2271774 (A), April 27, 1994.

(63)   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. 2,5-Diketopiperazines for the Treatment of Obesity. WO2004048345 A3, July 15, 2004.

(64)   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. Pyridyl-2,5-Diketopiperazines as Potent, Selective, and Orally Bioavailable Oxytocin Antagonists:

Synthesis, Pharmacokinetics, and In Vivo Potency. *J. Med. Chem.* **2012**, *55* (2), 783–796.

(65)   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. 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.* **2007**, *50* (8), 1787–1798.

(66)   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. 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.* **2008**, *51* (10), 2924–2932.

(67)   Szardenings, A. K.; Harris, D.; Lam, S.; Shi, L.; Tien, D.; Wang, Y.; Patel, D. V.; Navre, M.; Campbell, D. A. Rational Design and Combinatorial Evaluation of Enzyme Inhibitor Scaffolds: Identification of Novel Inhibitors of Matrix Metalloproteinases. *J. Med. Chem.* **1998**, *41* (13), 2194–2200.

(68)   Szardenings, A. K.; Antonenko, V.; Campbell, D. A.; DeFrancisco, N.; Ida, S.; Shi, L.; Sharkov, N.; Tien, D.; Wang, Y.; Navre, M. Identification of Highly Selective Inhibitors of Collagenase-1 from Combinatorial Libraries of Diketopiperazines. *J. Med. Chem.* **1999**, *42* (8), 1348–1357.

(69)   Campbell, D.; Look, G. C.; Szardenings, A. K.; Patel, D. V. Collagenase-1 and Stromelysin-1 Inhibitors, Pharmaceutical Compositions Comprising Same and Methods of Their Use. US5932579 A, August 3, 1999.

(70)   Qiao, Y.; Gao, J.; Qiu, Y.; Wu, L.; Guo, F.; Lo, K. K.-W.; Li, D. Design, Synthesis, and Characterization of Piperazinedione-Based Dual Protein Inhibitors for Both Farnesyltransferase and Geranylgeranyltransferase-I. *Eur. J. Med. Chem.* **2011**, *46* (6), 2264–2273.

(71)   Zhu, Y.; Tang, M.; Shi, X.; Zhao, Y. Quantum Chemical Study of Cyclic Dipeptides. *Int. J. Quantum Chem.* **2007**, *107* (3), 745–753.

(72)   MacDonald, J. C.; Whitesides, G. M. Solid-State Structures of Hydrogen-Bonded Tapes Based on Cyclic Secondary Diamides. *Chem. Rev.* **1994**, *94* (8), 2383–2420.

(73)   Mendham, A. P.; Potter, B. S.; Palmer, R. A.; Dines, T. J.; Mitchell, J. C.; Withnall, R.; Chowdhry, B. Z. Vibrational Spectra and Crystal Structure of the Di-Amino Acid Peptide cyclo(L-Met-L-Met): Comparison of Experimental Data and DFT Calculations. *J. Raman Spectrosc.* **2010**, *41* (2), 148–159.

(74)   Cohen, P. Protein Kinases--the Major Drug Targets of the Twenty-First Century? *Nat. Rev. Drug Discov.* **2002**, *1* (4), 309–315.

(75)   Chène, P. ATPases as Drug Targets: Learning from Their Structure. *Nat. Rev. Drug Discov.* **2002**, *1* (9), 665–673.

(76)   Montalbetti, C. A. G. N.; Falque, V. Amide Bond Formation and Peptide Coupling. *Tetrahedron* **2005**, *61* (46), 10827–10852.

(77)   Xu, C.-P.; Xiao, Z.-H.; Zhuo, B.-Q.; Wang, Y.-H.; Huang, P.-Q. Efficient and Chemoselective Alkylation of Amines/Amino Acids Using Alcohols as Alkylating Reagents under Mild Conditions. *Chem. Commun.* **2010**, *46* (41), 7834–7836.

(78)   Murahashi, S.-I.; Shimamura, T.; Moritani, I. Conversion of Alcohols into Unsymmetrical Secondary or Tertiary Amines by a Palladium Catalyst. Synthesis of N-Substituted Pyrroles. *J. Chem. Soc. Chem. Commun.* **1974**, No. 22, 931–932.

(79) Zubrzak, P.; Leplawy, M. T.; Kowalski, M. L.; Szkudlińska, B.; Paneth, P.; Silberring, J.; Suder, P.; Zabrocki, J. Correlating Biological Activity with Calculated Geometric Motifs in Cyclolinopeptide A Analogs. *J. Phys. Org. Chem.* **2004**, *17* (6–7), 625–630.

(80) Hannachi, J.-C.; Vidal, J.; Mulatier, J.-C.; Collet, A. Electrophilic Amination of Amino Acids with N-Boc-Oxaziridines: Efficient Preparation of N-Orthogonally Diprotected Hydrazino Acids and Piperazic Acid Derivatives. *J. Org. Chem.* **2004**, *69* (7), 2367–2373.

(81) Verardo, G.; Geatti, P.; Pol, E.; Giumanini, A. G. Sodium Borohydride: A Versatile Reagent in the Reductive N-Monoalkylation of α-Amino Acids and α-Amino Methyl Esters. *Can. J. Chem.* **2002**, *80* (7), 779–788.

(82) Park, J. D.; Kim, D. H. Cysteine Derivatives as Inhibitors for Carboxypeptidase A: Synthesis and Structure-Activity Relationships. *J. Med. Chem.* **2002**, *45* (4), 911–918.

(83) Dunn, P. J.; Haener, R.; Rapoport, H. Stereoselective Synthesis of 2,3-Diamino Acids. 2,3-Diamino-4-Phenylbutanoic Acid. *J. Org. Chem.* **1990**, *55* (17), 5017–5025.

(84) Li, X.; Zhao, M.; Tang, Y.-R.; Wang, C.; Zhang, Z.; Peng, S. N-[2-(5,5-Dimethyl-1,3-Dioxane-2-Yl)ethyl]amino Acids: Their Synthesis, Anti-Inflammatory Evaluation and QSAR Analysis. *Eur. J. Med. Chem.* **2008**, *43* (1), 8–18.

(85) Liao, W.; Chen, Y.; Liu, Y.; Duan, H.; Petersen, J. L.; Shi, X. 1,2,3-Triazole-Boranes: Stable and Efficient Reagents for Ketone and Aldehyde Reductive Amination in Organic Solvents or in Water. *Chem. Commun.* **2009**, No. 42, 6436–6438.

(86) Mao, F.; Ni, W.; Xu, X.; Wang, H.; Wang, J.; Ji, M.; Li, J. Chemical Structure-Related Drug-Like Criteria of Global Approved Drugs. *Mol. Basel Switz.* **2016**, *21* (1), 75.

(87) Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* **2002**, *1* (5), 376–386.

(88) Mann, M. Functional and Quantitative Proteomics Using SILAC. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (12), 952–958.

(89) Mi, H.; Huang, X.; Muruganujan, A.; Tang, H.; Mills, C.; Kang, D.; Thomas, P. D. PANTHER Version 11: Expanded Annotation Data from Gene Ontology and Reactome Pathways, and Data Analysis Tool Enhancements. *Nucleic Acids Res.* **2017**, *45* (D1), D183–D189.

(90) Mi, H.; Thomas, P. PANTHER Pathway: An Ontology-Based Pathway Database Coupled with Data Analysis Tools. In *Protein Networks and Pathway Analysis*; Nikolsky, Y., Bryant, J., Eds.; Methods in Molecular Biology; Humana Press, 2009; pp 123–140.

(91) DeClerck, Y. A.; Mercurio, A. M.; Stack, M. S.; Chapman, H. A.; Zutter, M. M.; Muschel, R. J.; Raz, A.; Matrisian, L. M.; Sloane, B. F.; Noel, A.; Hendrix, M. J.; Coussens, L.; Padarathsingh, M. Proteases, Extracellular Matrix, and Cancer. *Am. J. Pathol.* **2004**, *164* (4), 1131–1139.

(92) López-Otín, C.; Matrisian, L. M. Emerging Roles of Proteases in Tumour Suppression. *Nat. Rev. Cancer* **2007**, *7* (10), 800–808.

(93) Kauvar, L. M.; Higgins, D. L.; Villar, H. O.; Sportsman, J. R.; Engqvist-Goldstein, A.; Bukar, R.; Bauer, K. E.; Dilley, H.; Rocke, D. M. Predicting Ligand Binding to Proteins by Affinity Fingerprinting. *Chem. Biol.* **1995**, *2* (2), 107–118.

(94) Schneider, G.; Tanrikulu, Y.; Schneider, P. Self-Organizing Molecular Fingerprints: A Ligand-Based View on Drug-like Chemical Space and off-Target Prediction. *Future Med. Chem.* **2009**, *1* (1), 213–218.

(95)   Peragovics, Á.; Simon, Z.; Tombor, L.; Jelinek, B.; Hári, P.; Czobor, P.; Málnási-Csizmadia, A. Virtual Affinity Fingerprints for Target Fishing: A New Application of Drug Profile Matching. *J. Chem. Inf. Model.* **2013**, *53* (1), 103–113.

(96)   Wang, Y.; Zeng, J. Predicting Drug-Target Interactions Using Restricted Boltzmann Machines. *Bioinformatics* **2013**, *29* (13), i126–i134.

(97)   Alvarsson, J.; Eklund, M.; Engkvist, O.; Spjuth, O.; Carlsson, L.; Wikberg, J. E. S.; Noeske, T. Ligand-Based Target Prediction with Signature Fingerprints. *J. Chem. Inf. Model.* **2014**, *54* (10), 2647–2653.

(98)   Mervin, L. H.; Afzal, A. M.; Drakakis, G.; Lewis, R.; Engkvist, O.; Bender, A. Target Prediction Utilising Negative Bioactivity Data Covering Large Chemical Space. *J. Cheminformatics* **2015**, *7* (1), 51.

(99)   Cao, R.; Wang, Y. Predicting Molecular Targets for Small-Molecule Drugs with a Ligand-Based Interaction Fingerprint Approach. *ChemMedChem* **2016**, *11* (12), 1352–1361.

(100) Wang, Z.; Liang, L.; Yin, Z.; Lin, J. Improving Chemical Similarity Ensemble Approach in Target Prediction. *J. Cheminformatics* **2016**, *8*.

(101) Urich, R.; Wishart, G.; Kiczun, M.; Richters, A.; Tidten-Luksch, N.; Rauh, D.; Sherborne, B.; Wyatt, P. G.; Brenk, R. De Novo Design of Protein Kinase Inhibitors by in Silico Identification of Hinge Region-Binding Fragments. *ACS Chem. Biol.* **2013**, *8* (5), 1044–1052.

(102) Rajappa, S.; Natekar, M. V. Piperazine-2,5-Diones and Related Lactim Ethers. In *Advances in Heterocyclic Chemistry*; Katritzky, A. R., Ed.; Academic Press, 1993; Vol. 57, pp 187–289.

(103) Kishi, M.; Pan, Y. A.; Crump, J. G.; Sanes, J. R. Mammalian SAD Kinases Are Required for Neuronal Polarization. *Science* **2005**, *307* (5711), 929–932.

(104) Bright, N. J.; Carling, D.; Thornton, C. Investigating the Regulation of Brain-Specific Kinases 1 and 2 by Phosphorylation. *J. Biol. Chem.* **2008**, *283* (22), 14946–14954.

(105) Barnes, A. P.; Lilley, B. N.; Pan, Y. A.; Plummer, L. J.; Powell, A. W.; Raines, A. N.; Sanes, J. R.; Polleux, F. LKB1 and SAD Kinases Define a Pathway Required for the Polarization of Cortical Neurons. *Cell* **2007**, *129* (3), 549–563.

(106) Crump, J. G.; Zhen, M.; Jin, Y.; Bargmann, C. I. The SAD-1 Kinase Regulates Presynaptic Vesicle Clustering and Axon Termination. *Neuron* **2001**, *29* (1), 115–129.

(107) Inoue, E.; Mochida, S.; Takagi, H.; Higa, S.; Deguchi-Tawarada, M.; Takao-Rikitsu, E.; Inoue, M.; Yao, I.; Takeuchi, K.; Kitajima, I.; Setou, M.; Ohtsuka, T.; Takai, Y. SAD: A Presynaptic Kinase Associated with Synaptic Vesicles and the Active Zone Cytomatrix That Regulates Neurotransmitter Release. *Neuron* **2006**, *50* (2), 261–275.

(108) Lyn-Adams, C. L. The Regulation of Tau-Dependent Neurodegeneration by Brain Selective/SAD Kinases. phd, University of Warwick, 2011.

(109) Alvarado-Kristensson, M.; Rodríguez, M. J.; Silió, V.; Valpuesta, J. M.; Carrera, A. C. SADB Phosphorylation of Gamma-Tubulin Regulates Centrosome Duplication. *Nat. Cell Biol.* **2009**, *11* (9), 1081–1092.

(110) Carrera, A. C.; Alvarado-Kristensson, M. SADB Kinases License Centrosome Replication. *Cell Cycle Georget. Tex* **2009**, *8* (24), 4005–4006.

(111) Chen, D.; Vogel, J. SAD Kinase Keeps Centrosomes Lonely. *Nat. Cell Biol.* **2009**, *11* (9), 1047–1048.

(112) Lu, R.; Niida, H.; Nakanishi, M. Human SAD1 Kinase Is Involved in UV-Induced DNA Damage Checkpoint Function. *J. Biol. Chem.* **2004**, *279* (30), 31164–31170.

(113) Margaret, B. L.; A, B. R.; Arthur, B.; S, F. L.; Susana, N.-B.; Kevin, W. Brsk1s as Modifiers of the Pten/Akt Pathway and Methods of Use. WO2006099185 A2, September 21, 2006.

(114) Nakagawa, O.; Fujisawa, K.; Ishizaki, T.; Saito, Y.; Nakao, K.; Narumiya, S. ROCK-I and ROCK-II, Two Isoforms of Rho-Associated Coiled-Coil Forming Protein Serine/Threonine Kinase in Mice. *FEBS Lett.* **1996**, *392* (2), 189–193.

(115) Pireddu, R.; Forinash, K. D.; Sun, N. N.; Martin, M. P.; Sung, S.-S.; Alexander, B.; Zhu, J.-Y.; Guida, W. C.; Schönbrunn, E.; Sebti, S. M.; Lawrence, N. J. Pyridylthiazole-Based Ureas as Inhibitors of Rho Associated Protein Kinases (ROCK1 and 2). *MedChemComm* **2012**, *3* (6), 699–709.

(116) Amin, E.; Dubey, B. N.; Zhang, S.-C.; Gremer, L.; Dvorsky, R.; Moll, J. M.; Taha, M. S.; Nagel-Steger, L.; Piekorz, R. P.; Somlyo, A. V.; Ahmadian, M. R. Rho-Kinase: Regulation, (Dys)function, and Inhibition. *Biol. Chem.* **2013**, *394* (11), 1399–1410.

(117) Amano, M.; Chihara, K.; Kimura, K.; Fukata, Y.; Nakamura, N.; Matsuura, Y.; Kaibuchi, K. Formation of Actin Stress Fibers and Focal Adhesions Enhanced by Rho-Kinase. *Science* **1997**, *275* (5304), 1308–1311.

(118) Kawano, Y.; Fukata, Y.; Oshiro, N.; Amano, M.; Nakamura, T.; Ito, M.; Matsumura, F.; Inagaki, M.; Kaibuchi, K. Phosphorylation of Myosin-Binding Subunit (MBS) of Myosin Phosphatase by Rho-Kinase in Vivo. *J. Cell Biol.* **1999**, *147* (5), 1023–1038.

(119) Amano, M.; Ito, M.; Kimura, K.; Fukata, Y.; Chihara, K.; Nakano, T.; Matsuura, Y.; Kaibuchi, K. Phosphorylation and Activation of Myosin by Rho-Associated Kinase (Rho-Kinase). *J. Biol. Chem.* **1996**, *271* (34), 20246–20249.

(120) Fukata, Y.; Amano, M.; Kaibuchi, K. Rho-Rho-Kinase Pathway in Smooth Muscle Contraction and Cytoskeletal Reorganization of Non-Muscle Cells. *Trends Pharmacol. Sci.* **2001**, *22* (1), 32–39.

(121) Wojciak-Stothard, B.; Ridley, A. J. Rho GTPases and the Regulation of Endothelial Permeability. *Vascul. Pharmacol.* **2002**, *39* (4–5), 187–199.

(122) Wójciak-Stothard, B.; Potempa, S.; Eichholtz, T.; Ridley, A. J. Rho and Rac but Not Cdc42 Regulate Endothelial Cell Permeability. *J. Cell Sci.* **2001**, *114* (Pt 7), 1343–1355.

(123) Farah, S.; Agazie, Y.; Ohan, N.; Ngsee, J. K.; Liu, X. J. A Rho-Associated Protein Kinase, ROKalpha, Binds Insulin Receptor Substrate-1 and Modulates Insulin Signaling. *J. Biol. Chem.* **1998**, *273* (8), 4740–4746.

(124) Mueller, B. K.; Mack, H.; Teusch, N. Rho Kinase, a Promising Drug Target for Neurological Disorders. *Nat. Rev. Drug Discov.* **2005**, *4* (5), 387–398.

(125) Shi, J.; Wei, L. Rho Kinases in Cardiovascular Physiology and Pathophysiology: The Effect of Fasudil. *J. Cardiovasc. Pharmacol.* **2013**, *62* (4), 341–354.

(126) Loirand, G.; Guérin, P.; Pacaud, P. Rho Kinases in Cardiovascular Physiology and Pathophysiology. *Circ. Res.* **2006**, *98* (3), 322–334.

(127) Satoh, K.; Fukumoto, Y.; Shimokawa, H. Rho-Kinase: Important New Therapeutic Target in Cardiovascular Diseases. *Am. J. Physiol. - Heart Circ. Physiol.* **2011**, *301* (2), H287–H296.

(128) Surma, M.; Wei, L.; Shi, J. Rho Kinase as a Therapeutic Target in Cardiovascular Disease. *Future Cardiol.* **2011**, *7* (5), 657–671.

(129) Bourguignon, L. Y.; Zhu, H.; Shao, L.; Zhu, D.; Chen, Y. W. Rho-Kinase (ROK) Promotes CD44v(3,8-10)-Ankyrin Interaction and Tumor Cell Migration in Metastatic Breast Cancer Cells. *Cell Motil. Cytoskeleton* **1999**, *43* (4), 269–287.

(130) Somlyo, A. V.; Bradshaw, D.; Ramos, S.; Murphy, C.; Myers, C. E.; Somlyo, A. P. Rho-Kinase Inhibitor Retards Migration and in Vivo Dissemination of Human Prostate Cancer Cells. *Biochem. Biophys. Res. Commun.* **2000**, *269* (3), 652–659.

(131) Liao, J. K.; Seto, M.; Noma, K. Rho Kinase (ROCK) Inhibitors. *J. Cardiovasc. Pharmacol.* **2007**, *50* (1), 17–24.

(132) Li, R.; Martin, M. P.; Liu, Y.; Wang, B.; Patel, R. A.; Zhu, J.-Y.; Sun, N.; Pireddu, R.; Lawrence, N. J.; Li, J.; Haura, E. B.; Sung, S.-S.; Guida, W. C.; Schonbrunn, E.; Sebti,

S. M. Fragment-Based and Structure-Guided Discovery and Optimization of Rho Kinase Inhibitors. *J. Med. Chem.* **2012**, *55* (5), 2474–2478.

(133) Patel, R. A.; Forinash, K. D.; Pireddu, R.; Sun, Y.; Sun, N.; Martin, M. P.; Schönbrunn, E.; Lawrence, N. J.; Sebti, S. M. RKI-1447 Is a Potent Inhibitor of the Rho-Associated ROCK Kinases with Anti-Invasive and Antitumor Activities in Breast Cancer. *Cancer Res.* **2012**, *72* (19), 5025–5034.

(134) Akama, T.; Dong, C.; Virtucio, C.; Sullivan, D.; Zhou, Y.; Zhang, Y.-K.; Rock, F.; Freund, Y.; Liu, L.; Bu, W.; Wu, A.; Fan, X.-Q.; Jarnagin, K. Linking Phenotype to Kinase: Identification of a Novel Benzoxaborole Hinge-Binding Motif for Kinase Inhibition and Development of High-Potency Rho Kinase Inhibitors. *J. Pharmacol. Exp. Ther.* **2013**, *347* (3), 615–625.

(135) Sebti, S. M.; Hamilton, A. D. Rock Inhibitors and Uses Thereof. WO/2008/079945, July 4, 2008.

(136) Breitenlechner, C.; Gaßel, M.; Hidaka, H.; Kinzel, V.; Huber, R.; Engh, R. A.; Bossemeyer, D. Protein Kinase A in Complex with Rho-Kinase Inhibitors Y-27632, Fasudil, and H-1152P: Structural Basis of Selectivity. *Structure* **2003**, *11* (12), 1595–1607.

(137) Bonn, S.; Herrero, S.; Breitenlechner, C. B.; Erlbruch, A.; Lehmann, W.; Engh, R. A.; Gassel, M.; Bossemeyer, D. Structural Analysis of Protein Kinase A Mutants with Rho-Kinase Inhibitor Specificity. *J. Biol. Chem.* **2006**, *281* (34), 24818–24830.

(138) Prade, L.; Engh, R. A.; Girod, A.; Kinzel, V.; Huber, R.; Bossemeyer, D. Staurosporine-Induced Conformational Changes of cAMP-Dependent Protein Kinase Catalytic Subunit Explain Inhibitory Potential. *Structure* **1997**, *5* (12), 1627–1637.

(139) Cavasotto, C. N.; Abagyan, R. A. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol.* **2004**, *337* (1), 209–225.

(140) Shriner, R. L.; Furrow, C. L. Diacetyl-D-Tartaric Anhydride. *Org. Synth.* **1955**, *35*, 49.

(141) Dobashi, Y.; Hara, S. A Chiral Stationary Phase Derived from (R,R)-Tartramide with Broadened Scope of Application to the Liquid Chromatographic Resolution of Enantiomers. *J. Org. Chem.* **1987**, *52* (12), 2490–2496.

(142) Bell, K. Selective Aminolysis of Benzoates and Acetates of α-Hydroxy Acids and Phenols With Benzylamine and Butan-1-Amine. *Aust. J. Chem.* **1987**, *40* (10), 1723–1735.

(143) Villa Gonzalez, S.; Carlsen, P. Tartaric Acid Amides by the Gabriel Route. *Eur. J. Org. Chem.* **2007**, *2007* (21), 3495–3502.

(144) Nakamura, K.; Hara, S.; Dobashi, Y. Chiral Polysiloxanes Derived from (R,R)-Tartramide for the Gas Chromatographic Separation of Enantiomers. *Anal. Chem.* **1989**, *61* (18), 2121–2124.

(145) Basavaiah, D.; Rama Krishna, P. Synthesis of Chiral α-Aryl-α-Hydroxyacetic Acids: Substituent Effects in Pig Liver Acetone Powder (PLAP) Induced Enantioselective Hydrolysis. *Tetrahedron* **1995**, *51* (8), 2403–2416.

(146) Paquette, L. A.; Ra, C. S. Cleavage of Carbon-Carbon Bonds with High Stereochemical Control. 5. Course of the Haller-Bauer Reaction of Cyclic .alpha.-Phenyl Ketones. *J. Org. Chem.* **1988**, *53* (21), 4978–4985.

(147) El-Nezhawy, A. O. H.; El-Diwani, H. I.; Schmidt, R. R. O-(2-Oxopyrrolidin-5-Yl)trichloroacetimidates as Amidoalkylating Agents − Synthesis of (+)-Lentiginosine. *Eur. J. Org. Chem.* **2002**, *2002* (24), 4137–4142.

(148) Chen, M.-J.; Tsai, Y.-M. Intramolecular α-Acylamino Radical Cyclizations with Acylsilanes in the Preparation of Polyhydroxylated Alkaloids: (+)-Lentiginosine, (+)-1,8a-Di-Epi-Lentiginosine, and (+)-1,2-Di-Epi-Swainsonine. *Tetrahedron Lett.* **2007**, *48* (36), 6271–6274.

(149) Almeida, J. F.; Grande, M.; Moran, J. R.; Anaya, J.; Mussons, M. L.; Caballero, M. C. Synthesis of the 3-Hydroxy Oxiracetam Enantiomers, Potential Nootropic Drugs. *Tetrahedron Asymmetry* **1993**, *4* (12), 2483–2494.

(150) Oba, M.; Koguchi, S.; Nishiyama, K. A Concise Approach to Homochiral 3,4-Dihydroxyglutamic Acids. *Tetrahedron* **2002**, *58* (46), 9359–9363.

(151) Arakawa, Y.; Yoshifuji, S. Synthesis of (3S, 4S)-3,4-Dihydroxyprolines from L-Tartaric Acid. *Chem. Pharm. Bull. (Tokyo)* **1991**, *39* (9), 2219–2224.

(152) Hwang, D. J.; Kim, S. N.; Choi, J. H.; Lee, Y. S. Dicaffeoyl- or Digalloyl Pyrrolidine and Furan Derivatives as HIV Integrase Inhibitors. *Bioorg. Med. Chem.* **2001**, *9* (6), 1429–1437.

(153) Ruśkowski, P.; Synoradzki, L.; W\lostowski, M. Tartaric Acid and Its O-Acyl Derivatives. Part 8. Direct Synthesis of Novel N-Substituted Mono-and Diacyltartrimides: Unusual Reaction Course. *Arkivoc* **2011**, *9*, 142–154.

(154) Yoda, H.; Shirakawa, K.; Takabe, K. Chiral Cyclic Imides with c2-Symmetry. Novel Reagents for the Synthesis of Optically Pure Lactones Containing Three Contiguous Tertiary Centers. *Tetrahedron Lett.* **1991**, *32* (28), 3401–3404.

(155) Klaver, W. J.; Hiemstra Henk; Speckamp, W. N. Synthesis and Absolute Configuration of the Aristotelia Alkaloid Peduncularine. *J. Am. Chem. Soc.* **1989**, *111* (7), 2588–2595.

(156) Chamberlin, A. R.; Chung, J. Y. L. Synthesis of Optically Active Pyrrolizidinediols: (+)-Heliotridine. *J. Am. Chem. Soc.* **1983**, *105* (11), 3653–3656.

(157) LaPlanche, L. A.; Rogers, M. T. Cis and Trans Configurations of the Peptide Bond in N-Monosubstituted Amides by Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* **1964**, *86* (3), 337–341.

(158) Ramachandran, G. N.; Mitra, A. K. An Explanation for the Rare Occurrence of Cis Peptide Units in Proteins and Polypeptides. *J. Mol. Biol.* **1976**, *107* (1), 85–92.

(159) Ramachandran, G. N.; Sasisekharan, V. Conformation of Polypeptides and Proteins. *Adv. Protein Chem.* **1968**, *23*, 283–438.

(160) Stewart, D. E.; Sarkar, A.; Wampler, J. E. Occurrence and Role of Cis Peptide Bonds in Protein Structures. *J. Mol. Biol.* **1990**, *214* (1), 253–260.

(161) Weiss, M. S.; Jabs, A.; Hilgenfeld, R. Peptide Bonds Revisited. *Nat. Struct. Mol. Biol.* **1998**, *5* (8), 676–676.

(162) Pal, D.; Chakrabarti, P. Cis Peptide Bonds in Proteins: Residues Involved, Their Conformations, Interactions and Locations. *J. Mol. Biol.* **1999**, *294* (1), 271–288.

(163) Jabs, A.; Weiss, M. S.; Hilgenfeld, R. Non-Proline Cis Peptide Bonds in Proteins. *J. Mol. Biol.* **1999**, *286* (1), 291–304.

(164) Hamelberg, D.; McCammon, J. A. Fast Peptidyl Cis−trans Isomerization within the Flexible Gly-Rich Flaps of HIV-1 Protease. *J. Am. Chem. Soc.* **2005**, *127* (40), 13778–13779.

(165) Tchaicheeyan, O. Is Peptide Bond Cis/Trans Isomerization a Key Stage in the Chemo-Mechanical Cycle of Motor Proteins? *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **2004**, *18* (7), 783–789.

(166) Odefey, C.; Mayr, L. M.; Schmid, F. X. Non-Prolyl Cis-Trans Peptide Bond Isomerization as a Rate-Determining Step in Protein Unfolding and Refolding. *J. Mol. Biol.* **1995**, *245* (1), 69–78.

(167) Vanhove, M.; Raquet, X.; Palzkill, T.; Pain, R. H.; Frère, J.-M. The Rate-Limiting Step in the Folding of the Cis-Pro167Thr Mutant of TEM-1 β-Lactamase Is the Trans to Cis Isomerization of a Non-Proline Peptide Bond. *Proteins Struct. Funct. Bioinforma.* **1996**, *25* (1), 104–111.

(168) Wheeler, K. A.; Hawkins, A. R.; Pain, R.; Virden, R. The Slow Step of Folding of Staphylococcus Aureus PC1 β-Lactamase Involves the Collapse of a Surface Loop Rate

Limited by the Trans to Cis Isomerization of a Non-Proline Peptide Bond. *Proteins Struct. Funct. Bioinforma.* **1998**, *33* (4), 550–557.

(169) Agarwal, P. K. Cis/Trans Isomerization in HIV-1 Capsid Protein Catalyzed by Cyclophilin A: Insights from Computational and Theoretical Studies. *Proteins Struct. Funct. Bioinforma.* **2004**, *56* (3), 449–463.

(170) Ledvina, A. R.; Chung, T. W.; Hui, R.; Coon, J. J.; Tureček, F. Cascade Dissociations of Peptide Cation-Radicals. Part 2. Infrared Multiphoton Dissociation and Mechanistic Studies of Z-Ions from Pentapeptides. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (8), 1351–1363.

(171) Baldoni, H. A.; Zamarbide, G. N.; Enriz, R. D.; Jauregui, E. A.; Farkas, Ö.; Perczel, A.; Salpietro, S. J.; Csizmadia, I. G. Peptide Models XXIX. Cis–trans Isomerism of Peptide Bonds: Ab Initio Study on Small Peptide Model Compound; the 3D-Ramachandran Map of Formylglycinamide. *J. Mol. Struct. THEOCHEM* **2000**, *500* (1–3), 97–111.

(172) Moure, A.; Sanclimens, G.; Bujons, J.; Masip, I.; Alvarez-Larena, A.; Pérez-Payá, E.; Alfonso, I.; Messeguer, A. Chemical Modulation of Peptoids: Synthesis and Conformational Studies on Partially Constrained Derivatives. *Chem. - Eur. J.* **2011**, *17* (28), 7927–7939.

(173) Laursen, J. S.; Engel-Andreasen, J.; Fristrup, P.; Harris, P.; Olsen, C. A. Cis-Trans Amide Bond Rotamers in β-Peptoids and Peptoids: Evaluation of Stereoelectronic Effects in Backbone and Side Chains. *J. Am. Chem. Soc.* **2013**, *135* (7), 2835–2844.

(174) Kurtz, J.; Berger, A.; Katchalski, E. Mutarotation of Poly-L-Proline. *Nature* **1956**, *178* (4541), 1066.

(175) Gutowsky, H. S.; Holm, C. H. Rate Processes and Nuclear Magnetic Resonance Spectra. II. Hindered Internal Rotation of Amides. *J. Chem. Phys.* **1956**, *25* (6), 1228–1234.

(176) Steinberg, I. Z.; Berger, A.; Katchalski, E. Reverse Mutarotation of Poly-L-Proline. *Biochim. Biophys. Acta* **1958**, *28*, 647–648.

(177) Steinberg, I. Z.; Harrington, W. F.; Berger, A.; Sela, M.; Katchalski, E. The Configurational Changes of Poly-L-Proline in Solution. *J. Am. Chem. Soc.* **1960**, *82* (20), 5263–5279.

(178) Neuman, R. C.; Jonas, V. Studies of Chemical Exchange by Nuclear Magnetic Resonance. VI. Comparison of Carbon-Nitrogen Rotational Barriers in Amides, Thioamides, and Amidinium Ions. *J. Phys. Chem.* **1971**, *75* (23), 3532–3536.

(179) Neuman, R. C.; Jonas, V. Studies of Chemical Exchange by Nuclear Magnetic Resonance. X. Inherent Carbon-Nitrogen Rotational Barriers in Amides, Thioamides, and Amidinium Ions. *J. Org. Chem.* **1974**, *39* (7), 929–931.

(180) Taha, A. N.; True, N. S. Experimental 1H NMR and Computational Studies of Internal Rotation of Solvated Formamide. *J. Phys. Chem. A* **2000**, *104* (13), 2985–2993.

(181) Kang, Y. K.; Park, H. S. Internal Rotation about the C–N Bond of Amides. *J. Mol. Struct. THEOCHEM* **2004**, *676* (1–3), 171–176.

(182) Lauvergnat, D.; Hiberty, P. C. Role of Conjugation in the Stabilities and Rotational Barriers of Formamide and Thioformamide. An Ab Initio Valence-Bond Study. *J. Am. Chem. Soc.* **1997**, *119* (40), 9478–9482.

(183) Mantz, Y. A.; Gerard, H.; Iftimie, R.; Martyna, G. J. Ab Initio and Empirical Model MD Simulation Studies of Solvent Effects on the Properties of N-Methylacetamide along a Cis−trans Isomerization Pathway. *J. Phys. Chem. B* **2006**, *110* (27), 13523–13538.

(184) Luque, F. J.; Orozco, M. Theoretical Study of N-Methylacetamide in Vacuum and Aqueous Solution: Implications for the Peptide Bond Isomerization. *J. Org. Chem.* **1993**, *58* (23), 6397–6405.

(185) Wiberg, K. B.; Rablen, P. R.; Rush, D. J.; Keith, T. A. Amides. 3. Experimental and Theoretical Studies of the Effect of the Medium on the Rotational Barriers for N,N-Dimethylformamide and N,N-Dimethylacetamide. *J. Am. Chem. Soc.* **1995**, *117* (15), 4261–4270.

(186) Villani, V.; Alagona, G.; Ghio, C. Ab Initio Studies on N-Methylacetamide. *Mol. Eng.* **1998**, *8* (2), 135–153.

(187) Mantz, Y. A.; Branduardi, D.; Bussi, G.; Parrinello, M. Ensemble of Transition State Structures for the Cis−Trans Isomerization of N-Methylacetamide. *J. Phys. Chem. B* **2009**, *113* (37), 12521–12529.

(188) Yoder, C. H.; Gardner, R. D. Multiple-Substituent Parameter Analysis of the Effects of Substituents at Nitrogen on the Barriers to Rotation in Amides. *J. Org. Chem.* **1981**, *46* (1), 64–66.

(189) Fischer, G. Chemical Aspects of Peptide Bond Isomerisation. *Chem. Soc. Rev.* **2000**, *29* (2), 119–127.

(190) Nashed, Y. E.; Mitra, A. K. Synthesis and Characterization of Novel Dipeptide Ester Prodrugs of Acyclovir. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **2003**, *59* (9), 2033–2039.

(191) Tsume, Y.; Borras Bermejo, B.; Amidon, G. L. The Dipeptide Monoester Prodrugs of Floxuridine and Gemcitabine-Feasibility of Orally Administrable Nucleoside Analogs. *Pharm. Basel Switz.* **2014**, *7* (2), 169–191.

(192) Kohan, H. G.; Kaur, K.; Jamali, F. Synthesis and Characterization of a New Peptide Prodrug of Glucosamine with Enhanced Gut Permeability. *PLOS ONE* **2015**, *10* (5), e0126786.

(193) De, A. *Application of Peptide-Based Prodrug Chemistry in Drug Development*; SpringerBriefs in Pharmaceutical Science & Drug Development; Springer New York: New York, NY, 2013.

(194) Muller, P. Glossary of Terms Used in Physical Organic Chemistry (IUPAC Recommendations 1994). *Pure Appl. Chem.* **2009**, *66* (5), 1077–1184.

(195) Hirsch, J. A. Table of Conformational Energies—1967. In *Topics in Stereochemistry*; Allinger, N. L., Eliel, E. L., Eds.; John Wiley & Sons, Inc., 1967; pp 199–222.

(196) Charton, M. Steric Effects. I. Esterification and Acid-Catalyzed Hydrolysis of Esters. *J. Am. Chem. Soc.* **1975**, *97* (6), 1552–1556.

(197) Charton, M. The Upsilon Steric Parameter — Definition and Determination. In *Steric Effects in Drug Design*; Topics in Current Chemistry; Springer Berlin Heidelberg, 1983; pp 57–91.

(198) Icli, S.; Kandil, K. A.; Thankachan, C.; Tidwell, T. T. Steric Crowding and the Reactivity of Substituted Tert-Butyl Perbenzoates. *Can. J. Chem.* **1975**, *53* (7), 979–985.

(199) Marcelin, G.; Brooks, P. R. Steric Hindrance in Potassium Atom-Oriented Molecule Reactions. Methyl Iodide and Tert-Butyl Iodide. *J. Am. Chem. Soc.* **1975**, *97* (7), 1710–1715.

(200) Bartell, L. S.; Bradford, W. F. Molecular Structures of Neopentane and Di-Tert-Butylmethane by Vapor-Phase Electron Diffraction. *J. Mol. Struct.* **1977**, *37* (1), 113–126.

(201) Xia, P.; Wang, C.; Qi, C. Theoretical Study on the Cyclization Mechanism of Dipeptides. *Chin. J. Chem.* **2013**, *31* (6), 813–818.

(202) Gani, O. A.; Thakkar, B.; Narayanan, D.; Alam, K. A.; Kyomuhendo, P.; Rothweiler, U.; Tello-Franco, V.; Engh, R. A. Assessing Protein Kinase Target Similarity: Comparing Sequence, Structure, and Cheminformatics Approaches. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2015**, *1854* (10, Part B), 1605–1616.

(203) Thakkar, B. S.; Albrigtsen, M.; Svendsen, J. S.; Andersen, J. H.; Engh, R. A. Biofocussed Chemoprospecting: An Efficient Approach for Drug Discovery. *Chem. Biol. Drug Des.* **2017**, *in press*, 1–13.

(204) Schrödinger Release 2016-1: Maestro, Schrödinger, LLC, New York, NY, 2016.

(205) Schrödinger Release 2016-1: MacroModel, Schrödinger, LLC, New York, NY, 2016.

(206) Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A. Jaguar: A High-Performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences. *Int. J. Quantum Chem.* **2013**, *113* (18), 2110–2142.

(207) Schrödinger Release 2016-1: Jaguar, Schrödinger, LLC, New York, NY, 2016.

(208) Halgren, T. A.; Lipscomb, W. N. The Synchronous-Transit Method for Determining Reaction Pathways and Locating Molecular Transition States. *Chem. Phys. Lett.* **1977**, *49* (2), 225–232.

(209) Fukui, K. Formulation of the Reaction Coordinate. *J. Phys. Chem.* **1970**, *74* (23), 4161–4163.

(210) Services | Premier Screen | International Centre for Kinase Profiling http://www.kinase-screen.mrc.ac.uk/services/premier-screen (accessed Sep 16, 2016).

(211) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. *Anal. Chem.* **1996**, *68* (5), 850–858.

(212) Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized P.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, *26* (12), 1367–1372.

Appendix

The appendix details the results from various bioactivity screenings performed on the libraries synthesized as a part of this research project.

A. Bioactivity assays

1. Kinase profiling

The Kinase profiling of B1 and D8 was carried out at International Centre for Kinase Profiling, Dundee. The detailed results are depicted in the following charts.

Kinase profiling for **B1** (100 μM)

Kinase profiling for **B1** (100 μM)

Kinase profiling for **B1** (100 μM)

Kinase profiling for **B1** (100 μM)
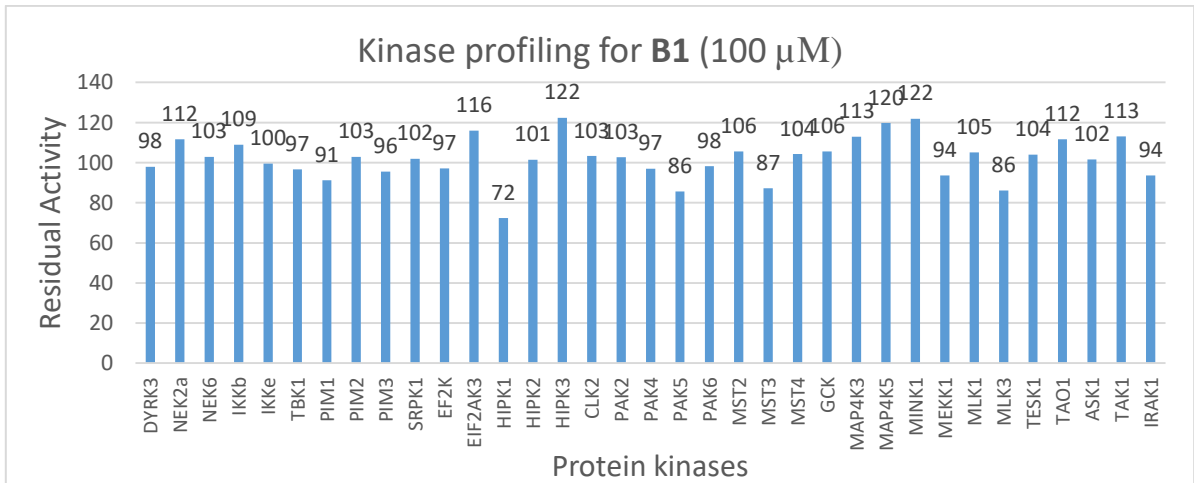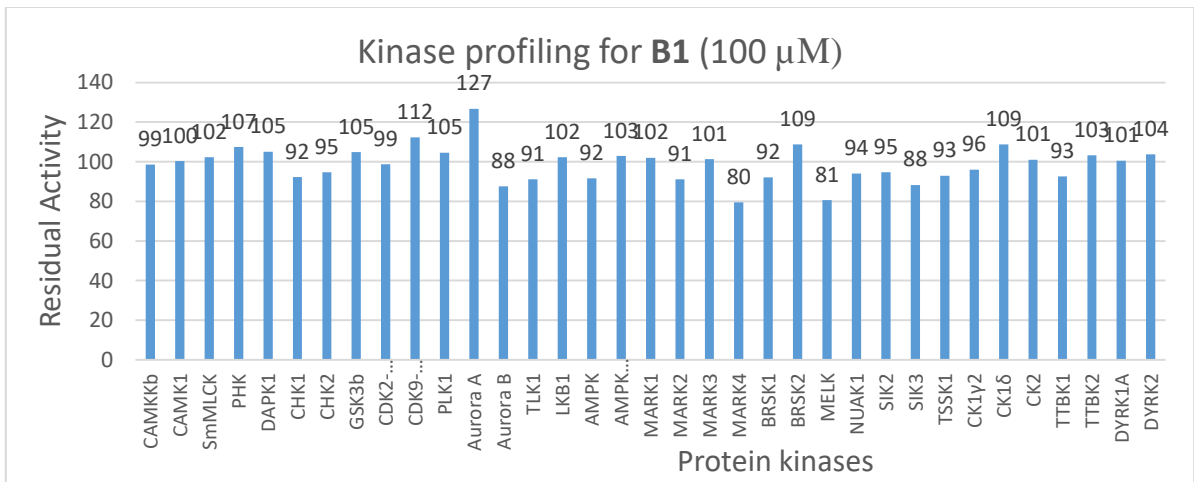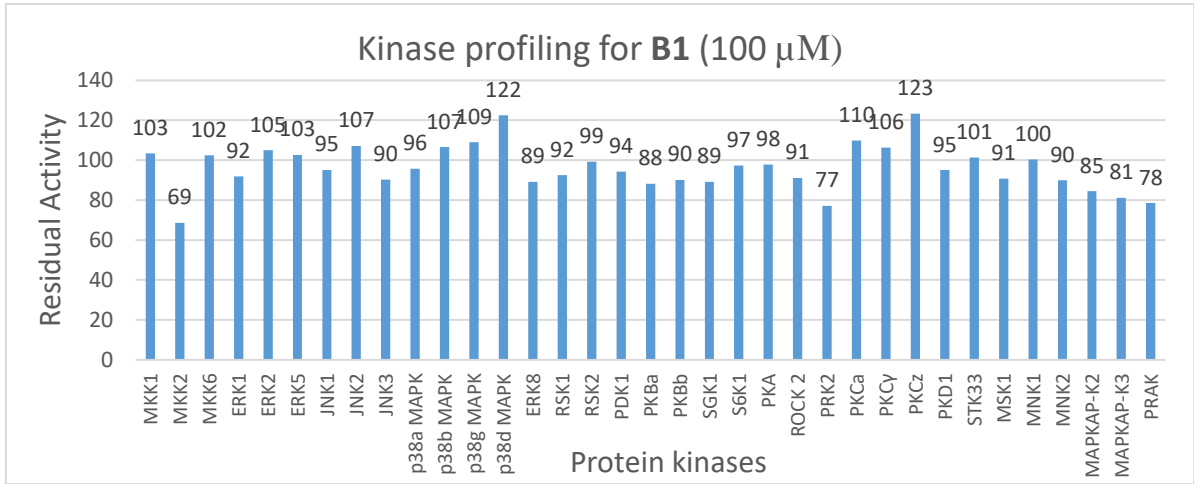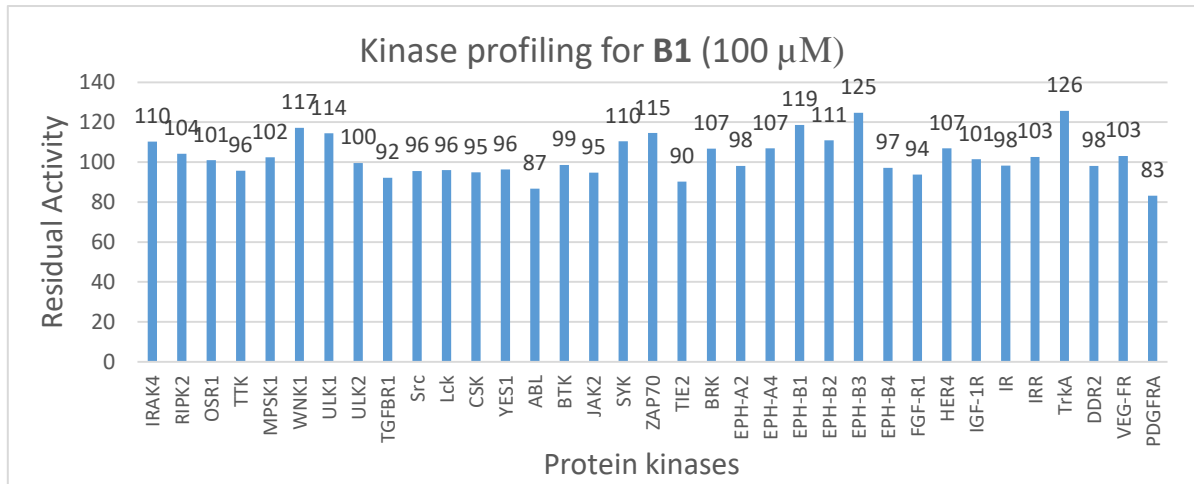
B1 did not show very significant inhibition of kinases except minor inhibition of MKK2 (31 % inhibition) and HIPK1 (28 % inhibition). On the other hand, **D8** showed interesting results. It showed inhibition of a few scattered kinases.

Kinase profiling for **D8** (100 μM)

Kinase profiling for **D8** (100 μM)

Kinase profiling for **D8** (100 µM)

Kinase profiling for **D8** (100 µM)
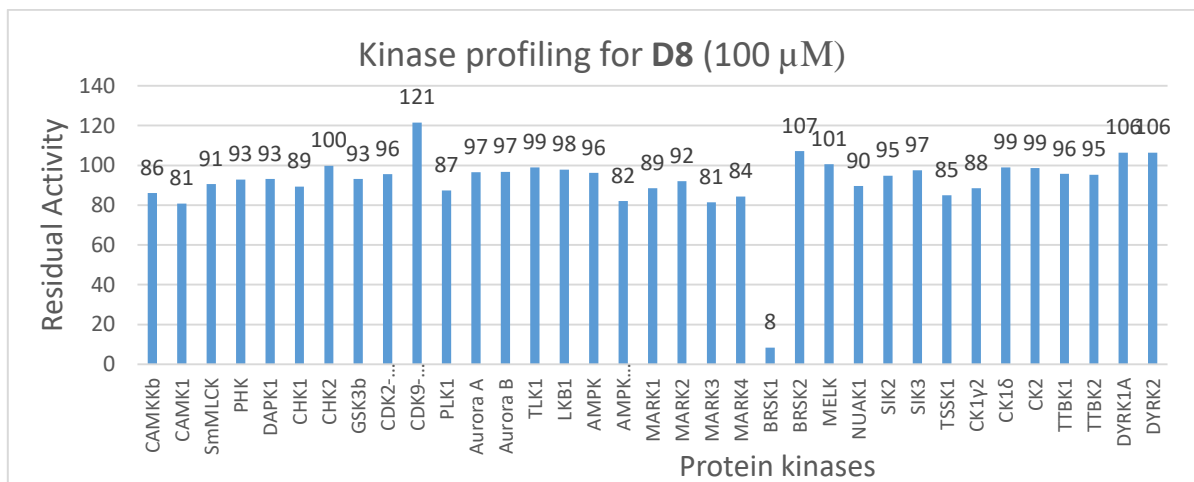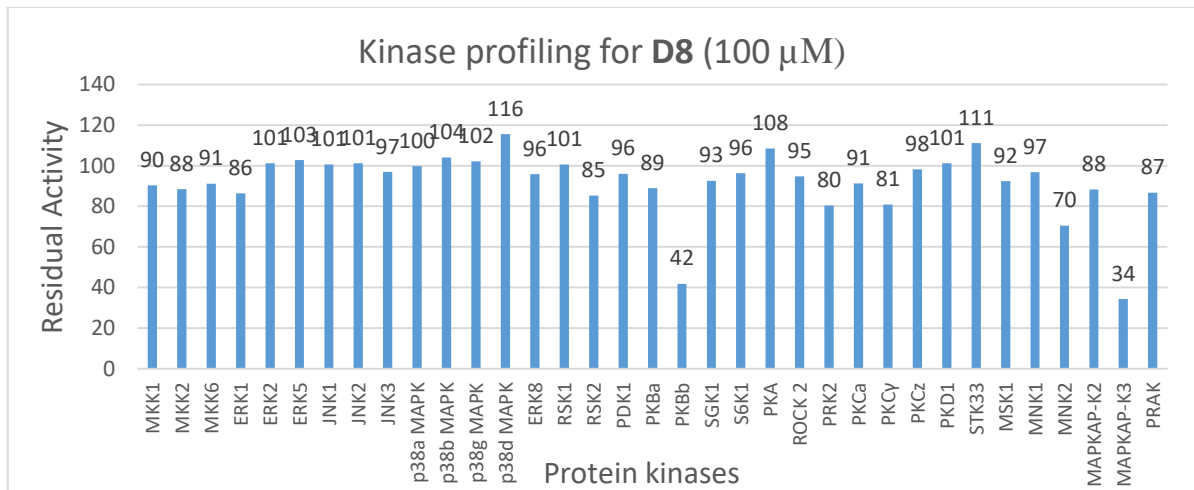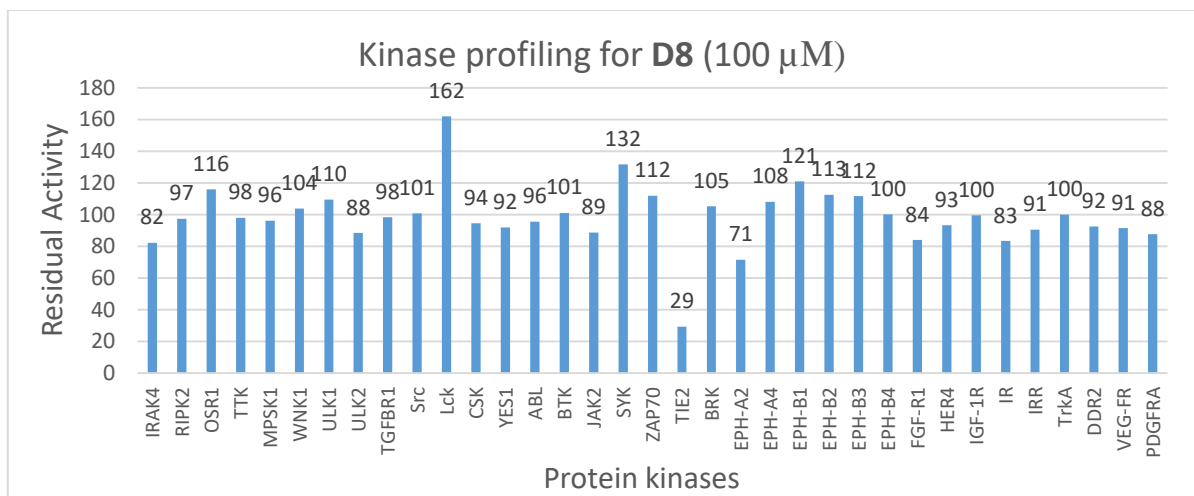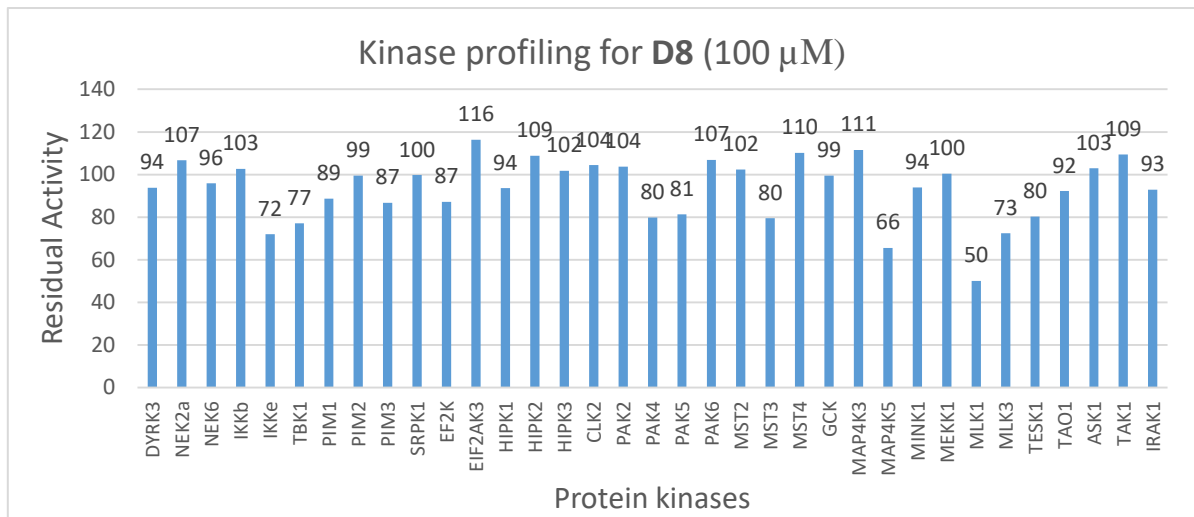
Thus, **C8** showed 92 % inhibition of brain specific kinase 1 (BRSK1, also known as SAD-1), 71 % inhibition of TIE2, 66 % inhibition of MAPKAP-K3, 58 % inhibition of PKB-b, and 50 % inhibition of MLK1. Surprisingly, it also showed 62 % activation of Lck. The inhibitions were very selective, yet scattered. The following kinome tree (Figure 45) explains how scattered the observed inhibition were.
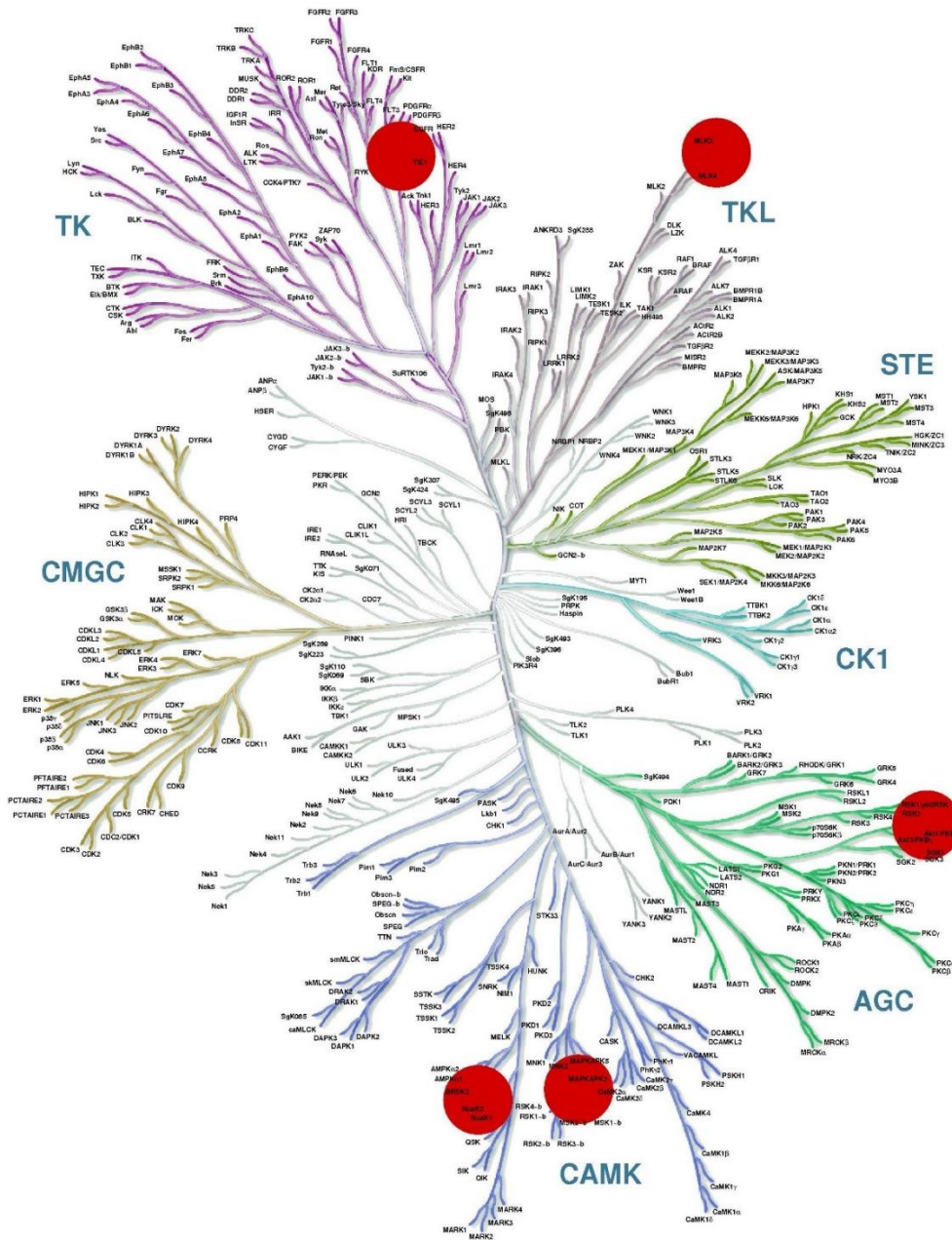
Figure 45. Kinases inhibited by D8 at 100 µM marked in kinome tree - Illustration reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com).

As the inhibition of kinases, especially BRSK1, shown by **D8** was significant, it was important to determine the IC_{50} values for the inhibition. Hence, the compound was sent again with 2 more similar compounds, **D3** and **C4**, to determine IC_{50} values. However, this time, the results were disappointing as **D8** showed only 46 % inhibition at 100 µM concentration, resulting IC_{50} greater than 100 µM, while **D3** showed only minor inhibition while **C4** did not show any significant inhibition of BRSK1 at 100 µM (Figure 46).
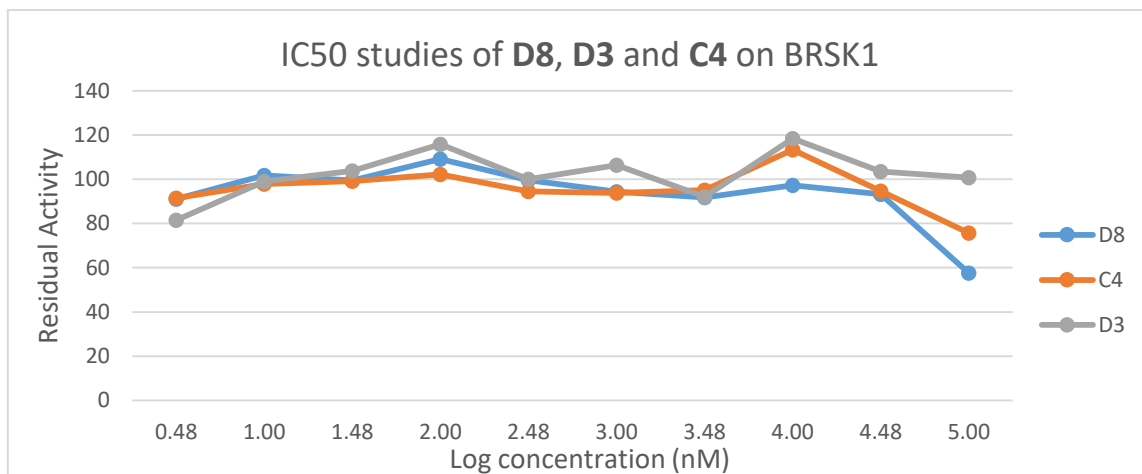
Figure 46. IC50 studies of D8, D3 and C4 on BRSK1

Despite the less reproducibility, the inhibition shown by **D8** is still interesting. As of today, BRSK1 still stands as very less explored enzyme. It requires further research to explore its importance and applications.

Along with above three compounds for IC$_{50}$ studies, three more compounds, **D15**, **D20** and **D22** were also sent for kinase profiling. One noticeable inhibition was on MAPKAP-K3, which, despite weak, was consistent for all four DKP compounds tested. However, the activation of ERK2 and p38g MAPK was surprising.

Kinase profiling for **D15** (100 µM)

Kinase profiling for **D15** (100 µM)

Kinase profiling for **D15** (100 µM)

Kinase profiling for **D20** (100 µM)

Kinase profiling for **D20** (100 µM)

Kinase profiling for **D20** (100 µM)

Kinase profiling for **D20** (100 μM)

Kinase profiling for **D22** (100 μM)

Kinase profiling for **D22** (100 μM)

Kinase profiling for **D22** (100 µM)

Kinase profiling for **D22** (100 µM)
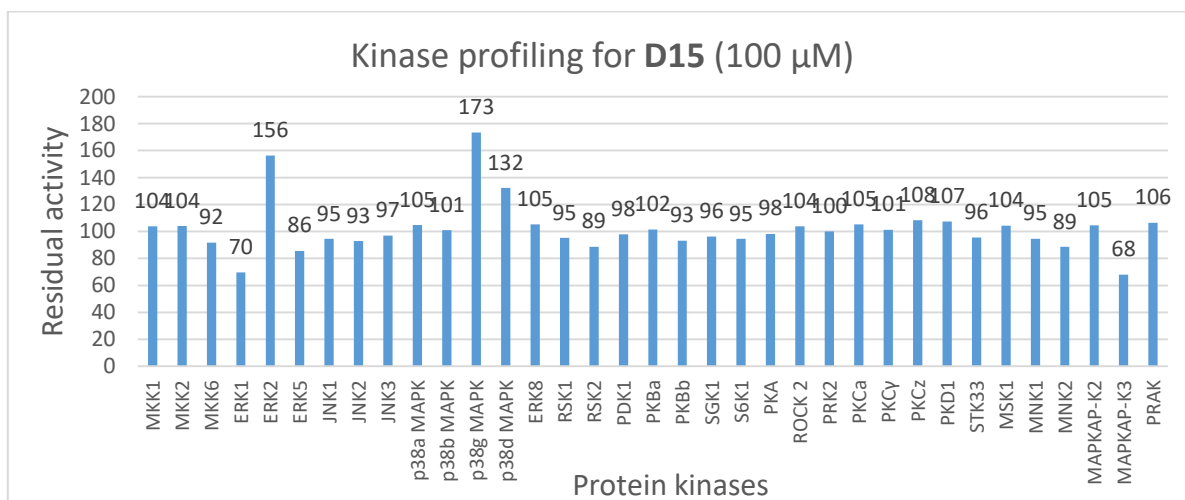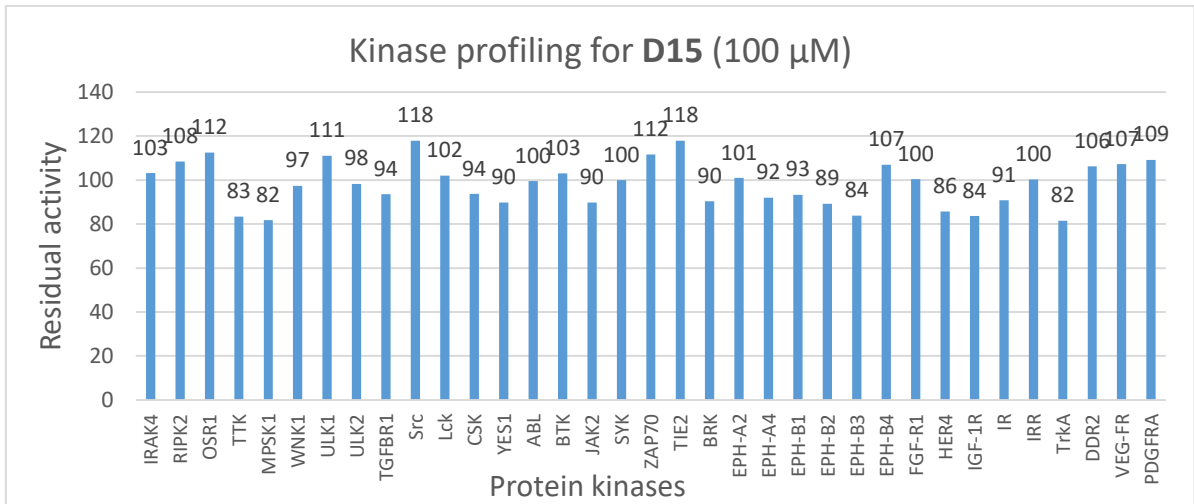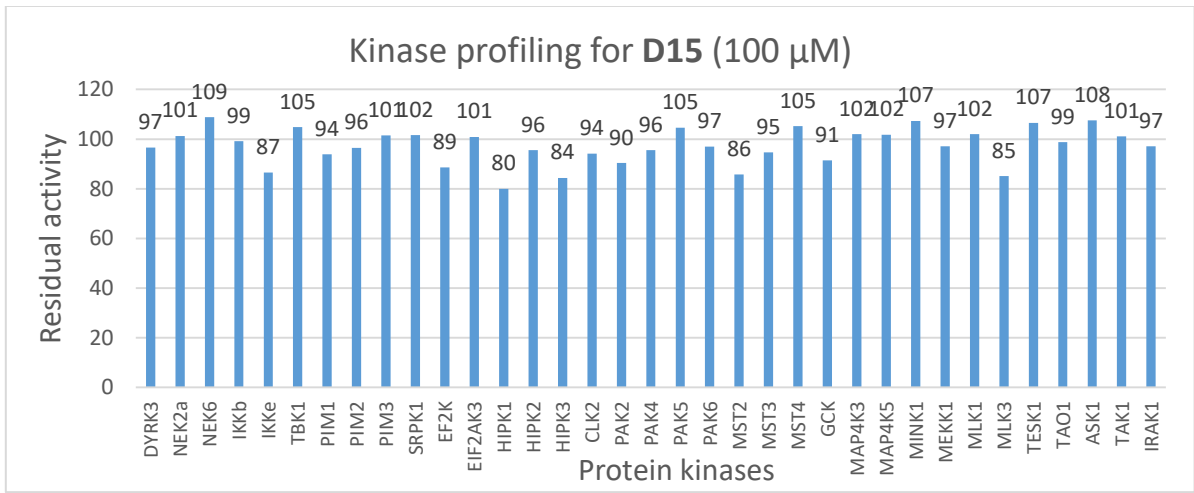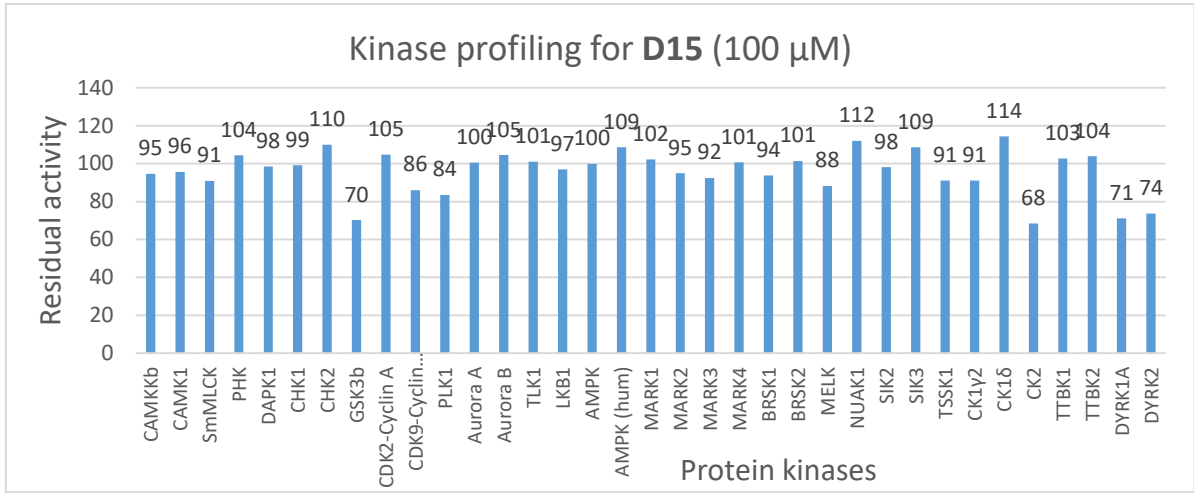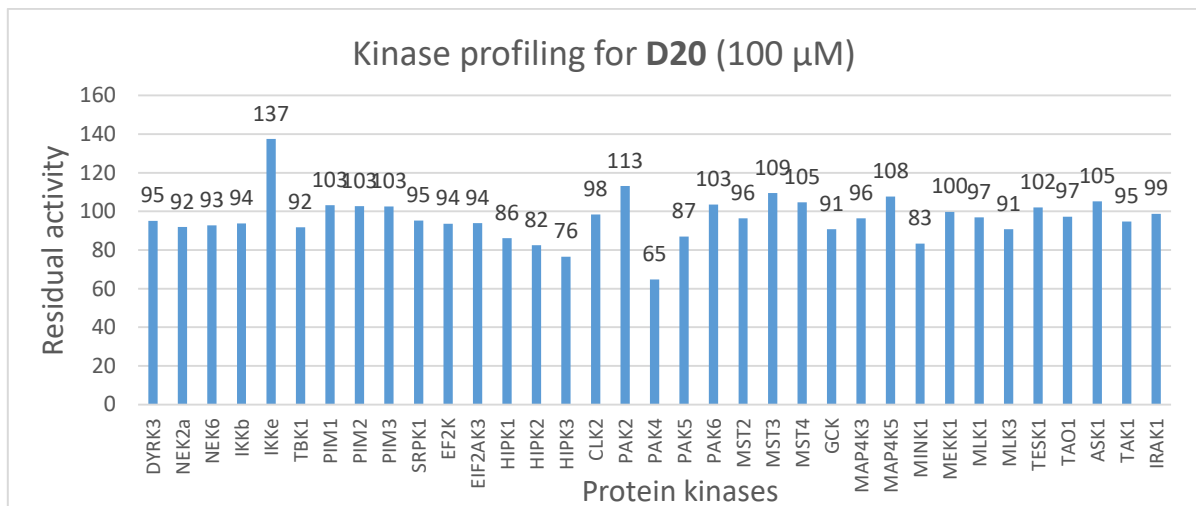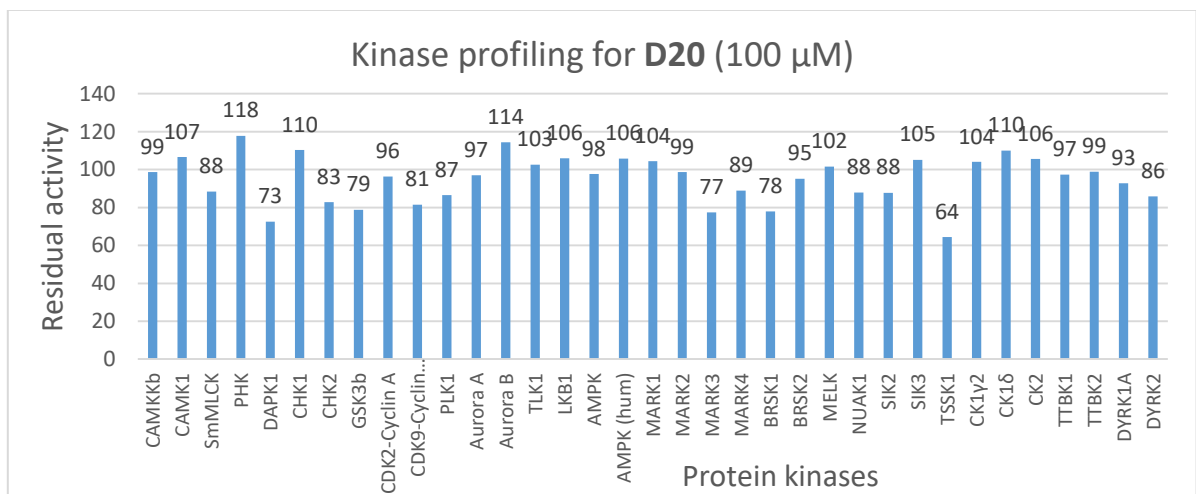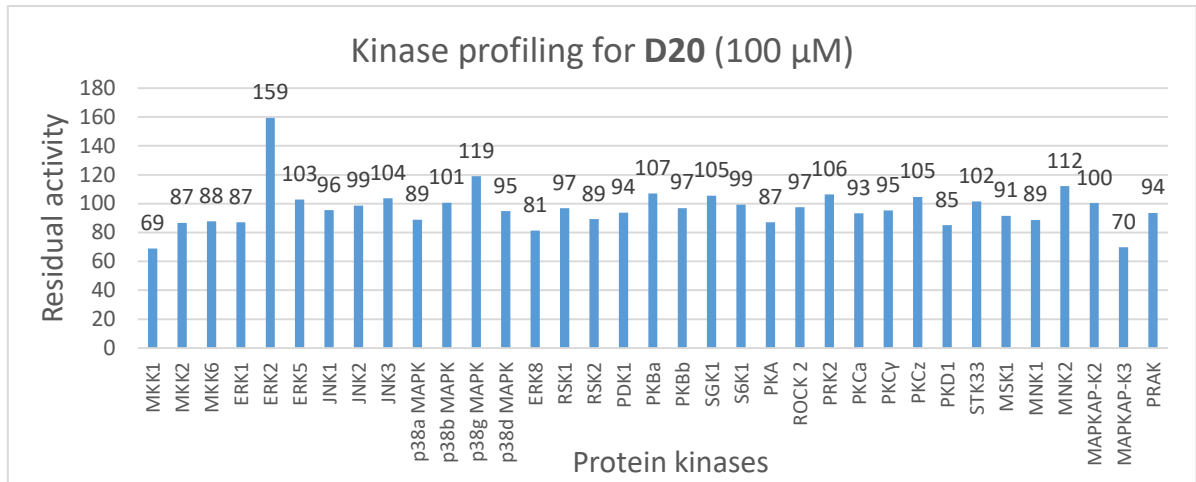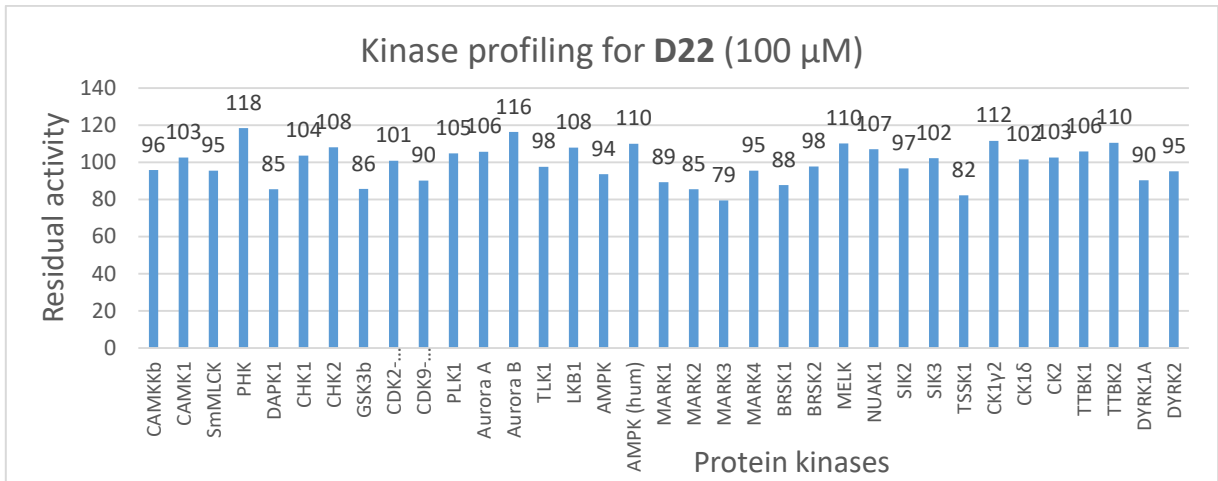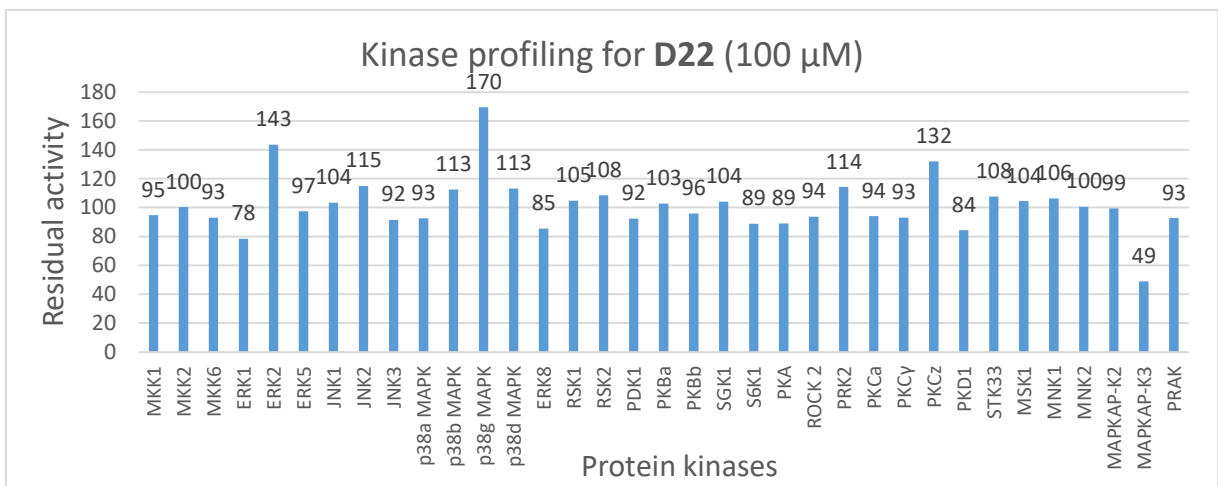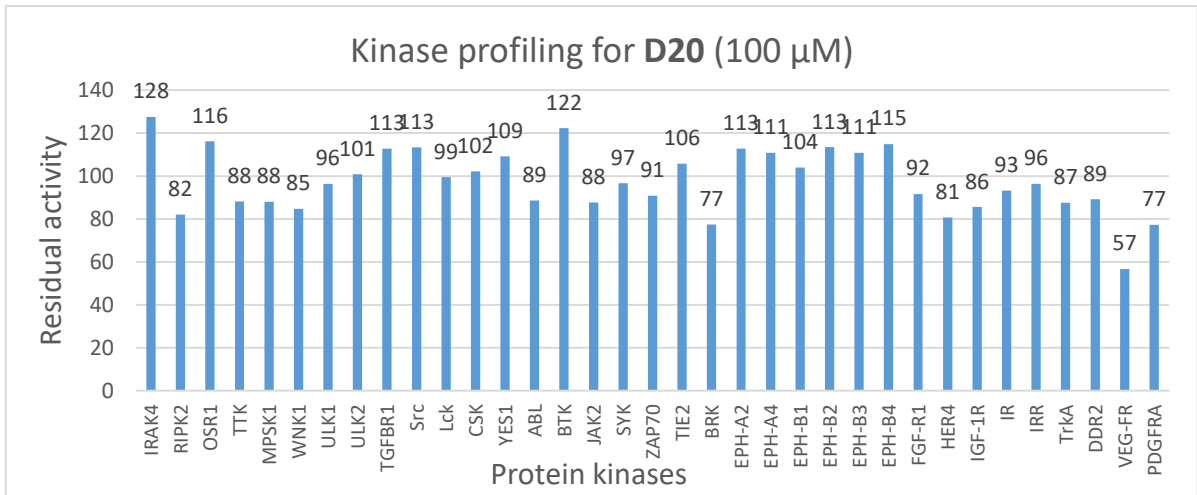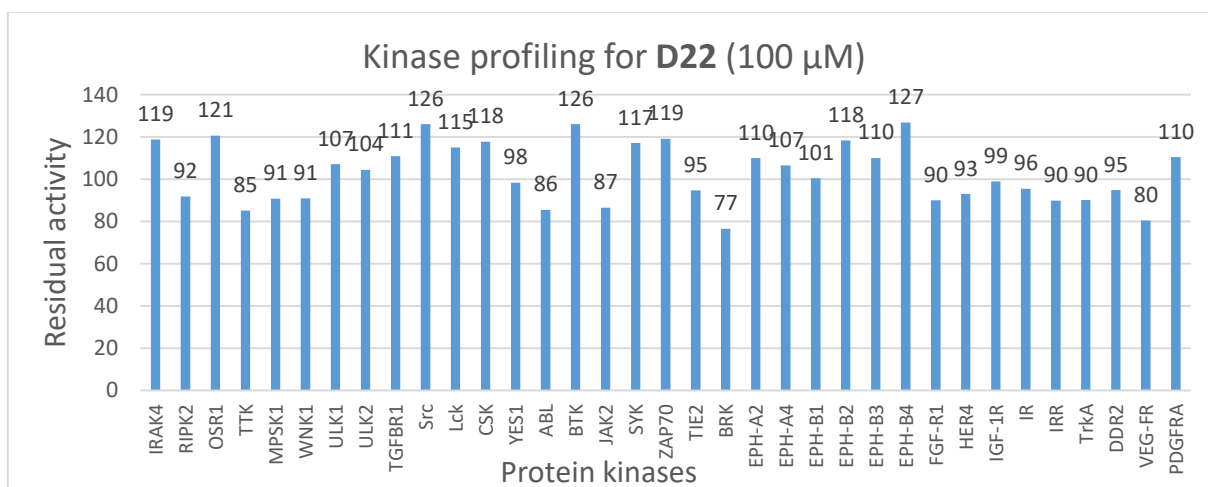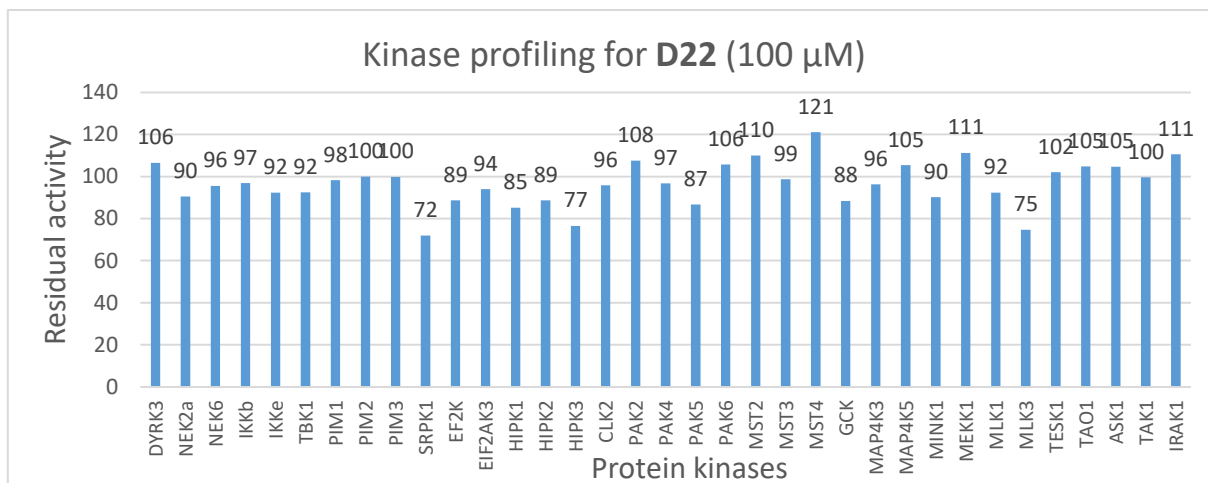
Overall, the kinase profiling indicated a possibility of hit for specific kinases, which can be developed with further research and testing.

2. Assays on cancer cell lines

a. Assays on melanoma A2058 cell line

A2058 melanoma cell line is derived from human skin. It is a common cell line used for anti-cancer activity assays. The compounds belonging to all three series were tested on this cell line at MabCent.

Assay on A2058 melanoma cell-line (50 μM) — *N*-substituted dipeptide ester compounds

Assay on A2058 melanoma cell-line (50 μM) — DKP compounds including cyclic dimers

Assay on A2058 melanoma cell-line (50 μM) — Compounds from TABs series

Assay on HT29 colon cancer cell-line (50 µM)

N-substituted dipeptide ester compounds

Assay on HT29 colon cancer cell-line (50 µM)

DKP compounds including cyclic dimers

Assay on HT29 colon cancer cell-line (50 µM)

Compounds from TABs series

Assay on MRC5 lung cell-line (50 μM) — *N*-substituted dipeptide ester compounds

Assay on MRC5 lung cell-line (50 μM) — DKP compounds including cyclic dimers

Assay on MRC5 lung cell-line (50 μM) — Compounds from TABs series

Assay on MCF7 breast cancer cell-line (50 μM)

N-substituted dipeptide ester compounds

Assay on MCF7 breast cancer lung cell-line (50 μM)

DKP compounds including cyclic dimers

Assay on MCF7 breast cancer cell-line (50 μM)

Compounds from TABs series

261

A few compounds from *N*-substituted dipeptide esters (B11, B12, B18, B25 and B28) and DKP (D13 and D14) series show significant loss of viability at 50 µM. The compounds showing more than 50 % viability loss were then subjected to IC_{50} studies.

IC50 studies of B18 and D14 on A2058 melanoma cell-line
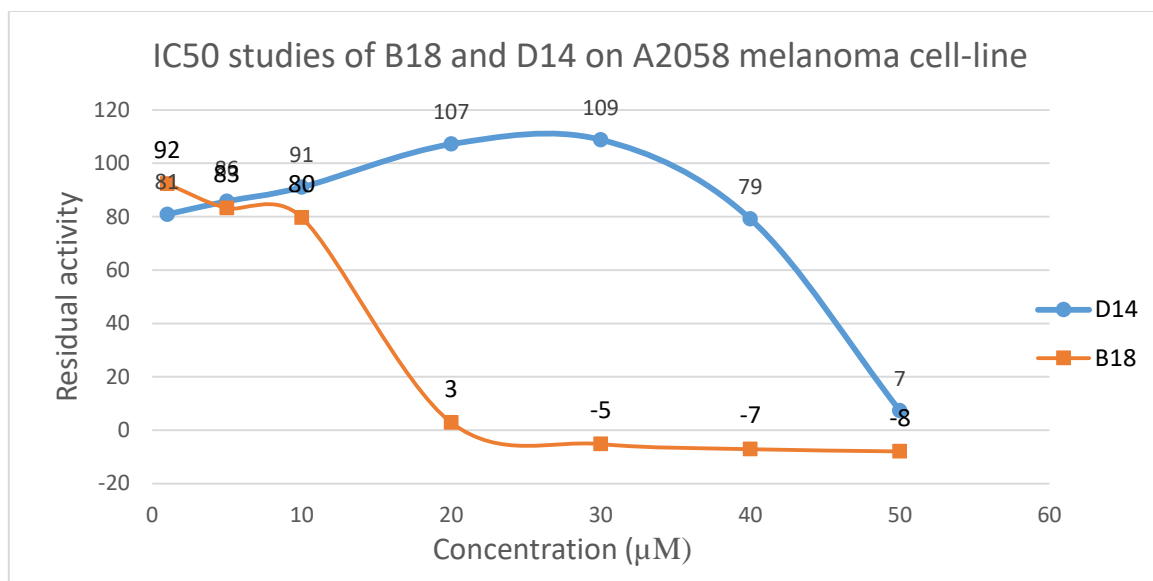
The IC_{50} value of B18 has been measured at 13.09 µM, while IC50 value of C14 has been measured at 43.72 µM for A2058 cell line. The IC_{50} value for the other compounds and cell lines could not be determined due to high variations because of precipitation of compounds.

3. Anti-fouling assay

As the DKP scaffold was developed from synoxazolidinones, barettin and ianthelline, we also expected the compounds to show anti-fouling activity. Currently, the only available assay for anti-fouling activity is the bio-film assay. In this assay, the activity of the compound is determined by whether it can inhibit the biofilm formation of *Staphylococcus epidermidis.*

The biofilm assays were carried out at MarBio. To confirm as an active compound, the optical density value should come less than or equal to 0.25. However, no compound from all three series showed any significant anti-fouling activity.

Anti-fouling activity assay (50 µM)

N-substituted dipeptide ester compounds

Anti-fouling activity assay (50 µM)

DKP compounds including cyclic dimers

Anti-fouling activity assay (50 µM)

Compounds from TABs series
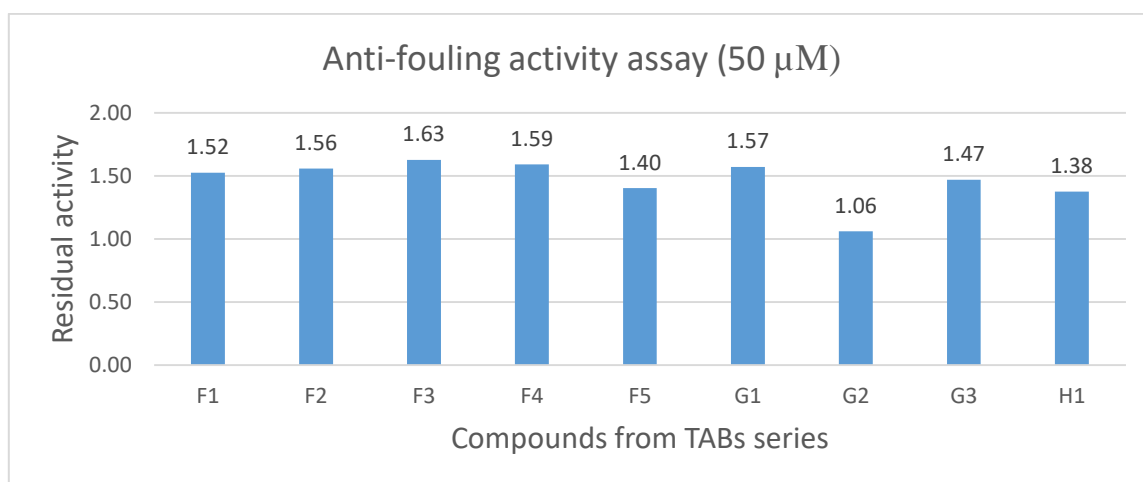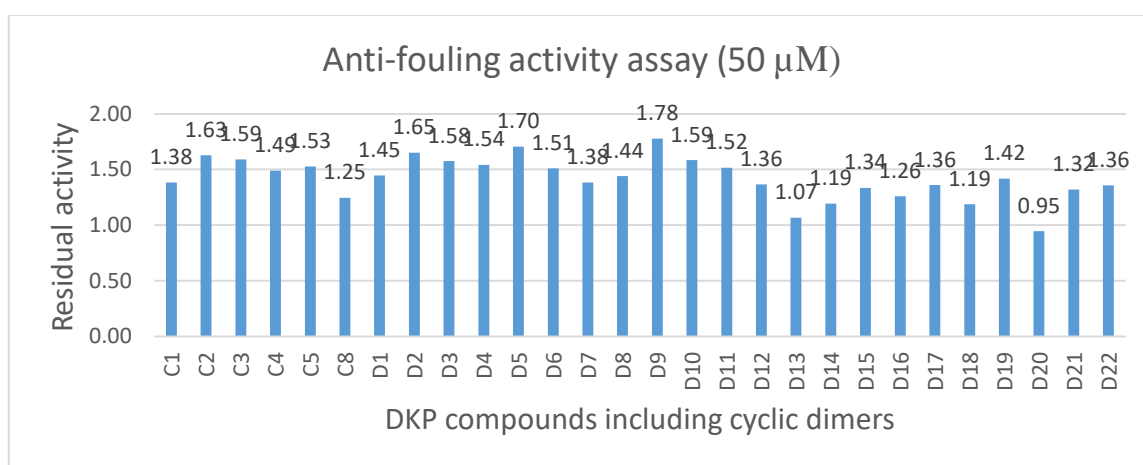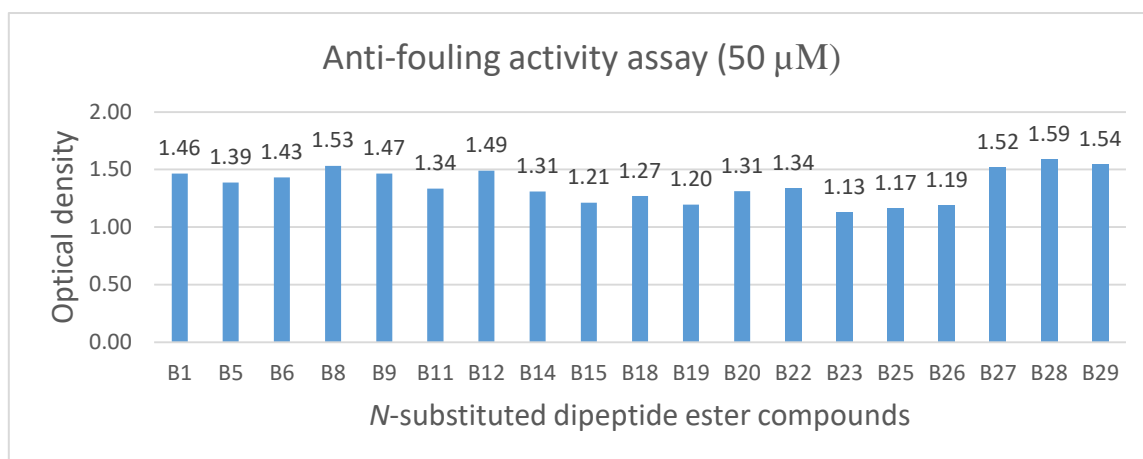
4. Cellular lipid peroxidation antioxidant assay (CLPAA)

CLPAA is used for testing of anti-oxidative effect of compounds. It measures the intracellular anti-oxidative protection against free radicals. The assays were carried out at MabCent. Compounds from all 3 series were tested. The residual activity less than 50 % is usually considered as an indication of an active compound. But no compound was found active.

263

Cellular Lipid Peroxidation Antioxidant Assay (50 µM)

Residual activity — *N*-substituted dipeptide ester compounds

Compound	Residual activity
B1	104
B5	101
B6	107
B8	81
B9	91
B11	89
B12	101
B14	125
B15	127
B18	89
B19	120
B20	83
B21	80
B22	89
B23	113
B25	129
B26	104
B27	105
B28	103
B29	98

Cellular Lipid Peroxidation Antioxidant Assay (50 µM)

Residual activity — DKP compounds

Compound	Residual activity
C1	88
C2	88
C3	93
C4	75
C5	85
C6	99
C7	99
C8	100
C9	101
C10	97
C11	83
C12	120
C13	99
C14	112
C15	120
C16	109
C17	91
C18	101
C19	92
C20	121
C21	118
C22	100

Cellular Lipid Peroxidation Antioxidant Assay (50 µM)

Residual activity — Cyclized dimer compounds

Compound	Residual activity
D1	112
D2	89
D3	93
D4	93
D5	96
D8	123

5. Anti-bacterial activity assays

The anti-bacterial assays were carried out at MabCent. Compounds from all three series were tested. The assays were done against 5 bacterial species:

1) *Enterococcus faecalis*
2) *Escherichia coli*
3) *Pseudomonas aeruginosa*

4) *Staphylococcus aureus*

5) *Streptococcus bovis*

For the antibacterial assays, cell density less than 0.05 is considered an indication of an active compound. However, no compound from all three series was found to be active in any anti-bacterial assay.

a.   Assays on *Enterococcus faecalis*

Assay on *Enterococcus faecalis* (50 μM)

Assay on *Enterococcus faecalis* (50 μM)

Assay on *Enterococcus faecalis* (50 μM)

Assay on *Escherichia coli* (50 μM)

N-substituted dipeptide ester compounds

Assay on *E Coli* (50 μM)

DKP compounds including cyclic dimers

Assay on *E Coli* (50 μM)

Compounds from TABs series
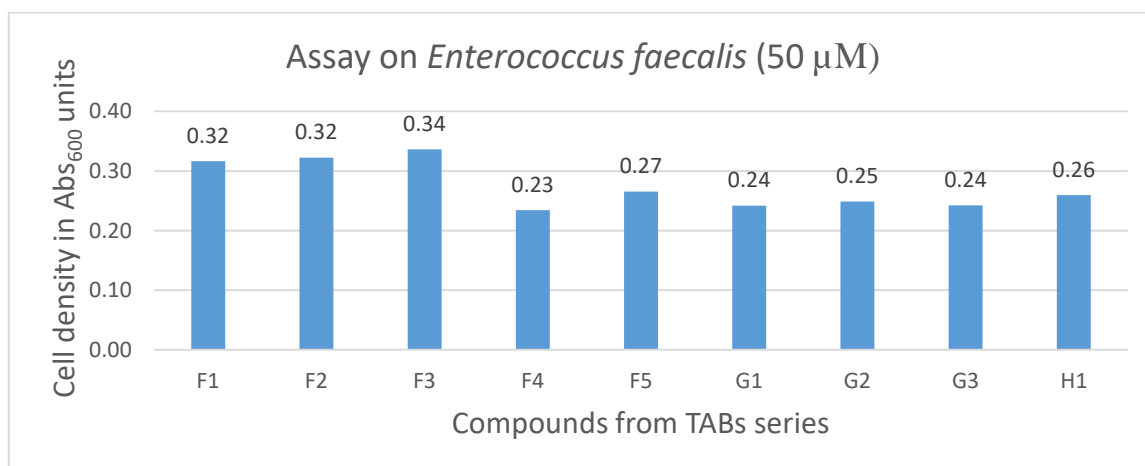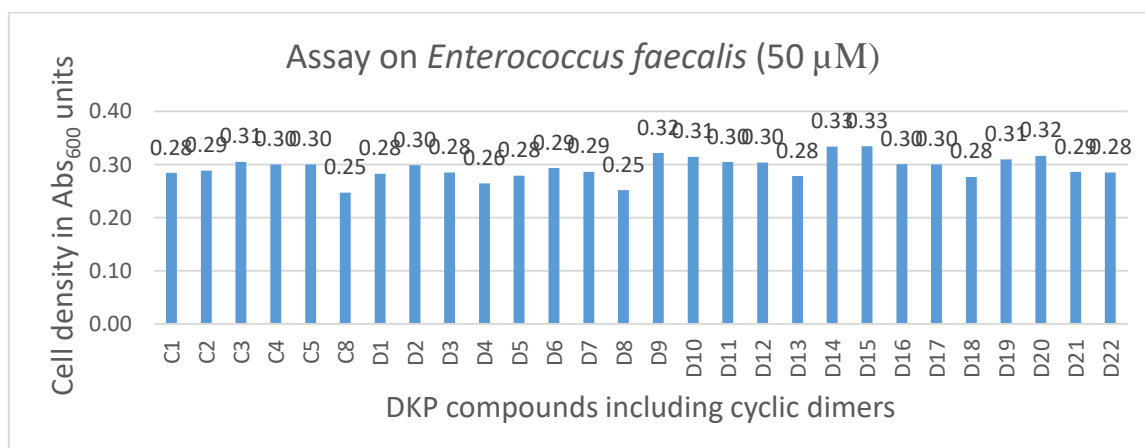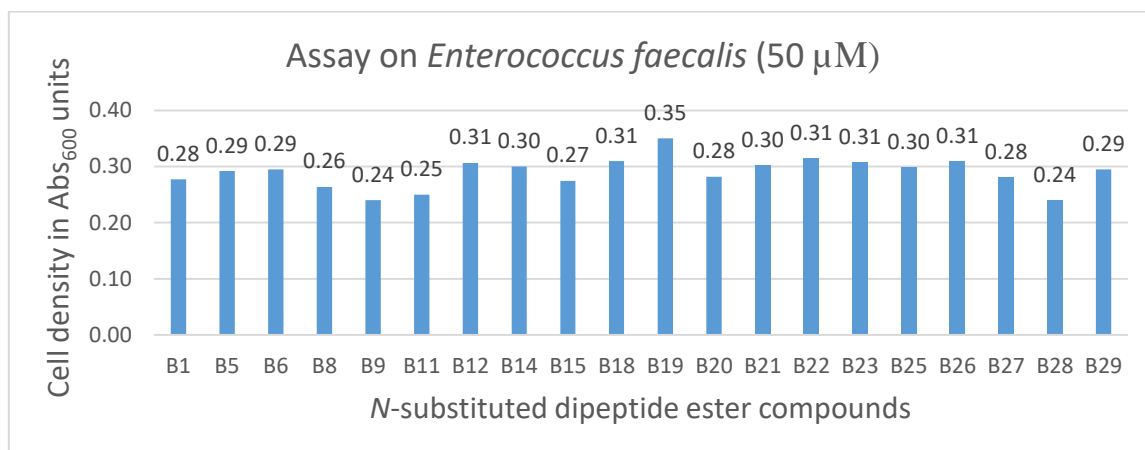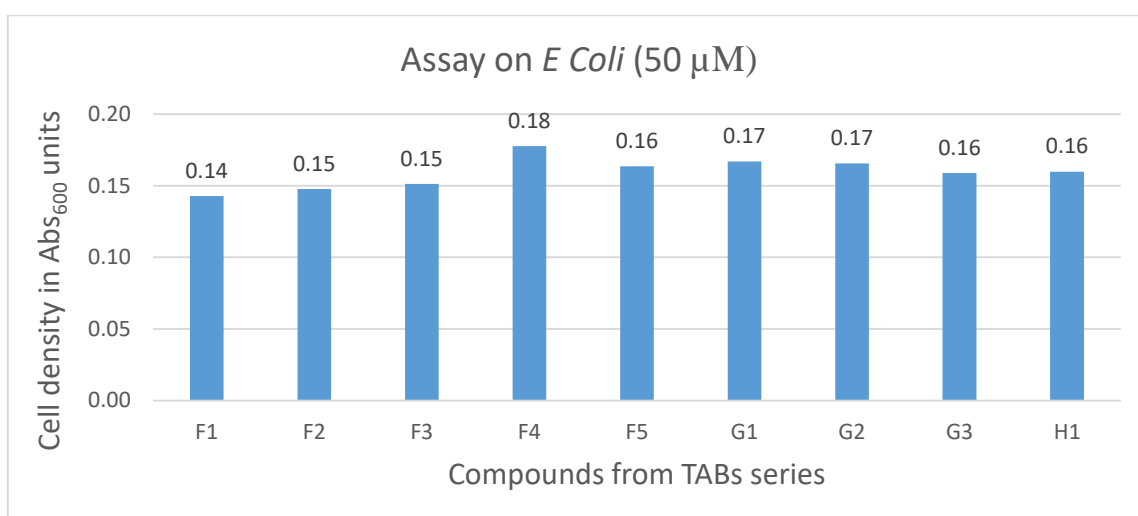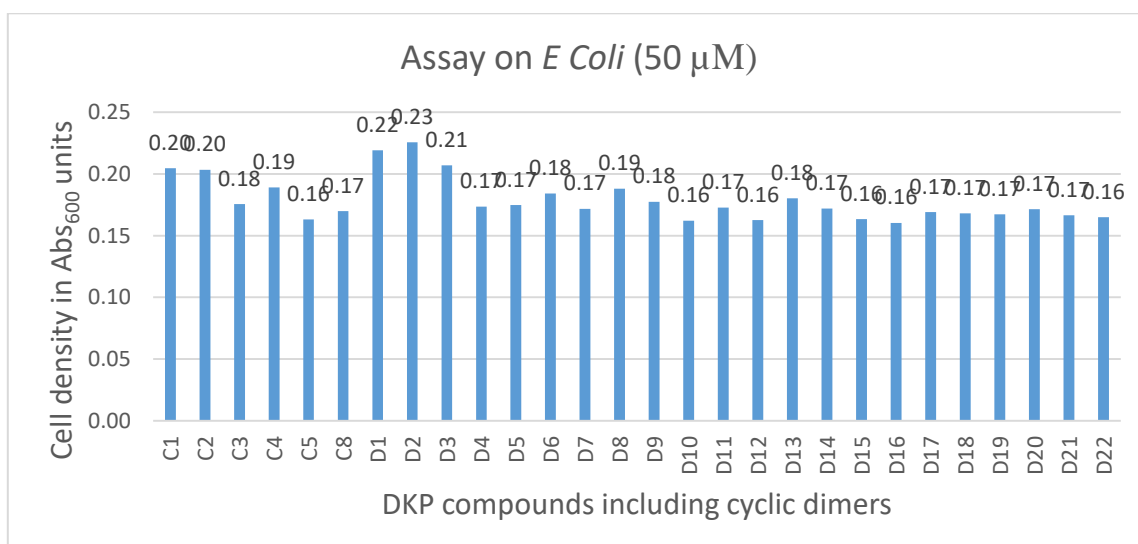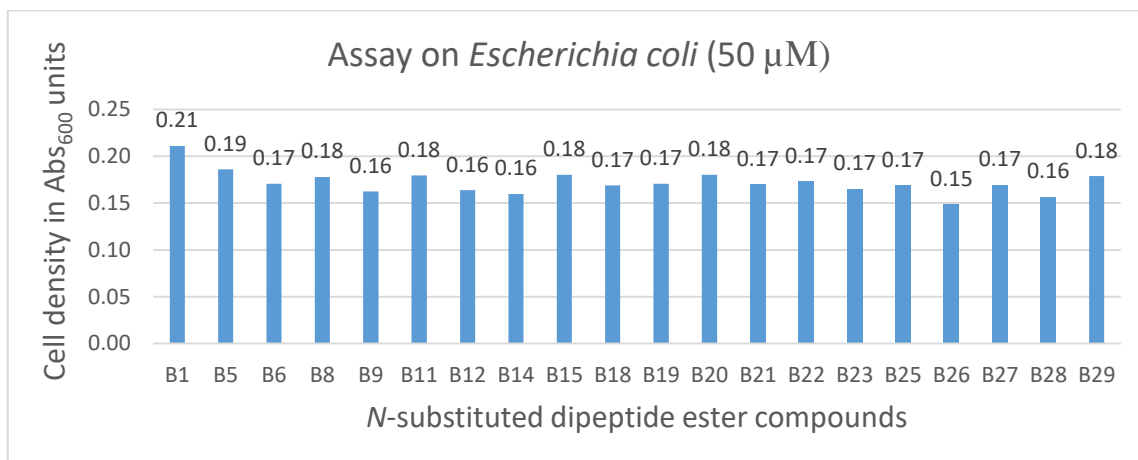
c.   Assays on Pseudomonas aeruginosa

Assay on *Pseudomonas aeruginosa* (50 μM)

Cell density in Abs$_{600}$ units

Values for *N*-substituted dipeptide ester compounds:
B1: 0.32, B5: 0.27, B6: 0.23, B8: 0.21, B9: 0.36, B11: 0.34, B12: 0.28, B14: 0.26, B15: 0.22, B18: 0.27, B19: 0.24, B20: 0.26, B21: 0.27, B22: 0.27, B23: 0.25, B25: 0.25, B26: 0.25, B27: 0.30, B28: 0.31, B29: 0.39

N-substituted dipeptide ester compounds

Assay on *Pseudomonas aeruginosa* (50 μM)

Cell density in Abs$_{600}$ units

Values for DKP compounds including cyclic dimers:
C1: 0.32, C2: 0.25, C3: 0.25, C4: 0.25, C5: 0.41, C8: 0.34, D1: 0.25, D2: 0.30, D3: 0.27, D4: 0.25, D5: 0.28, D6: 0.29, D7: 0.26, D8: 0.25, D9: 0.26, D10: 0.28, D11: 0.25, D12: 0.24, D13: 0.25, D14: 0.27, D15: 0.26, D16: 0.26, D17: 0.27, D18: 0.23, D19: 0.28, D20: 0.25, D21: 0.26, D22: 0.25

DKP compounds including cyclic dimers

Assay on *Pseudomonas aeruginosa* (50 μM)

Cell density in Abs$_{600}$ units

Values for compounds from TABs series:
F1: 0.27, F2: 0.27, F3: 0.27, F4: 0.31, F5: 0.31, G1: 0.35, G2: 0.34, G3: 0.30, H1: 0.29

Compounds from TABs series

Assay on *Staphylococcus aureus* (50 µM)

Assay on *Staphylococcus aureus* (50 µM)

Assay on *Staphylococcus aureus* (50 µM)

e.   Assays on *Streptococcus bovis*

Assay on *Streptococcus bovis* (50 µM)

N-substituted dipeptide ester compounds

Cell density in Abs$_{600}$ units

B1: 0.28, B5: 0.23, B6: 0.30, B8: 0.19, B9: 0.33, B11: 0.30, B12: 0.31, B14: 0.21, B15: 0.19, B18: 0.24, B19: 0.23, B20: 0.20, B21: 0.21, B22: 0.20, B23: 0.19, B25: 0.28, B26: 0.24, B27: 0.24, B28: 0.27, B29: 0.36

Assay on *Streptococcus bovis* (50 µM)

DKP compounds including cyclic dimers

Cell density in Abs$_{600}$ units

C1: 0.20, C2: 0.27, C3: 0.25, C4: 0.21, C5: 0.37, C8: 0.31, D1: 0.22, D2: 0.25, D3: 0.19, D4: 0.26, D5: 0.25, D6: 0.26, D7: 0.22, D8: 0.20, D9: 0.29, D10: 0.23, D11: 0.20, D12: 0.19, D13: 0.22, D14: 0.27, D15: 0.20, D16: 0.18, D17: 0.22, D18: 0.18, D19: 0.23, D20: 0.21, D21: 0.21, D22: 0.17

Assay on *Streptococcus bovis* (50 µM)

Compounds from TABs series

Cell density in Abs$_{600}$ units

F1: 0.24, F2: 0.24, F3: 0.18, F4: 0.27, F5: 0.24, G1: 0.27, G2: 0.25, G3: 0.29, H1: 0.27

Cell density less than 0.05 is required to consider a compound as active. Thus, no compound was found to be active in any anti-bacterial assay.

269

6. SILAC studies

a. Downregulated proteins

Table 36. Downregulated proteins with significant change

T: Gene names	Ratio H/L normalized 0.5 µM	Ratio H/L normalized 2.5 µM	Ratio H/L normalized 5 µM	C: Ratio H/L normalized 0.5 µM B significant	C: Ratio H/L normalized 2.5 µM B significant	C: Ratio H/L normalized 5 µM B significant
FABP7	-0.5621	-0.3640	-0.6763	+	+	+
LIMCH1	-0.5074	-0.3293	-0.1321	+	+	
PDS5A	-0.3963	-0.2584	-0.0140	+	+	
UBAP2L	-0.2782	-0.1897	-0.0523	+	+	
MYH10	-0.5008	-0.3426	0.1881	+	+	
ITGAV	-0.3043	-0.2144	0.0742	+	+	
TAGLN	-0.9489	-0.6920	-0.6865	+	+	+
MAD1L1	-0.3066	-0.2323	0.0138	+	+	
NUP155	-0.2613	-0.2049	-0.0436	+	+	
SF3B2	-0.2105	-0.1657	-0.0250	+	+	
SUPT16H	-0.2624	-0.2070	-0.0246	+	+	
NOMO1; NOMO3; NOMO2	-0.2514	-0.1992	-0.0963	+	+	
PEG10	-0.2421	-0.1980	-0.1218	+	+	
ARPC5L	-0.4846	-0.3987	-0.0824	+	+	
MSH6	-0.2677	-0.2211	0.0410	+	+	
MAP4	-0.2307	-0.1954	0.0563	+	+	
ANLN	-0.3106	-0.2659	-0.0379	+	+	
CEP170	-0.4256	-0.3747	-0.0038	+	+	
CWC22	-0.4077	-0.3743	-0.6052	+	+	+
RBX1	-0.3960	-0.3669	-0.3472	+	+	+
CBX3	-0.2604	-0.2479	-0.1395	+	+	
DIS3	-0.2088	-0.2279	-0.0177	+	+	
TIMP3	-0.5664	-0.6344	-0.3649	+	+	+
LAMB1	-0.3160	-0.3616	0.4140	+	+	
CDH2	-0.2845	-0.3415	-0.3549	+	+	+
SLC7A1	-0.2959	-0.3610	0.2004	+	+	
GSN	-0.3108	-0.4203	-0.5218	+	+	+
ATP1A1	-0.2069	-0.2857	-0.0621	+	+	
FLNA	-0.1403	-0.2795	0.5093	+	+	
HSP90AB2P	-0.2117	-0.5089	-0.1183	+	+	+
NCOR1	-0.3628	-2.2587	-0.1888	+	+	
MANF	-0.8355	-0.0998	-0.2980	+		+
MRPL43	-0.6707	NaN	-1.0579	+		+
SNCA	-0.8705	-0.2178	-0.1751	+		+
MYL6	-0.3437	0.0751	-0.1465	+		+

CALM2; CALM1; CALM3	-1.4699	-0.0341	-0.3694	+		+
CAPS	-0.9050	NaN	-1.2556	+		+
EIF1; EIF1B	-0.2897	0.1209	-0.2657	+		+
RPLP2	-0.7308	-0.0269	-0.2263	+		+
RPS10;RPS10P5	-0.2200	-0.0985	-0.1917	+		+
NUBP1	-0.6034	NaN	-0.2983	+		+
MTPN	-0.3305	0.3017	-0.2202	+		+
ACTC1;ACTA1	-0.2175	-0.0252	-0.2021	+		+
UBE2L3	-0.4292	-0.0550	-0.1536	+		+
FABP5	-0.1504	0.0243	-0.2501	+		+
HSPA14	-0.3443	NaN	-0.3578	+		+
GPNMB	-0.2913	-0.1639	-0.2252	+		+
PFDN5	-0.8685	NaN	-0.3107	+		+
COPS8	-1.2720	0.1849	-0.4648	+		+
VPS16	-0.2410	-0.1694	-0.4128	+		+
ACSL1	0.0606	-0.2169	-0.4471		+	+
CALD1	-0.1622	-0.1826	-0.1577		+	+
VKORC1	-0.0842	-0.7549	-0.6919		+	+
PUS1	-0.2180	-0.7775	-0.2791		+	+
PYGL	-0.1142	-0.2125	-0.1553		+	+
CRKL	-0.2517	-0.2740	-0.3389		+	+
SNW1	-0.0859	-0.2379	-0.2299		+	+
CIRBP	-0.0971	-1.7229	-0.5395		+	+
UAP1	-0.1603	-0.2406	-0.3402		+	+
ELMO2	NaN	-0.5779	-0.4515		+	+
DPP9	0.1734	-0.7812	-0.4670		+	+
AKAP1	NaN	-0.7588	-1.0041		+	+
STAM	NaN	-0.3678	-0.4854		+	+
CHMP4B	NaN	-0.3399	-0.4884		+	+
RAI14	-0.0497	-0.7148	-0.5126		+	+
SPCS1	NaN	-0.6461	-0.3987		+	+

NaN = The peptides corresponding to the protein could not be identified in both heavy and light media samples.

b. Upregulated proteins

Table 37. Upregulated proteins with significant change

T: Gene names	Ratio H/L normalized 0.5 μM	Ratio H/L normalized 2.5 μM	Ratio H/L normalized 5 μM	C: Ratio H/L normalized 0.5 μM B significant	C: Ratio H/L normalized 2.5 μM B significant	C: Ratio H/L normalized 5 μM B significant
RPL36	0.3639	0.1378	-0.0947	+	+	
ATP5I	0.4084	0.1932	-0.0865	+	+	
RPL39P5; RPL39	0.4193	0.2227	-0.1828	+	+	

TMEM258	0.5213	0.2889	-0.2764	+	+	
EXOSC2	0.3619	0.2228	0.1643	+	+	
DAD1	0.2844	0.1805	-0.2871	+	+	
ACBD3	0.3042	0.1980	-0.0050	+	+	
GABARAP; GABARAPL1	0.6808	0.4592	NaN	+	+	
HSPA1B; HSPA1A	0.2020	0.1375	0.1843	+	+	+
SERPINB1	0.3340	0.2434	0.1130	+	+	
RPS23	0.3043	0.2238	-0.0605	+	+	
SORBS2	0.4691	0.3463	0.0259	+	+	
DPM3	0.3275	0.2494	-0.2162	+	+	
ROMO1	0.5096	0.3936	NaN	+	+	
MMP14	0.3975	0.3080	0.3268	+	+	+
LMNB2	0.2552	0.1995	0.1511	+	+	
FAH	0.2432	0.1911	0.0592	+	+	
SEC61G	0.3337	0.2674	-0.4032	+	+	
IER3IP1	0.3613	0.2957	-0.2918	+	+	
	0.1626	0.1333	-0.1003	+	+	
ME1	0.3996	0.3318	0.3717	+	+	+
CTSD	0.2683	0.2297	0.2460	+	+	+
TMEM59	0.4868	0.4273	-0.2674	+	+	
APOBEC3C	0.2292	0.2040	0.2095	+	+	+
ANPEP	0.1391	0.1290	0.2989	+	+	+
SQSTM1	0.5540	0.5170	0.5628	+	+	+
TGM2	0.4270	0.3985	0.5245	+	+	+
KYNU	0.3226	0.3024	0.2428	+	+	+
RCN1	0.2183	0.2086	0.0303	+	+	
SERPINB6	0.1738	0.1692	0.0849	+	+	
G6PD	0.2185	0.2203	0.2347	+	+	+
KCTD12	0.3465	0.3516	0.2678	+	+	+
RPL35A	0.2149	0.2181	0.0072	+	+	
AGPAT9	0.5789	0.5943	NaN	+	+	
AKR1B1	0.3334	0.3449	0.3279	+	+	+
IFITM2; IFITM3; IFITM1	0.2499	0.2606	-0.0918	+	+	
SNRPG; SNRPGP15	0.3871	0.4088	-0.0110	+	+	
ERLIN2	0.2727	0.2915	0.1991	+	+	+
SPATA5L1	0.3092	0.3362	-0.0936	+	+	
S100A6	0.2965	0.3298	-0.0707	+	+	
GCLC	0.3342	0.3733	0.2624	+	+	+
MYLK	0.3999	0.4495	0.2707	+	+	
SERPINB9	0.2521	0.2862	0.2176	+	+	+
FAU	0.2038	0.2320	-0.0504	+	+	
HIST2H3A	0.1319	0.1587	-0.0138	+	+	

PGD	0.1551	0.1914	0.1776	+	+	+
MGST1	0.1666	0.2058	-0.0654	+	+	
HSPE1	0.1528	0.1931	-0.0862	+	+	
HMOX1	0.3582	0.4582	0.4652	+	+	+
TXN	0.1718	0.2247	-0.0633	+	+	
CKB	0.2063	0.2701	0.1794	+	+	
RPS28	0.1626	0.2228	-0.1141	+	+	
TUBB4A	0.2281	0.3211	-0.0477	+	+	
EPHX1	0.1871	0.2800	0.1974	+	+	+
USMG5	0.1900	0.2921	-0.0817	+	+	
LGALS1	0.1586	0.2449	0.0195	+	+	
SDCBP	0.1809	0.2799	0.1519	+	+	+
MFF	0.2648	0.4345	NaN	+	+	
PIGU	0.2954	0.6002	0.0807	+	+	
TXNRD1	0.1170	0.1025	0.1406	+		+
PLOD2	0.1909	0.1328	0.2170	+		+
RNF40	0.5578	NaN	0.7058	+		+
ZNF207	0.2994	NaN	0.5037	+		+
RAB27B	0.5577	NaN	0.5121	+		+
PTGES	1.0086	NaN	0.4841	+		+
PAPSS2	0.2463	0.0990	0.2463	+		+
TFPI2	0.4029	0.2700	0.5492	+		+
PTDSS1	0.3705	-0.1155	0.3481	+		+
TRA2B	0.1656	0.1693	0.1747	+		+
ATL3	0.1685	0.1051	0.1990	+		+
HTRA1	0.3151	0.1677	0.5779	+		+
EXOSC4	0.3183	NaN	0.5040	+		+
SERINC1	0.5131	0.1762	0.3665	+		+
EBF2	0.0006	0.4249	0.4813		+	+
SDCBP	0.2171	0.2826	0.5773		+	+
AIP	0.2365	0.2535	0.2470		+	+
KPNA6	-0.0616	0.1964	0.2209		+	+
FTL	-0.0495	0.3507	0.3473		+	+
NQO1	0.0656	0.1634	0.2085		+	+
PSM8; PSMB8	0.1866	0.2297	0.2313		+	+
GCLM	0.1239	0.2450	0.2468		+	+
TFCP2	NaN	1.0218	0.3476		+	+
DECR1	0.0942	0.2116	0.2221		+	+
RDH11	0.0982	0.2328	0.2675		+	+
NDRG1	0.2290	0.4208	0.4699		+	+
RIC8A	0.1312	0.2888	0.4033		+	+
PLXNA1	NaN	0.5035	0.4231		+	+

NaN = The peptides corresponding to the protein could not be identified in both heavy and light media samples.