



**Some pages of this thesis may have been removed for copyright restrictions.**

If you have discovered material in AURA which is unlawful e.g. breaches copyright, (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please read our [Takedown Policy](#) and [contact the service](#) immediately

# Molecular Modelling Assisted Design and Synthesis of Cyclodialdine Derivatives as DNA Gyrase Inhibitors

Bo Shan

Doctor of Philosophy

ASTON

SEPTE

This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with the author and that no quotation from the thesis and no information derived from it may be published without proper acknowledgement.

ASTON UNIVERSITY

**Molecular Modelling Assisted Design and Synthesis of Cyclothialidine  
Derivatives as DNA Gyrase Inhibitors**

A thesis submitted by Bo Shan BSc for the degree of Doctor of Philosophy

**Abstract:**

Since cyclothialidine was discovered as the most active DNA gyrase inhibitor in 1994, enormous efforts have been devoted to make it into a commercial medicine by a number of pharmaceutical companies and research groups worldwide. However, no serious breakthrough has been made up to now. An essential problem involved with cyclothialidine is that though it demonstrated the potent inhibition of DNA gyrase, it shows low activity against Gram-negative organisms. This is attributable to its inability to penetrate the outer membrane of negative organisms. In that equation, the hydrophobicity is a major factor influencing the cyclothialidine activity. The analogues synthesized new analogues have failed to prove that. In the structure based drug design stage, we designed a group of open chain cyclothialidine derivatives by applying the SPROUT programme and completed the syntheses. Improved activity is found in a few analogues and a 3D pharmacophore of the DNA gyrase B is proposed to lead to synthesis of the new derivatives for development of potent antibiotics.

## Acknowledgements

© 2004 Pearson Education, Inc.

To Mum, Dad and my wife Hongyang

## Acknowledgements

I would like to thank my supervisors Dr. Yongfeng Wang and Dr. Carl Schwa for their kindly help and advice as well as the final checking of this thesis.

Thanks to Karen Farrow for her technical support and Mike Davies for keeping the lab tidying and running.

Thanks to Dr. Peter Lambert for the help in biological testing and for his calm advice and support.

A special thanks to Dr. Dan Rathbone for his installation of the molecular modelling program and inspiring me in the first operation, Dr. John Williams for his UNIX system directing, Dr. Bill Fraser for his HPLC helping. Last but not least, to Dr. Kejun Zhao for his great help in the synthesis work.

Thanks to all my friends in the medicinal chemistry group, especially Kejun Zhi, Dan Fei, Yi Ge and those in the lab for their friendship and company. Thanks also to my working partner, Jiayun Pang, for helping me in the chemical synthesis.

Finally, thanks to my Mum and Dad for their continued supporting and my wife Hongyang Li. I cannot submit this thesis without them.

Special thanks go to Northpharm Institute for the finance support to the project and to UK government for Overseas Research Awards Scheme (ORS) to help with fees.

## Abbreviations

<b>ACN</b>	acetonitrile
<b>ATP</b>	adenosine triphosphate
<b>DCC</b>	dicyclohexylcarbodiimide
<b>DCM</b>	dichloromethane
<b>DEAD</b>	diethylazodicarboxylate
<b>DMF</b>	dimethylformamide
<b>DMSO</b>	dimethylsulphoxide
<b>DNA</b>	deoxyribonucleic acid
<b>EDC</b>	1-(3-Dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride
<b>HOBt</b>	1-hydroxybenzotriazole
<b>MS</b>	mass spectrometry
<b>NBS</b>	N-bromosuccinimide
<b>NMR</b>	nuclear magnetic resonance
<b>R<sub>f</sub></b>	retention coefficient
<b>RNA</b>	ribonucleic acid
<b>TBAF</b>	tetra-butylammonium fluoride
<b>TBDMS</b>	tert-butyldimethylsilyl
<b>TEA</b>	triethylamine
<b>TFA</b>	trifluoroacetic acid
<b>THF</b>	tetrahydrofuran
<b>TLC</b>	thin layer chromatography
<b>TMS</b>	trimethylsilyl

# CONTENTS

<b>Title</b>	1
<b>Abstract</b>	2
<b>Dedication</b>	3
<b>Acknowledgements</b>	4
<b>Abbreviations</b>	5
<b>Contents</b>	6
<b>List of Figures</b>	10
<b>List of Tables</b>	11
<b>Chapter 1. Introduction</b>	12
(1). Modern drug design in the pharmaceutical industry	13
(2). Molecular modelling of a new compound	16
(2.1) Overview	16
(2.2) Current assessment	19
(2.3) Future issues	21
(2.4) The concepts in molecular modelling	23
(2.4.1) The three dimensional pharmacophores	23
(2.4.2) Conformation analysis	24
(2.4.3) Molecular docking	25
(2.4.4) Quantitative Structure-Activity Relationship (QSAR)	27
(2.4.4.1) Multiple linear regressions	30

- (2.4.4.2) Non-linear models: neural network and genetic algorithms
- (2.4.4.3) Principal component analysis
- (2.4.4.4) Partial least squares
- (2.4.4.5) Partial least squares and molecular field analysis
- (2.4.5) The protein and ligand interaction
- (2.4.6) Structure based ligand design
- (2.4.7) Pharmacophore identification
- (2.4.8) The protein modelling
- (2.5) The molecular modelling program
  - (2.5.1) Overviews
  - (2.5.2) The SPROUT program
    - (2.5.2.1) Program overview
    - (2.5.2.2) Primary constraints
- (3). Biological function and chemical synthesis of Cyclothialidine
  - (3.1) Topoisomerases
    - (3.1.1) Topoisomerases I
    - (3.1.2) Topoisomerases II
    - (3.1.3) DNA gyrase
  - (3.2) Cyclothialidine, an inhibitor of DNA gyrase

## **Chapter 2. Methods**

- (1). Cyclothialidine synthesis route
- (2). The Mannich reaction and Mannich base reduction
  - (2.1) The Mannich reaction



- (2.2) The Mannich base reduction
- (3). Peptide synthesis
  - (3.1) Coupling Reagents
    - (3.1.1) DCC (dicyclohexylcarbodiimide) reagent
    - (3.1.2) 1-(3-Dimethylaminopropyl)-3-ethylcarbodiimide hydrochloric (EDC) reagent.
  - (3.2) Active esters
- (4). Carboxylic acid protection
- (5). Hydroxyl group protection
- (6). Bromination
- (7). The sulfide preparation
- (8). Deprotection of 4-nitrobenzyl ester
- (9). Cyclisation
- (10). Broth microdilution method for the determination of MIC values
- (11). The enzyme-ligand design method
  - (11.1) CAnGAROO
  - (11.2) HIPPO
  - (11.3) EleFAnT
  - (11.4) SPIDeR
  - (11.5) ALLigaTOR
- (12). Quantitative Structure-Activity Relationships (QSARs) methods
  - (12.1) TSAR Program
  - (12.2) Data reduction
  - (12.3) Multiple Regression Analysis
  - (12.4) The Cross Validation

### **Chapter 3. Experimental**

- (1). Synthesis of Cyclothialidine analogues
- (2). Synthesis of open chained gyrase inhibitors

### **Chapter 4. Results and discussion**

- (1). The whole plan.
- (2). QSAR analysis
- (3). Structure-based ligand design
  - (3.1) Figure 19b leading synthesis
  - (3.2) Figure 19a leading synthesis

### **Chapter 5. References**

## List of Figures

Figure 1. Classification of ligand design software	43
Figure 2. The program's procession	52
Figure 3. DNA gyrase's structure	58
Figure 4. The activity of <i>E.coli</i> DNA gyrase	59
Figure 5. Structures of Coumarins	61
Figure 6. The structure of Cyclothialidine	63
Figure 7. Retrosynthetic analysis	64
Figure 8. Ro-61-6653	64
Figure 9. The organic synthesis scheme	68&95
Figure 10. The two amino acids (in red) within the binding pocket of the B subunit	83
Figure 11. The four target sites making up the binding cleft	84
Figure 12. The fulfilled templates to the active sites	85
Figure 13. A complete skeleton produced at the end of SPIDeR	87
Figure 14. The crystal structure of Cyclothialidine with gyrase B	158
Figure 15. The 2D structure of Cyclothialidine with some active site	158
Figure 16. Rudolph's seco-Cyclothialidine structures	159
Figure 17. The key pharmacophore for gyrase B	160
Figure 18. The fulfilled templates to the correlated sites	161
Figure 19a and 19b. The designed structures by SPROUT	161&162
Figure 20. The superimposition of Figure 19a and Rudolph's structure	162
Figure 21. The superimposition of 55 and 66	168
Figure 22. The possible pharmacophore of gyrase B	169

## List of Tables

<b>Table 1. The structure and antibacterial activity of Cyclothialidine analogues</b>	<b>148</b>
<b>Table 2. The logarithm of mole activity</b>	<b>150</b>
<b>Table 3. The correlation matrix for all the parameters (partly displayed)</b>	<b>151</b>
<b>Table 4. The parameters and principal component matrix</b>	<b>152</b>
<b>Table 5. The new analogues' activity to gram-negative bacteria</b>	<b>155</b>
<b>Table 6. The synthesized linear chain analogues</b>	<b>164</b>
<b>Table 7. The synthesized side chain analogues</b>	<b>166</b>
<b>Table 8. The MICs to another eight organisms</b>	<b>167</b>

## **Chapter 1. Introduction**

## **Drug design and molecular modelling**

### **(1). Modern drug design in the pharmaceutical industry**

A key recent trend in the industry has been the integration of what used to be considered "development" activities into the early stages of drug discovery. The aim is to identify and promptly reject candidate molecules that are likely to fail in the later phases of discovery and development. Even advanced robots cannot make all the molecules under consideration. Researchers must try to concentrate on compounds with the best possible chance of becoming drugs. They must identify and / or design a subset of drug-like molecules from all the compounds that could be synthesized.

The ADME (Absorption, Distribution, Metabolism, Excretion) screens are used by most pharmaceutical companies to discard candidate molecules that are likely to fail later on, according to the motto "fail fast, fail cheap". Thus, absorption is checked using Caco-2 (a human intestinal epithelial cell line originally derived from a colorectal carcinoma) or MDCK (Madin-Darby canine kidney) cell monolayers. Susceptibility to metabolism is evaluated using liver microsomes or hepatocytes.

Though valuable, these experimental filters have disadvantages. They require physical samples of compounds for testing, and even with advanced technology they are time-consuming and resource-intensive. Thus there is a great need to develop and apply computational methods for predicting "drug-likeness". Such methods could be used to eliminate poor candidates from a set of virtual compounds or libraries before any synthesis is carried out. This process would be rapid and cost-effective.

Current computational methods for predicting drug-likeness will now be reviewed, with particular emphasis on:

- Drug-likeness in a general sense
- Intestinal absorption
- Blood-brain barrier (BBB) penetration.

It should be noted that only the prediction of absorption by passive mechanisms will be covered here. Although of much interest and importance, neither the prediction of active transport by carrier systems nor the metabolism of compounds will be considered.

#### *Prediction of general drug-likeness*

Initial attempts to achieve this goal created computational filters to remove chemically unsuitable compounds (1-4). Such filters frequently set limits on molecular weight (MW) and the number of rotatable bonds, and include substructure searches for toxic or reactive groups. Although these methods provide a very useful initial filtering of any data set before further data analysis, they attempt to eliminate definite non-drugs and not to define the features required in a good drug. A more sophisticated strategy for predicting drug-likeness compares two sets of compounds, one containing known drugs and the other consisting of compounds known (or presumed) not to be potential drugs. Computer methods are implemented to identify the properties or characteristics of the former set that differentiate them from the latter. The methods are of two main types, based either on genetic algorithms or on neural networks.

#### *Prediction of intestinal absorption*

For most drugs the oral route of administration is preferred, if the bioavailability is sufficient. Researchers have therefore attempted to define the physicochemical properties that favour intestinal absorption (5,6) and to develop computational methods to predict it (7).

### *Lipinski's rule of five*

This rule is probably the best-known method. It was devised by Lipinski and coworkers at Pfizer (Groton, NJ, USA) through analysis of 2245 drugs from the World Drug Index (WDI) believed to have entered Phase II trials. As implemented in the Pfizer registration system, the rule-of-five warns of possible absorption problems for compounds fulfilling any two of the following conditions:

- Molecular weight > 500
- Number of hydrogen-bond acceptors > 10
- Number of hydrogen-bond donors > 5
- Calculated log P > 5.0

To implement this rule, any oxygen (O) and nitrogen (N) atoms are defined as hydrogen-bond acceptors, and N-H and O-H groups as hydrogen-bond donors. Log P is a measure of lipophilicity, referred to the octanol-water partition coefficient of a compound. A variety of software for calculating log P is available, including C log P (Daylight Information Systems, Mission Viejo, CA, USA) and M log P (developed by Moriguchi and coworkers) (8). Some of these findings have been reinforced by an analysis of the CMC database by Amgen (Thousand Oaks, CA, USA), particularly concerning the preferred ranges of MW and log P.

### *Veber's criteria of two*

In 2002, Veber and coworkers at Glaxo SmithKline suggested (10) that compounds meeting only the following two criteria will have a high probability of good oral bioavailability in the rat:

- Number of rotatable bonds = 10 or fewer
- Polar surface area =  $140 \text{ \AA}^2$  or less (or 12 or fewer H-bond donors and acceptors)



Here a "rotatable bond" was defined as any single bond, not in a ring, bound to a non-terminal non-hydrogen atom. Amide C-N bonds were not included in this count because of their high rotational energy barrier. Hydrogen bond donors were defined as any heteroatom with at least one bonded hydrogen atom. Hydrogen bond acceptors were taken as any heteroatom without a formal positive charge, excluding halogens, pyrrole nitrogen, heteroaromatic oxygen and sulfur, and higher oxidation states of nitrogen, phosphorus, and sulfur but including the oxygen bonded to them.

### *Importance of BBB penetration*

The BBB is a complex cellular system which maintains the homeostasis of the CNS by separating the brain from the systemic blood circulation (11). Drugs targeted to the CNS must penetrate the BBB. However, if drugs aimed at other sites of action pass through the BBB, there is a risk of unwanted side effects.

## **(2). Molecular modelling of a new compound**

### **(2.1) Overview**

It is time-consuming and expensive to produce a new medicine, because it must (i) produce the desired response with minimal side effects and (ii) be demonstrably better than existing therapies. The first task is often to identify one or more lead compounds. A lead compound shows activity in an appropriate assay. This lead compound is then altered to enhance its potency and selectivity, to ensure that it and also its metabolites are non-toxic, and to provide appropriate transport characteristics ensuring that it can pass through cell membranes and reach its target.

Many drugs function by interacting with a biological macromolecule such as an enzyme, DNA, glycoprotein or receptor. The interaction between a drug and its target may be due entirely to non-bonded forces, but in some cases covalent

bonding may occur. Drugs which interact with receptor proteins may act as agonists, antagonists or inverse agonists. Agonists produce the same effect as the natural substrate or effector molecule, perhaps even more strongly. Antagonists block the effect of the natural ligand. Inverse agonists create the opposite effect.

For a ligand to bind tightly to its target, it often shows a high degree of complementarity, assessed and measured in various ways. Many ligands have shapes that match the region of the macromolecule to which they bind (the binding site). The ligand often forms hydrogen bonds with the receptor. Some receptors have hydrophobic "pockets", formed by groups of non-polar amino acids, into which an appropriately sized hydrophobic group on the ligand can fit. It is also crucial to remember, as mentioned earlier, that an administered drug must get to its target before it can bind to it. Frequently a drug must pass through cell membranes, which are hydrophobic, and so the drug must be sufficiently lipophilic (lipid loving) to partition into the membrane. The drug must also evade metabolism and excretion.

Finding novel lead compounds can be very difficult. Often this has happened by serendipity. For many years pharmaceutical companies have screened soil and other biological samples to find new leads, but extracting and purifying any active ingredients can be difficult. Modern combinatorial chemistry techniques can now generate a large number of compounds. Test compounds can be rapidly screened using highly automated, robotic techniques. As a complementary technique, molecular modelling has much to offer.

For an increasing number of target macromolecules a three-dimensional structure is available. Such structures may be directly obtained by X-ray crystallography or NMR, or inferred from the structures of related macromolecules by a theoretical method such as homology modelling. Even when such detailed information about the structure of a target receptor is not available, it may be possible to derive a *pharmacophore*, which is an abstract model that indicates the key features of a series of active ligands. Molecular modelling aims to suggest compounds likely to interact well with the receptor, or that contain the required functional groups of the pharmacophore.

Once a lead compound has been identified, it is modified to enhance its properties. Molecular modelling techniques can suggest what modification to make and can help in understanding the experimental binding results.

Calculations using quantum mechanics and molecular mechanics can provide information about electronic and conformational properties. Molecular dynamics can simulate the dynamical behaviour of the ligand-receptor system and can calculate the relative free energies of binding of alternative compounds.

The assumption that computational methods can speed the discovery of novel compounds and guide the optimisation of their properties (12) is now well justified. Computational methods have been primarily used in the discovery phase of drug development (13), usually focused on *in vitro* testing. Important goals are the reduction of screening costs or of the time required to identify a candidate (12). Lead discovery is a search. As our understanding of cellular biochemistry has improved, it has become possible to target specific enzymes through modification of their substrates or cofactors. This information can be used in many ways. For example, conventional screening of corporate databases, the National Cancer Institute database, or of novel natural products has been a route to more than half the best-selling drugs. Alternatively, the search can proceed via computational methods both in the presence and the absence of structural information about the target. Advances in chemistry and molecular biology have led to the creation of libraries of organic compounds, peptides, oligonucleotides, and even whole proteins for screening (14-18).

However, discovery is only the first stage in the long process leading to a marketable drug. The following steps can be identified: (a) identification of a lead compound, (b) optimisation of this lead, (c) evaluation of delivery and metabolic issues, (d) optimisation of synthesis, and (e) animal and clinical testing. Computational methods also have potential in evaluating the pharmacological and biopharmaceutical properties of molecules that relate to *in vivo* test outcomes for a molecule. There are also eventual computational issues involved in the metabolism and toxicology of drug candidates that may determine their clinical success.

Computer-assisted molecular design has contributed both to the design of potent *in vitro* lead molecules and to the development of clinically useful drugs (19-28). It is not always easy to identify such success. Details may not be released because of commercial confidentiality, and it is difficult to discover the true stories from a few sentences in a scientific publication or meeting abstract, a news release, or the Food and Drug Administration (FDA) database. The

successful search and discovery methods fall into several types. Database-searching methods find molecules that have shape and chemical interactions complementary to a receptor structure provided by the user (29-30). Three-dimensional substructure searches (pharmacophore searches) identify molecules that include specified functional groups or types of groups in defined geometric relations (31). Similarity searches rank molecules in a database by the number of features, which match those of a submitted query (32). As an alternative to these methods for searching a database of compounds, modelling based on the interaction site (33-34) and *de novo* compound design are gaining popularity (35). Complementary to these search and discovery methods is the statistical analysis of biological response data. Two-dimensional (2D) and three-dimensional (3D) quantitative structure-activity relationships (QSAR) seek to correlate the biological responses with some calculable molecular properties.

More often than not, more than one of these techniques must be combined to achieve success. For example, conformational analysis is often the first step in any or all of the techniques listed, and pharmacophore studies often are the precursor for 3D-QSAR studies and database searches. QSAR studies and structure-based design are mainly used in this thesis.

## **(2.2) Current assessment**

It is difficult to evaluate the contribution of computer-aided molecular design to drug discovery because it is just one part of a complex undertaking. Like any other tool, it should be judged by how well it works: correct identification of compounds that bind to the target; correlation of structure and activity; accuracy of geometrical relationships predicted by a pharmacophore model; predictability of a statistical model; correct ligand placement, ligand conformation and macromolecular conformation in docking calculations. The practical matters of cost and speed are also important.

Lead identification is the easiest task because a limited proportion of false-positive hits are acceptable and we will never know about false-negative ones. Present programs are relatively successful: typical hit rates of 1-10% are competitive with high throughput experimental screens, and computer screening is much cheaper. The false-negative issue is important and could be addressed by

running a substantial, diverse database of compounds through both computer and experimental screens. Searches for antagonists of molecular recognition or macromolecular conformation changes have been generally less successful, but some progress has been reported (36). Nucleic acid targets have also been investigated (37-39).

The prediction of binding free energy for diverse ligands bound to a common receptor has achieved a degree of success in line with the analysis above. When force-field terms are used, average errors usually are a few kilocalories per mole. Comparing scores with  $K_i$  or  $IC_{50}$  data spanning many orders of magnitude for diverse inhibitor-receptor complexes provides a very severe test of current methods. Much better agreement is obtainable by using the most computationally intensive techniques and limiting attention to a family of analogues. Present computational methods are best suited for extracting the more active compounds from a database. Compounds thus retrieved have shown selectivity for the target (40-42).

The most difficult task for a docking computation is to predict the binding geometry (or geometries) of a ligand, because this requires the comparison of free energy for several alternative binding modes. Computer methods usually suggest reasonable binding geometries (43-44), but they do not always find the dominant binding mode or forecast the conformational rearrangements that may happen. With the DOCK program, the best cases have shown displacements of 1-2 Å between predicted and experimental geometry, rising to about 5 Å in the worst cases. Problems arise from the conformational freedom of the ligand and receptor, possible alternative binding modes (configurational freedom), and the inclusion of water molecules or ions as part of the binding complex (45). Using recent advances in the X-ray crystallography of macromolecules, one can combine structural experiments with computations to assess the degrees of freedom in the system of interest (44). Some successful efforts at structure-based design have carried out one X-ray structure determination for every one to two compounds synthesized (46).

The success of database searching programs is difficult to evaluate rigorously. Ideally, all compounds in the database ought to be screened, ranked in order and compared with the rank ordering obtained computationally. This is not feasible because the chemical and chiral purity of every compound would have to be

checked, the equilibrium binding constant of every compound would have to be measured, and the computation would have to be adjusted to match the assay conditions. For directed combinatorial libraries, in which only a limited set of compounds are synthesised, evaluation is even more difficult.

Overall, computer-aided molecular design must be judged by its ability to identify molecules that can go forward into the clinic. Up to now it has been most effective for enzyme inhibitors. Even allowing for the capital cost of structure determination, it does seem to offer some speed-up and cost reduction because fewer syntheses are required. It is scientifically attractive because it offers testable structural hypotheses to guide the project.

### **(2.3) Future issues**

We can expect the techniques for database searching to keep on improving, as they have done recently. Specifically, the exploration of conformations and the evaluation of  $\Delta G_{\text{binding}}$  ought to get better through improvements in both algorithms and computers. Second- and third-generation force fields will move into general use (47-51). Increased processor speeds will extend the usefulness of free-energy perturbation computations. Within a single target system, it is reasonable to look forward to accuracy of  $\pm 1 \text{ kcal mol}^{-1}$ , which would facilitate routine searches of conformational space for both ligands and side chains. Finally, as computers become equipped with bigger physical memory, less time needs to be wasted on disk input and output, so that database searches will become more efficient.

Improved engineering of user interfaces to computer-aided discovery programs should make access to results easier in the future. Hits are now returned without consideration of availability, cost, synthetic accessibility, solubility, toxicity profiles, or likely impurities. These factors must be evaluated by an experienced chemist. Removal of compounds in chemical classes considered uninteresting, or because they contain biologically undesirable groups should be straightforward, and the final list could include only those compounds

that meet user-defined criteria. An expert system to evaluate synthetic feasibility would be valuable both in *de novo* design as well as database searching (52-53).

Combinatorial chemistry, which simultaneously provides many compounds with a common skeleton (54) provides an exciting opportunity for chemistry and computation to meet. Computation can help with optimal design for diversity and for directed libraries (55). Docking procedures can help with the selection of scaffolds and side chains. Large libraries of virtual compounds can be examined prior to synthesis.

There have been clear examples in which priorities for synthesis have been usefully set by considering the results of free-energy perturbation calculations (60). The word "iterative" appears frequently in design studies. Models must evolve as more high-quality experimental results become available. In the coupling of calculations to combinatorial chemistry (54), techniques developed for the design of chemical libraries, monomer selection and analyses, have led to *in vitro* successes (61-62). Computational techniques for the optimisation of chemical libraries have also advanced (63-64).

A crucial biological challenge to structure-based drug design is the rapid emergence of resistance to anti-infective drugs, such as virulent resistance to even sub-nanomolar inhibitors of retroviral enzymes (56). Mutations to enzymes are generally expected to make them less efficient (57), and this is true for point mutations of the HIV protease (58). However, the virus is able to produce a wide range of multiple mutations located away from the active site, some of which appear to restore wild-type activity while reducing sensitivity to any single protease inhibitor. Even inhibitors designed to interact mainly with backbone atoms (59) have generated significant resistance. These events challenge the foundations of structure-based design: that drugs and natural substrates compete for the same enzyme pockets. The rules for this competition have not been clearly enunciated. It may be necessary to explore large-scale kinetic barriers for the entire enzyme.

Another issue is that the design of agonists is much more difficult than the design of antagonists. Antagonists can block processes in many ways, but the design of agonists generally requires fairly precise knowledge of the biological mechanism. A promising set of targets is the family of pharmacological receptors that use dimerization to generate a signal.

However, most of the success stories in drug design reported up to now seem to have relied on fairly fundamental methods such as molecular graphics, interaction energies and molecule-docking, coupled with a medicinal chemist's intuition. The success of these simple methods is due to the ease of including them in a multi-disciplinary approach, which allows early incorporation of expertise on synthesis and bioavailability. It seems that structure-based approaches possess the greatest overall potential. From the point of view of this thesis, it is encouraging that structural information (available in the case to be studied) is so helpful and that it has not been an absolute requirement to use the very most complicated and expensive software approaches.

## **(2.4) The concepts in molecular modelling**

### **(2.4.1) The three-dimensional pharmacophores**

In drug design, the pharmacophore refers to a set of features that is common to a series of active molecules. Typical features include hydrogen bond donors and acceptors, positively and negatively charged groups, and hydrophobic groups. A three-dimensional (3-D) pharmacophore defines the spatial relationships between such groups, often expressed as distances and distance ranges. If necessary, other measures of geometry such as angles and planes may be included. Our improving ability to study the conformations of ligands has stimulated interest in the influence of the three-dimensional structure of molecules on their chemical and biological activity. Pharmacophore mapping aims to determine possible 3D pharmacophores for a series of active compounds. Such a pharmacophore can then be used to find or suggest other active molecules.

Two problems must be overcome when calculating 3D pharmacophores. Unless all the molecules are completely rigid, one must consider their conformational properties. Secondly, one must determine those combinations of pharmacophoric groups that are common to the molecules and can be positioned in a common orientation in space. More than one pharmacophore may be possible; some algorithms generate hundreds of possible pharmacophores which are then evaluated to find out which best fits the data. The important underlying



assumption is that all of the molecules bind to the macromolecular target in a common manner.

#### **(4.2) Conformation analysis**

The physical, chemical and biological properties of a molecule are critically affected by the 3D structures, or conformations, that it can adopt.

Conformational analysis is the study of this topic. It was pioneered by D. H. R. Barton, who showed in 1950 that the reactivity of substituted cyclohexanes was influenced by the equatorial or axial location of the substituents (65).

Conformational analysis advanced as conformations were actually determined by NMR and infrared spectroscopy and X-ray crystallography.

Conformations are traditionally defined as those spatial arrangements of the atoms of a molecule that can be interconverted purely by rotation about a single bond. This definition is slightly relaxed to accommodate the small alterations in bond lengths and angles that accompany conformational changes, and to recognise that rotations can occur about bonds in conjugated systems that have an intermediate bond order between single and double.

A key purpose of conformational analysis is the conformational search to identify the preferred conformation(s) of a molecule that determine its behaviour. This usually involves the location of conformations that are at minimum points on the energy surface. Therefore energy minimisation calculations are an essential tool in conformational analysis. Energy minimisation methods only locate the minimum point that is closest to the starting structure. Thus it is vital to have a separate efficient algorithm which generates the starting structures for subsequent minimisation. Solvation effects may have an important influence on conformational equilibria. Computationally efficient methods are now available for calculating the free energy of solvation of a conformation, which can be considered along with the intramolecular energy.

Ideally this procedure should locate every minimum-energy conformation on the energy surface. However, this may be impractical because there are too many

minima, and it is only feasible to search for all accessible minima. For proteins, in particular, there are too many minima. Under such circumstances, it is often assumed that the conformation with the very lowest value of the energy function (the global minimum) corresponds to the native (i.e. naturally occurring) conformation. Even though the global minimum energy conformation has the lowest energy, it may not have the highest population because the statistical weight of each structure is affected by the vibrational energy levels. Furthermore, the global minimum energy conformation does not necessarily represent the active structure. In some cases the active structure may not correspond to any minimum on the energy surface of the isolated molecule. A molecule may even need to adopt more than one conformation in order to exert its biological function. For instance, a substrate may bind to an enzyme in one conformation and change to a different conformation prior to reaction.

### **(2.4.3) Molecular docking**

Molecular docking attempts to predict the structure (or structures) of the intermolecular complex formed when two or more molecules associate. Docking is widely used to suggest how inhibitors bind to proteins. Most docking algorithms generate many possible structures, which therefore must be scored to identify the most interesting ones. Thus the “docking problem” involves the generation and evaluation of plausible structures of intermolecular complexes (66).

Sometimes a manual approach to the docking problem, using interactive computer graphics, can be very effective. We need to have a good idea of the expected binding mode, perhaps from the already known binding mode of a closely related ligand. However, even in such cases caution is required; X-ray crystallography has shown that very similar inhibitors may adopt significantly different binding modes. Automatic docking algorithms suffer less from bias and usually consider many more possibilities.

Such algorithms take various forms, differing in the number of degrees of freedom they ignore. The earliest and simplest algorithms treated both molecules, i.e. the small molecule ligand and the protein or DNA target, as completely rigid

bodies. Sphere centres are defined at appropriate positions in the binding site. Ligand atoms are matched to these sphere centres to find matching sets (cliques) in which all the distances between the ligand atoms in the set are equal (within a user-specified tolerance) to the distances between corresponding sphere centres. The orientation of the ligand in the site is confirmed by means of a least-squares fit of its atoms to the sphere centres. This orientation is checked to verify the absence of unacceptable steric interactions between ligand and receptor. If this orientation is acceptable, the interaction energy is calculated as the "score" for that binding mode. Alternative orientations are generated by matching different sets of atoms and sphere centres. The top scoring (lowest energy) orientation is retained for further analysis.

Methods for conformationally flexible docking have to consider the conformational degrees of freedom, usually only for the ligand. The receptor is still assumed to be rigid. All common methods for searching conformational space have at some stage been incorporated into a docking algorithm. The internal conformation of the ligand is changed by rotating a particular bond, or else the entire molecule is randomly translated or rotated. The energy of the ligand within the binding site is calculated using molecular mechanics, and the change is accepted or rejected (67). Thus both the orientation and the internal conformation will evolve. If a population of structures is considered, the score of each docked structure acts as the fitness function to determine which individuals continue into the next generation.

Docking can be based on distance geometry, although the major obstacle to its use is the need to generate conformations of the ligand within the binding site. One way to achieve this is to incorporate a modified penalty function that restrains the ligand conformation to keep it within the binding site.

The ideal docking method would explore the conformational degrees of freedom for the receptor as well as the ligand. The most "natural" way to allow for the flexibility of the binding site appears to be a molecular dynamics simulation of the ligand-receptor complex. However, such calculations demand a lot of computational resources and are practically useful only for the refinement of structures produced by other docking methods. Molecular dynamics does not sample the full range of binding modes well except for very small, very mobile

ligands. Otherwise the energy barriers that separate one binding mode from another are often too large to overcome.

#### **(2.4.4) Quantitative Structure-Activity Relationship (QSAR)**

It is well known that the pharmaceutical activity of a compound is related to its structural properties. Commonly these relationships are expressed in statistical terms and used to predict the behaviour of untested molecules. Such a relationship between the activity of a compound and its structural and physico-chemical properties is called a quantitative structure-activity relationship (QSAR).

The statistical methods which can be applied to determine QSARs are well established. Problems arise in the majority of cases because the action of the compounds is not fully known and it is often difficult to decide which molecular parameters to include in the statistical analysis. For predictions based on regression analysis, it is generally necessary to reduce the number of molecular descriptors originally considered, in order to achieve a valid relationship.

Procedures to do this include:

- Selection of descriptors based on correlation with observed activity
- Factor analysis, such as principal component analysis (PCA)
- Partial least squares (PLS)
- Weighted least squares (WLS)
- Clustering

For the first three methods a QSAR can be generated by multiple linear regression or principal component regression (PCR) between the selected variables and the observed activity.

QSAR analyses are usually applied to a series of molecules with a common core framework but different substituents (R-groups). The analysis aims to correlate the biological activity or physico-chemical properties of the molecules with molecular parameters associated with the substituents. Traditional, or 2D, QSAR considers descriptors such as electronic, lipophilic and steric parameters,

while 3D QSAR incorporates modelling parameters, such as inter-atomic distances or molecular surface areas.

In drug design, a molecule usually needs qualities beyond *in vitro* potency. As already emphasised, a potent enzyme inhibitor is of little pharmaceutical use if it cannot reach its target. The *in vivo* activity of a molecule often depends on many factors. A QSAR can help to decide which molecular features are conducive to activity, thus guiding the search for modified compounds with enhanced properties. The mathematical relationship between these numerical properties and the activity is often described by an equation of the general form:

$$V = f(p)$$

V is the activity in question, p is structure-derived properties of the molecule, and f is some function. An early example of a structure-activity relationship was a correlation between the potencies of narcotics and the partition coefficient of the compound between oil and water. This interpretation was that the narcotic effect is due to physical changes caused by the dissolution of the drug in the lipid component of cells.

The first use of QSAR to rationalise biological activity is usually attributed to Hansch (68). He developed equations which related biological activity to a molecule's electronic characteristics and hydrophobicity. For example:

$$\log (1/C) = k_1 \log P - k_2 (\log P)^2 + k_3 \sigma + k_4$$

C is the concentration of the compound required to produce a standard response in a given time, log P is the logarithm of the partition coefficient of the compound between 1-octanol and water which was chosen by Hansch as a suitable measure of relative hydrophobicity,  $\sigma$  is the Hammett substituent parameter and  $k_1$  though  $k_4$  are constants.

The hydrophobic component was considered to model the ability of the drug to pass through cell membranes. This equation contributed the recognition that there

would be an optimal value of the hydrophobicity: too low and the drug would not partition into the cell membranes; too high and the compound would partition into the membranes but tend to remain there rather than proceeding to the actual target. This explains the dependence of the activity upon  $\log P$ . An alternative way to express the Hansch equation uses a parameter  $\pi$ . This is the logarithm of the partition coefficient of a compound with substituent X relative to a parent compound in which the substituent is hydrogen:

$$\pi = \log (P_X/P_H)$$

Thus

$$\text{Log } (1/C) = k_1\pi - k_2\pi^2 + k_3\sigma + k_4$$

Hansch used the Hammett substituent parameter as a measure of the electronic characteristics of the molecules. Hammett showed that the positions of equilibrium and the reaction rates of the series of related compounds:

$$\text{Log } (k/k_0) = \rho\sigma \quad \text{or} \quad \log (K/K_0) = \rho\sigma$$

$k_0$  and  $K_0$  are the rate constant and equilibrium constant respectively for a 'reference' compound (usually a hydrogen-substituted compound). The substituent parameter  $\sigma$  depends only upon the nature of the substituent and whether it is *meta* or *para* to the carboxyl group. The reaction constant  $\rho$  is fixed for a given process under specified experiment conditions. The 'standard' reaction is the dissociation of the benzoic acids which have  $\rho = 1$ . A full discussion of linear free energy relationships can be found in many physical organic textbooks.

The QSAR equation derivation involves a number of steps. First, it is obviously necessary that the compounds should be synthesized and their biological activities should be determined. QSARs can be derived for very different sets of compounds, but it is more common to consider a related series of compounds that only are different from just one part of the molecule. These

differences can then be expressed as appropriate substituent constants. When we plan which compounds to be synthesised, it is important to cover the range of properties that may affect the activity. The series of compounds with almost the same partition coefficients are not always being chosen. The properties chosen for inclusion in the QSAR equation should ideally be uncorrelated with each other. Many different parameters have been used in QSAR equations, most of which are designed to represent the hydrophobic, electronic or steric characteristics of the molecule. Some of these parameters are properties of the entire molecule, such as partition coefficient. Others such as the Hammett constant  $\sigma$  are substituent values, indicating the value relative to a standard substituent (usually hydrogen). Steric effect can be modelled using parameters that are computed from the geometry of each substituent. Molar refractivity (MR) is also used to model steric effects. The refractive index does not vary much from one organic compound to another and because the molecular weight divided by the density equals the volume, MR indicates the steric bulk of a molecule.

Molecular modelling methods can also be used to calculate the values of some of the properties. Many QSAR equations contain parameters that are related to the electronic structure of the molecule such as the dipole moment, the atomic charges and the orbital energies (especially the HOMO and LUMO energies). These parameters must be obtained theoretically using quantum mechanics.

#### **(2.4.4.1) Multiple Linear Regressions**

The most widely used technique for deriving QSAR equation is multiple linear regression, which uses least-square fitting to find the best combination of coefficients in the QSAR equation. We can illustrate the least-square technique using the simple case where the activity is a function of just one property. The equation is therefore derived:

$$y = mx + c$$

y is known as the dependent variable (the observation) and x is the independent variable (the parameter). For example, y might be the activity and x might be the log P. The objective for a regression is to find the coefficients m and c that minimise the sum of the squared deviations of the observations from the fitted equation. The coefficients m and c in the linear regression equation are given by:

$$m = \frac{\sum_{i=1}^n (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sum_{i=1}^n (x_i - \langle x \rangle)^2} \quad ; \quad c = \langle y \rangle - m \langle x \rangle$$

The regression equation passes through the point ( $\langle x \rangle$ ,  $\langle y \rangle$ ), where  $\langle x \rangle$  and  $\langle y \rangle$  are the means of the dependent and independent variables respectively. The quality of a regression equation is often reported as the  $R^2$  value. This indicates the fraction of the total variation in the dependent variables that is explained by the regression equation. To determine  $R^2$  the sum of squares of the deviations of the observed y values are calculated both from the mean  $\langle y \rangle$  and the predicted values from the regression equation,  $y_{p,i}$ :

$$SS_{\text{mean}} = \sum_{i=1}^n (y_i - \langle y \rangle)^2 \quad SS_{\text{error}} = \sum_{i=1}^n (y_i - y_{p,i})^2$$

$y_{p,i}$  is obtained by feeding the appropriate  $x_i$  value into the regression equation. The  $R^2$  is then given by:

$$R^2 = \frac{SS_{\text{mean}} - SS_{\text{error}}}{SS_{\text{mean}}}$$

$R^2$  can adopt values between 0.0 and 1.0; a value of 0.0 indicates that none of the variation in the observation is explained by variation in the independent variables



whereas a value of 1.0 indicates that all of the variation in the observations can be explained. A disadvantage of the standard  $R^2$  value is that it is dependent upon the number of independent variables, with high  $R^2$  value being obtained for larger data sets. More sophisticated statistical measures should ideally be used.

It is straightforward to extend this analysis to more than one independent variable: such calculations are difficult to perform by hand but can be performed using a statistical package.

There are some important criteria to consider when using multiple linear regressions. To achieve statistically significant results there should be sufficient data; it is often considered that at least five compounds are required for each parameter included in the regression analysis. The selected compounds should give a good spread of values of the parameters, which should be uncorrelated. Compounds, which have a value for some parameter that is greatly different from the remainder, should be examined very closely.

#### **(2.4.4.2) Non-linear models: neural networks and genetic algorithms**

Neural networks have been proposed as an alternative way to generate quantitative structure-activity relationships (69). A commonly used type of neural net contains layers of units with connections between all pairs of units in adjacent layers. Each unit is in a state represented by a real value between 0 and 1. The state of a unit is determined by the state of the units in the previous layer to which it is connected and the strengths of the weights on these connections. A neural net must first be trained to perform the desired task. To do this the network is presented with a set of sample inputs and outputs. Each input is fed along the connections to the nodes in the next layer where they are operated upon and the results fed into the next layer and so on. During the training period the network

adjusts the strengths of the connection until it finds the set of values giving the best agreement between the input and output. Once trained, the net can be used in a predictive fashion.

In QSAR, the inputs correspond to the value of the various parameters and the network is trained to reproduce the experimentally determined activities. Once trained, the activity of an unknown compound can be predicted by presenting the network with the relevant parameter values. Some encouraging results have been reported with neural networks, which have also been applied to a wide range of problems such as predicting the secondary structure of proteins and interpreting NMR spectra. One of their main advantages is the ability to incorporate non-linearity into the model. However, they do present some problems (70); for example, if there is too few data value then the network may simply memorise the data and have no prediction capability. Moreover, it is difficult to assess the importance of the individual terms and the networks can also require a considerable time to train.

Genetic algorithms can also be used to derive QSAR equations (71).

#### **(2.4.4.3) Principal component analysis**

Multiple linear regressions cannot deal with data sets where the variables are highly correlated and/or where the number of the variables exceeds the number of data values. Two methods are widely used to deal with such situations: principal components regression (PCR) and partial least squares (PLS). In principal components regression the variables are subjected to a principal components analysis and then a regression analysis is performed using the first few principal components. It is a more rigorous technique employed for reducing the data and a form of factor analysis, which is a statistical technique used to identify a relatively small number of factors that can be used to represent relationships among sets of interrelated variables. Hence PCR allows the reduction of the dimensionality of problems that involve numerous variables. The original data is recast in a new coordinate system that requires only a few variables to account for most of the

variation in the data. This new coordinate system is determined one component at a time, such that each successive component explains as much of the remaining variation in the data as possible. Generally, the first few components account for the majority of the variability in the data, and later components can often be ignored.

If there is always a high degree of correlation between the  $x$  and the  $y$  values, we could define a new variable,  $z = x + y$ , express most of the variation in the data as the values of this new variable  $z$ . The new variable is called a *principal component*. In general, a principal component is a linear combination of the variables:

$$p_i = \sum_{j=1}^v c_{i,j} x_j$$

$p_i$  is the  $i$ th principal component and  $c_{i,j}$  is the coefficient of the variable  $x_j$ . There are  $v$  such variables. The first principal component of a data set corresponds to that linear combination of the variables which gives the ‘best fit’ straight line through the data when they are plotted in the  $v$ -dimensional space. More specifically, the first principal component maximises the variance in the data so that the data have their greatest ‘spread’ of data values along the first principal component. The second and subsequent principal components account for the maximum variance in the data not already accounted for by previous principal components. Each principal component corresponds to an axis in a  $v$ -dimensional space, and each principal component is orthogonal to all the other principal components. There can clearly be as many principal components as there are dimensions in the original data, and indeed in order to explain all of the variation in the data it is usually necessary to include all the principal components. However, in many cases only a few principal components may be required to explain a significant proportion of the variation in the data. If only one or two principal components can explain most of the data then a graphical representation is possible.

The principal components are calculated using standard matrix techniques (72). The first step is to calculate the variance-covariance matrix. If there are  $s$

observations, each of which contains  $v$  values, then the data set can be represented as a matrix  $\mathbf{D}$  with  $v$  rows and  $s$  columns. The variance-covariance matrix  $\mathbf{Z}$  is:

$$\mathbf{Z} = \mathbf{D}^T \mathbf{D}$$

The eigenvectors of  $\mathbf{Z}$  are the coefficients of the principal components. As  $\mathbf{Z}$  is a square symmetric matrix its eigenvectors will be orthogonal. The eigenvalues and their associated eigenvectors can be obtained by solving the secular equation  $|\mathbf{Z} - \lambda\mathbf{I}| = 0$  or by matrix diagonalisation. The first principal component corresponds to the largest eigenvalue, the second principal component to the second largest eigenvalue, and so on. The  $i$ th principal component accounts for a proportion  $\lambda_i / \sum_{j=1}^v \lambda_j$  of the total variance in the data. The first  $m$  principal components therefore account for  $\sum_{j=1}^m \lambda_j / \sum_{j=1}^v \lambda_j$  of the total variation in the data.

#### (2.4.4.4) Partial least squares

Whilst PCA is a method, which transforms the data matrix into a form that explains the variance in the independent variables in a few components, PLS (partial least squares projections to latent structures) attempts to explain the variance in the observed (dependent) variables (73). Both PLS and PCA generate an orthogonal set of eigenvectors. However, PCA extracts the largest eigenvector first, whereas the first eigenvector extracted by PLS is the one most parallel to the observed data. In the form most commonly applied, PLS derives a linear relationship between the observed data and the field data. Each data element is transformed by a loading or coefficient.

Multiple linear regression requires that the data matrix has many more observed points than data fields, although in QSAR work this is often not the case. PLS is a method, which can overcome this limitation. It is common to carry out a PLS calculation on a few tens of observed values, with data matrices containing

hundreds or even thousands of columns of data, and from this derive a linear relationship between the observed data and the calculated data. It could be argued that with a large number of data fields, it would be possible to arrive at a linear relationship between any observed and calculated data, regardless of whether such a relationship really existed. To overcome this problem the concept of cross-validation has been introduced (74). In cross-validation, observed data and associated calculated fields are systematically ignored during the PLS calculation. The ignored field data is then used in the resultant equation to predict the observed data. The difference between the observed and calculated values is an indication of the predictive power of the final PLS relationship.

The PLS method expresses a dependent variable ( $y$ ) in terms of linear combination of the original independent ( $x$ ) variables as follows:

$$y = b_1t_1 + b_2t_2 + b_3t_3 + \dots b_mt_m$$

where

$$t_1 = c_{11}X_1 + c_{12}X_2 + \dots c_{1p}X_p$$

$$t_2 = c_{21}X_1 + c_{22}X_2 + \dots c_{2p}X_p$$

$$t_m = c_{m1}X_1 + c_{m2}X_2 + \dots c_{mp}X_p$$

$t_1$ ,  $t_2$ , etc. are called latent variables (or components) and are constructed in such a way that they form an orthogonal set. The use of orthogonal linear combinations of the  $x$  values is very similar to principal component analysis. The major difference is that the latent variables in partial least squares are constructed to explain not only the variation in the  $x$  data but also to maximise the degree to which the variation in the observations can be explained.

It is also possible to calculate the degree to which each component explains the variance in each variable, and how far each component explains the variation in the dependent variable.

#### (2.4.4.5) Partial least squares and molecular field analysis

One of the most popular uses of the partial least squares method in molecular modelling and drug design is comparative molecular field analysis (CoMFA), first described by Cramer and co-workers (75). The starting point for a CoMFA analysis is a set of conformations, one for each molecule in the set. Each conformation should be the presumed active structure of the molecule. The conformations must be overlaid in the suggested binding mode. The molecular fields surrounding each molecule are then calculated by placing appropriate probe groups at points on a regular lattice that encompasses the molecule. The results of this analysis can be represented as a matrix,  $S$ , in which each row corresponds to one of the molecules and the columns are the energy values at the grid points. If there are  $N$  points in the grid and  $P$  probe groups are used, then there will be  $N \times P$  such columns. The table is completed by adding an additional column that contains the activity of the molecule.

A correlation between the biological activity and the field values is then determined. The general form of the equation that we desire is:

$$\text{Activity} = C + \sum_{i=1}^N \sum_{j=1}^P c_{ij} S_{ij}$$

$c_{ij}$  is the coefficient for the column in the matrix that corresponds to placing probe group  $j$  at grid point  $i$ . This problem is massively overdetermined as there may be thousands of grid points but often fewer than 30 compounds. However, a successful analysis may often be performed using partial least squares.

The maximum number of latent variables is the smaller of  $x$  values or the number of molecules. However, there is an optimum number of latent variables in the model beyond which the predictive ability of the model does not increase. A PLS model is often evaluated according to its ability to predict the activity of compounds not used to derive the model (rather than how well the model reproduces the activity of the compounds actually used to construct the model). A variety of statistical methods can be used to achieve this. One of the most commonly used techniques is cross-validation, in which the predictive ability of

the model is determined by dividing the data into a number of groups. A series of models are derived by leaving out one of the groups. The omitted data is then used to test the model by calculating the differences between the predicted and actual values. This measures the predictive ability of the model, which is usually reported as the PRESS (predictive residual sum of squares).

$$PRESS = \sum_{i=1}^p (y_{obs} - y_{pre})^2$$

$$SSQ = \sum_{i=1}^p (y_{obs} - y_{ms})^2$$

$$R_s^2 = 1 - \frac{PRESS}{SSQ}$$

$y_{obs}$  is the observed activity for structure  $i$ ,  $y_{pre}$  is the predicted activity for structure  $i$ ,  $y_{ms}$  is the mean observed activity for structures omitted in sample  $s$ ,  $p$  is the number of structure omitted in sample  $s$ . The PRESS value is a measure of the reliability of the predictions made using the regression equation. The last component is then removed and a new set of regression coefficients is calculated and used to predict the observed values, and hence to determine a new PRESS value. This is repeated until the PRESS value increases. The minimum PRESS value and the associated coefficients are saved as the results for the current sample, and the whole process starts again with the next training set. Note that the summed deviations for a single sample, SSQs, depend on the mean observed value for the structures omitted in the sample (if three structures were omitted, SSQs would be calculated from the observed activities of the three structures).  $R_s^2$  is the R squared for cross validation. An alternative way to assess the significance of a model is to randomly reassign the activity, thereby associating the wrong activity with each set of grid values. When this is done then the predictive ability of the model should be significantly better for the true data than for any of the randomised sets.

The CoMFA approach generates a coefficient for each column in the data table. This coefficient indicates the significance of each grid point in explaining the activity. Such data can usually be represented as a three-dimensional surface that connects points having the same coefficients. These diagrams have been used to identify regions where changing the steric bulk would increase or decrease binding.

Since its introduction partial least squares has been widely used to calculate 3D QSARs. These studies have demonstrated the validity and usefulness of the approach but have also highlighted the sensitivity of the approach to several factors. These factors include the selection of the active compounds, the different types of probe that can be employed, the force field models to describe the interaction between the probe and each compound, the size and spacing of points in the grid, and indeed the way in which the PLS analysis is performed. One of the main requirements (and indeed limitations) of the CoMFA technique is that it required the structures of the molecules to be correctly overlaid in what is assumed to be the bioactive conformation (this in turn implies that the compounds have a common binding mode). The first application of CoMFA was to a series of steroid molecules binding to two different targets: human corticosteroid globulins and testosterone binding globulins. In this case the steroid nucleus of each molecule was least square fitted to the nucleus of the most active steroid.

#### **(2.4.5) The protein and ligand interaction**

Evaluation of ligand-receptor interactions is clearly a crucial aspect of selecting compounds. The most fundamental approaches, based on quantum mechanics, have rarely been used for routine studies because of the amount of computational resources required. As computer speed and memory availability continue to improve, *ab initio* calculations are becoming feasible for host-guest problems (76) and may one day extend to appropriately limited ligand-macromolecule interaction (77). The standard methods used for calculating interaction energies use molecular force fields. It is not easy to guess the accuracy of these approximate force-field calculations for molecular interactions in which



one component is a macromolecule. A second approach to evaluating molecular interactions uses heuristic modeling, either independently or in concert with the force-field terms (78). The goal is to represent aspects of the system that might not be well captured by the force-field approximations. For example, hydrogen bonding, strain energy, and the hydrophobic effect are often accounted for in a usual manner. Heuristic methods are especially useful when a complex concept can be approximated by a simple function, as in the use of buried surface area as a measure of the configurational entropy of the solvent (24).

Good quality structures of small molecules are important both as the foundation from which pharmacophore models are created and as the raw material for the database searching. The best starting point is experimental data, when available. The most popular computational approaches are conformation analysis (79-81), although distance geometry and molecular mechanics optimisation (82-83) are also used. An unresolved issue is how extensive a set of ligand conformations should be generated. For database searching, it is frequently assumed that a single low-energy conformation is sufficient. The databases of most interest are the Cambridge Structure File (84); the Available Chemical Directory distributed by MDL Information Systems, San Leandro CA; the National Cancer Institute Database (85); and sections of the Chemical Abstracts Registry. In the last few years, three-dimensional search methods have begun to explore the conformational flexibility of potential ligands. Although storing multiple conformations of a single molecule is conceptually simple, it is limited by the amount of storage space available and by the linear increase in search time with the number of molecules to be searched. It is also impossible to ensure that the conformation being sought will be among those stored, even with systematic search algorithms, because of the chance that the conformation of the bound ligand is not a minimum energy structure in the absence of the receptor. Instead of storing multiple conformations, some approaches encode the distances accessible to pairs of atom centres. These accessible distances are used during the search to select molecules that could match a particular query, and the matching conformations are generated for only those compounds. Other approaches adjust torsion angles to match the query. Flexible searching can significantly increase the amount of computer time required for searching. Because more hits are produced

in a given search, the amount of human time required for screening the results also increases.

A paper includes full flexibility of ligands and limited flexibility of the protein target (86). In searching for a pharmacophore, it is sufficient to determine whether a molecule contains a specified set of distances and angles. Docking methods must go further to explore the orientational and translational space established by the receptor geometry. In searching databases, care is required to balance the number of orientations sampled for each molecule against the compute time available to search all desired molecules. The time required for such searches has been studied by Gschwend and co-worker (87-88). They concluded that the use of a simplex minimizer to adjust orientations improved performance.

#### **(2.4.6) Structure-based ligand design**

De novo design methods present an alternative to database searching and they allow templates selection, lead optimisation, and the design of totally novel molecules. Even though these approaches require a commitment to synthetic chemistry, they can immediately generate novel compounds of a proprietary nature. Finding new leads is a demonstrated strength of database searching methods, but these leads must usually be optimised to improve their activity, increase bioavailability, or reduce toxicity. The needs offer an opportunity to develop computer methods that will guide medicinal chemists in choosing chemical modifications.

Thanks to recent advances in protein isolation and purification as well as in structure determination, information is becoming available on many more protein structures. New techniques in molecular biology and protein biochemistry have made it much easier to obtain large quantities of highly purified protein. Methods in crystallography and nuclear magnetic resonance have also gained rapidly in power. Thus there are many more targets of pharmaceutical interest for which structure information is available.

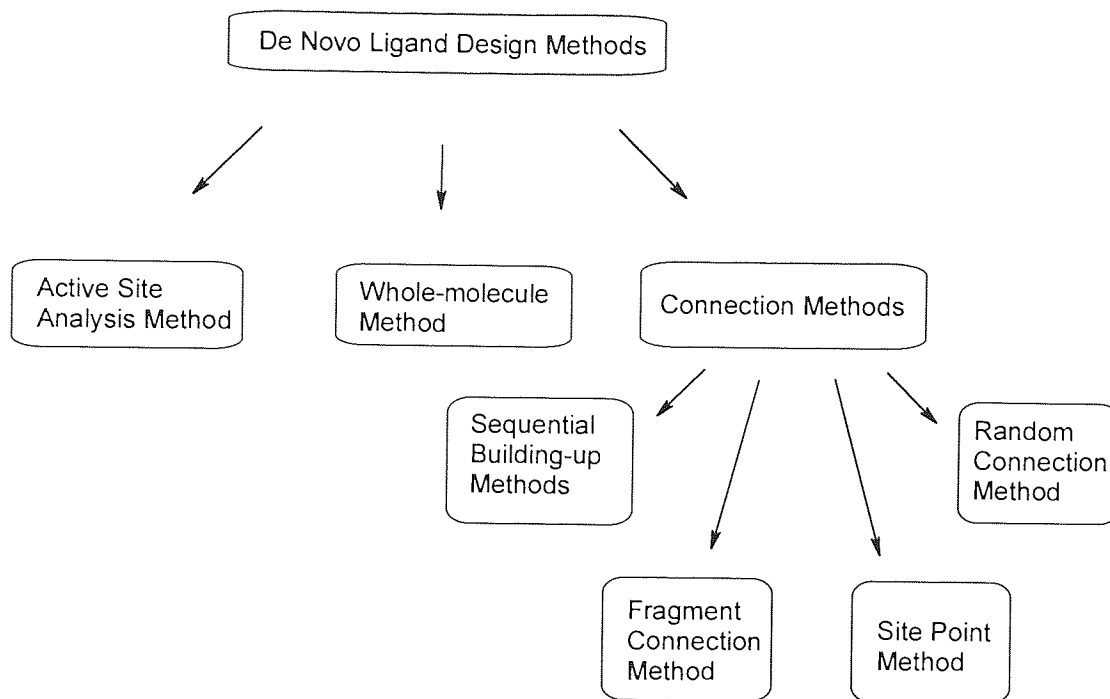
Given this increase in structures available to us either from accurate experiments or by intelligent guesswork, there is plainly a need for novel methods to exploit this information. Specifically, we require a set of computational tools that can analyse receptor sites and suggest compounds that may bind there. This approach is often called *de novo drug design*. Enzymes have been the targets for most reported work of this type because of the availability of three-dimensional structural information. However, because receptor agonists or antagonists have been the subjects of some research, the broader term ligand design is used in preference to inhibitor design. One must remember that computational chemistry is a diverse field. There are many other approaches besides *de novo* ligand design that may be applied to drug discovery. Indeed, *de novo* methods can only be successfully applied if other modelling techniques and topics are thoroughly understood, such as computational searching and analysis, force-field methods, hydrogen bonding, and the hydrophobic effect. At the most fundamental level, all *de novo* design methods can be classified into three types:

#### *Methods that analyse the active site*

Such methods identify which kinds of atoms and functional groups are best able to interact with features at various locations within the active site. The result of the analysis may well be a set of preferred positions for simple fragments, for example water or benzene. Because such a list of fragments is not an actual ligand molecule, some scientists would say that this work does not constitute true *de novo* design, but rather the initial preparation for subsequent *de novo* design.

#### *Methods that dock whole molecules*

These methods attempt to position each ligand molecule, one at a time, in the active site of the receptor, or match it to a pharmacophore model. They may consider conformational flexibility either by allowing the conformation of the ligand to change during the docking process, or by generating multiple conformations for the ligand and rapidly docking each one.



**Figure 1. Classification of ligand design software**

*Methods that connect molecular fragments or atoms together to produce a ligand*

Everyone agrees that this class constitutes true *de novo* design. As shown in **Figure 1**, these methods fall into a few different sub-classes, depending on the approach used:

- I. *Site-point connection methods*: Determine desirable locations for individual atoms (site points); at these locations place suitable fragments which can be joined.
- II. *Fragment connection methods*: Start with previously positioned fragments; without moving them, connect them with suitable "linkers" or "scaffolds".
- III. *Sequential buildup methods*: Construct a ligand atom-by-atom or fragment-by-fragment. Generally a small set of building blocks is used, and the construction process may be random.

IV. *Random connection methods*: This special class of techniques combines some features of the previous three methods, along with strategies for the introduction of randomness into processes such as bond connection.

Some methods do not fall neatly into these classes. As in the example above, it is possible to combine features of more than one method into multistep or hybrid approaches. Furthermore, some methods can be put to alternative uses. For instance, a whole-molecule technique may be used to fit small fragments from a library into optimal locations, and may thus be made to operate as a fragment placement method.

#### **(2.4.7) Pharmacophore identification**

In biological systems where the activity of a small molecule or drug depends on its interaction with a protein receptor, even the identity of this receptor may be unknown, and its full 3D structure is known only in relatively few cases. Rational drug design often must proceed by finding a *pharmacophore*, which has been defined in section (4.1). Usually there are only 3 or 4 atoms involved in a pharmacophore, forming interactions such as hydrogen bonds with the receptor. However, molecules with similar activity do not necessarily interact with the receptor in the same way. Agonists, which induce a physiological response, usually meet this criterion. However, antagonists, which merely block a physiological response, may do this in different ways, sometimes even interacting with different receptors.

Drugs, which react chemically with the receptor pose different problems for modelling than those that remain unchanged. Even though molecules of the former type may contain common functionality, pharmacophore modelling is of little use for them. Peptides are believed to interact with a receptor by often forming a much larger number of hydrogen bonds than is observed with small molecules. Because of this multitude of hydrogen bonds, peptides are usually not worth considering in pharmacophore identification, although small molecule peptide mimics which produce the same physiological response may be used.

### (2.4.8) The protein modelling

Insights into the 3D structure of a protein are of great value both in understanding its function and in the design of drugs to interact with it. Although, as mentioned in section 4.6, protein structure elucidation techniques have advanced rapidly in recent years, they still are often hampered by difficulties such as obtaining sufficient protein and growing crystals that diffract well. The experimental approaches still are involved and often time-consuming. Thus the rate of sequence determination is much higher than the rate of structure elucidation. Currently the Swiss-Prot database (89) contains more than 120,000 sequences, while the Protein Data Bank (PDB) (90) contains about 20,000 entries, not all of which are distinct proteins. In this context, it is obvious why methods to predict the 3D structure of a protein are of increasing interest.

Structure-based drug development relies on the atomic-level structures of the proteins determined by X-ray crystallography and nuclear magnetic resonance (NMR). However, computational methods play important roles in the derivation of the structure from the raw experimental data. While some of the techniques are unique to the processing of crystallographic or NMR data, standard molecular graphics, molecular mechanics, and molecular dynamics methods are heavily used to process the initial solution into the final experimental model. When an experimentally determined structure of the desired protein is not available, the computational methods of protein homology model building, secondary structure prediction, and *de novo* 3D protein structure prediction are often used to create approximate models.

Knowledge of protein structure is only the first step in understanding biochemical processes, which involve multistep dynamical events. Molecular dynamics provides a way to study molecular movement visually. Likewise the design of a molecule that binds well to a protein is only the first step on the way to a drug. High-resolution structures are often unavailable in the early stages of a drug design project, and functionally important residues are usually identified from "blind" site-directed mutagenesis experiments and, to a lesser extent, by structure-activity relationships arising from assays on small molecules and peptides.

Unfortunately, the 3D structure of a protein is determined not only by its amino acid sequence, but also by its environment at the time of folding. The rules governing this process are largely unknown. Thus the prediction of a protein's 3D structure solely from its amino acid sequence in the absence of a structural template (modelling *de novo*) has not yet been reliably achieved. Current efforts generally start with the prediction of secondary structure using well-known techniques which are based on the statistical distribution of amino acids in known 3D structures (91-92) and the fine analysis of multiple sequence alignments (93). Though progress has been made, the prediction of secondary structure is not reliable enough to support the prediction of tertiary structure. The most recent advances in the prediction of protein structure from amino acid sequence alone have utilised the growth in computer power to generate large numbers of random conformations selected by simplified force fields. This approach is expected to provide more reliable predictions of secondary and supersecondary structures, because it seems to work best for stretches of about 50 residues.

## **(2.5). The molecular modelling program**

### **(2.5.1) Overviews**

With the rapid development of computer hardware and networks, displaying molecular structure and protein modelling are both becoming available to ever more scientists. Besides the well-established commercial software packages from suppliers such as Tripos and MDL, which generally require powerful workstations, several systems such as Chem-X allow protein modelling and molecular visualization on personal computers. Such tools are enabling experimentalists to visualize their structures and perform simple modelling tasks without heavy investments. They are becoming a fundamental component of modern biomedical research.

Now that suitable algorithms are available, molecular design programs can search databases of 3D structures (94). Examples include ALADDIN, CAVEAT, and DOCK. The ALADDIN program (95) uses the concept of a pharmacophore to search a database for structures that match its criteria. The CAVEAT program

(96) uses vectors to create templates from molecular fragments to hold conformationally flexible molecules in a specified conformation. The DOCK program (97) is based on the idea of molecular shape. A binding site cavity is used to search a database for molecules, which have complementary shape.

Programs that search databases of 3D structures suffer from two main restrictions. Firstly, often a single conformation is stored for each structure, usually the one believed to be lowest in energy. However, it is known that drugs often bind in conformations other than the global minimum, and hence useful structures can be missed. The second restriction arises from the limited content of the databases. When the ALADDIN program was used to search different databases, there was very little overlap between the results, implying that only a fraction of the structure space had been investigated (98).

It remains very difficult to design structurally novel compounds. Lewis and Dean (99) first developed the concept of spacer skeletons to approach this problem. These topological devices can model more than one structure at a time. Initially their use was limited to planar ring systems, but they have been extended into 3D by using a diamond lattice as a spacer skeleton (100). This makes it possible to design acyclic structures, but it is still not a general solution since each atom in the lattice is  $sp^3$  hybridized and all torsion angles are staggered. Lewis (101) has described a new algorithm providing diversity in the atom chains it creates, by solving a series of trigonometric equations within geometric constraints appropriate for a given set of atom types. This method is intended to provide the necessary bridges between fragments that have been placed at the important interaction sites.

A number of other programs are also intended for the design of novel molecules using knowledge of the 3D structure of the target enzyme (102-107). Nishibata and Itai (102) described a program called LEGEND, which generates structures one at a time. Instead of an exhaustive procedure, the method involves the use of random numbers at every stage of the process, i.e. to select an anchor point for the first atom, to select a root atom each time for extending the structure by attaching a new atom, to determine the type of the atom and bond, and to determine torsion angles. GROW (103) is a program that generates peptides by connecting small molecular fragments. A large set of amino acid fragments can be used as templates. Each amino acid is considered in several conformations.



The search space is managed by a tree representation that is pruned according to an energy function evaluated by molecular mechanics. The user specifies the number of successor structures to be included in the tree each time an additional amino acid is added to extend the peptide. The main limitation of this program is its restriction to peptide-like compounds. The LUDI program (104) first recognizes interaction sites within the receptor site. Molecular fragments are fitted onto these interaction sites and finally connected by bridging fragments. The program BUILDER (105) produces novel structures by combining database searching, structure generation algorithms and interactive graphics modelling. An initial database search is performed with the DOCK program to find structures that fit the binding site sterically. The retrieved structures are then superimposed within this site, and the vertices from different molecules are linked by virtual bonds to produce a lattice. The user intervenes to specify the regions of interest, the appropriate parts of the lattice are displayed graphically, and an attempt is made to join fragments via paths through the lattice.

## **(2.5.2) The SPROUT Program**

### **(2.5.2.1) Program overview**

The aim of the SPROUT project is to build a general purpose program for the design of molecules appropriate to a wide range of applications, e.g., inhibitor design, the design of catalysts (particularly synthetic enzymes) and the design of agents for asymmetric synthesis. All of these systems rely on molecular recognition where molecules interact because they are complementary to each other. This complementarity arises from the steric and electrostatic properties of the molecules, for example, an inhibitor is able to bind to an enzyme because it has complementary shape and electrostatic properties to a binding site on the receptor. SPROUT uses information about one molecule to constrain the design of other molecules with which it can interact. The constraints should be sufficient to determine the nature of the molecules. However, full structural information is not always available and sometimes must be inferred. For example, whereas there are a growing number of enzymes whose 3D structures are available to the design of

enzyme inhibitors, in other cases the structure of the enzyme receptor site can be inferred by overlaying sets of active and inactive compounds (108).

The major interactions involved in the association of molecules are steric, electrostatic (including hydrogen bonding), dispersion or van der Waals, and hydrophobic (109). These interactions give rise to primary and secondary constraints for molecule design. The primary constraints are steric in nature and are dependent on the particular application and the information that is available, for example, in inhibitor design the 3D coordinates of the receptor may not be available. The secondary constraints arise from the electrostatic and hydrophobic properties of the known molecule.

Lewis and Dean (99) used this division of constraints to divide structure generation into two phases: *primary structure* and *secondary structure generation*. SPROUT also consists of two phases. Primary structure generation is defined, as the process of generating a 3D molecular graph consistent with the primary constraints on the system, i.e., the shape of the receptor site. Secondary structure generation is the process of converting the molecular graph into a molecular structure, i.e., the vertices and edges of the graph are replaced by atoms and bond appropriately to give the molecule the desired functionality. The secondary structure generation phase makes use of the secondary constraints, e.g., electrostatic and hydrophobic properties.

The primary constraints include the 3D shape of the receptor, which defines the volume the molecule must lie within. The volume is enclosed by a *boundary*, which then restricts the shape of potential ligands. The volume is fixed, for example, in the case of an enzyme the receptor site is assumed to be rigid. Within this volume there are *interaction sites*. These are regions which if occupied by an atom of the ligand can lead to favourable interactions between the ligand and the receptor for example, electrostatic or hydrophobic interactions. If the interaction sites are sufficiently localised they form part of the primary constraints and are used to direct primary structure generation. This is the case for hydrogen bond interactions where it is possible to identify regions that are small enough to be occupied by a single atom. These localised interaction sites are called *target sites* and satisfying these interaction site forms a requirement for primary structure generation. Other interaction, such as hydrophobic and van der Waals interactions,

are less directional and lead to more diffuse interaction regions within the cavity. These interactions are the secondary constraints on structure generation.

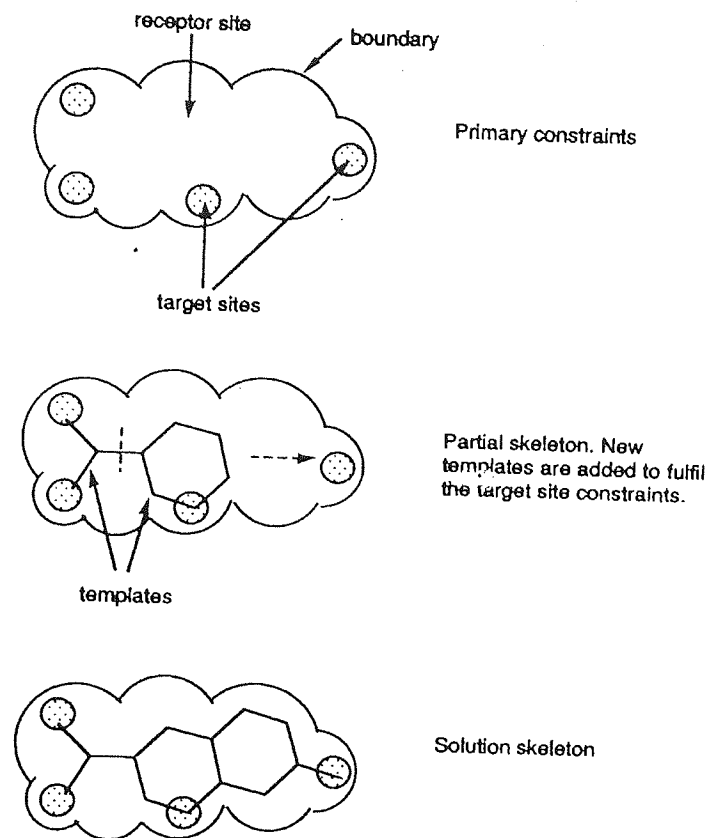
During primary structure generation atoms are placed at each of the target sites and linked to produce molecules, without violating the volume specified by the boundary. This type of structure generation is well known as a combinatorial problem (110) where attempts at finding solutions rapidly produce a large number of intermediate structures. Therefore, methods of controlling this combinatorial explosion are required.

One way of reducing the number of intermediate structures is to use molecular fragments as building blocks. This allows larger distances to be spanned in single step than if molecules are built one atom at a time. However, an enormous number of fragments are required to enable a wide range of molecules to be produced. Primary structure generation is mainly concerned with shape. This makes it possible to group molecular fragments by shape and connectivity so that initially structures are generated using a representative from each group. The representative fragment can be replaced by other fragments from the same group, once structures have been found that match the steric constraints. *Templates* are used as representative fragments. Templates are 3D sub graphs where the vertices represent atoms and are labelled only by their hybridisation state (and not element type). The edges of the sub graphs represent bonds. The vertices are labelled as  $sp^3$ ,  $sp^2$  or  $sp$  and have tetrahedral, trigonal or planar geometry, respectively. The geometry defines the positions where new templates can be joined. The bonds are labelled as single, double, triple or aromatic and the distances between the vertices are the corresponding carbon-carbon bond lengths. A number of molecular fragments can be produced from each template by replacing the vertices by any element that can adopt the appropriate hybridisation state. Bond angles and bond lengths are also adjusted appropriately. This adjustment results in only minor differences in shape between a molecular fragment and its representative template.

The primary structure generation phase joins templates together to produce *skeletons*. (This process is also called *skeleton generation*). Each skeleton represents a number of molecules because each component template represents a number of molecular fragments. Each molecule that can be produced from a skeleton adopts approximately the same shape as the skeleton from which it is derived. Therefore, these molecules satisfy the primary constraints, i.e., they have

the required shape. A skeleton that does not satisfy all the requirements is called a *partial skeleton*.

Skeleton generation begins by selecting a template and positioning it at one of the target sites thus satisfying one of the steric requirements. One vertex of the template is anchored at the centre of the target site. The template can be rotated about the anchoring vertex to occupy any position. A representative set of orientations is chosen and each orientation gives rise to a partial skeleton. New templates are added to build skeletons of increasing size. A skeleton is grown in the direction of the remaining target sites. A solution has been found when all the steric requirements are satisfied and no boundary violations have occurred. This process is illustrated in **Figure 2**. (143)



**Figure 2. The program's procession**

Templates are divided into cyclic and acyclic templates. Acyclic templates can be joined to cyclic or acyclic templates by forming a new bond between one vertex from each template. This type of join is called the *new bond* join. Rotation is possible about a new bond and so a number of conformations are produced. The current program produces the staggered conformations when two  $sp^3$  atom are joined by this method. Two cyclic templates can be joined by fusion, bridging, spiro joining or by forming a new bond between them. The secondary structure generation phase fulfils the requirements made by the secondary constraints, by making atom substitutions to produce the required electrostatic and hydrophobic properties.

*The main components of the program are identified as:*

- (1) Representation of the primary constraints, i.e., the steric constraints.
- (2) Construction of the library of 3D molecular fragment or templates.
- (3) Methods for joining templates into larger approximate structures, which are consistent with the primary constraints.
- (4) A strategy for controlling the combinatorial explosion inherent in structure generation.
- (5) Representation of the secondary constraints, e.g., electrostatic and hydrophobic effects.
- (6) Atom substitutions to convert the approximate structures into molecules that are consistent with the secondary constraints.
- (7) Organising and evaluating the resulting molecules, e.g., by using conformational analysis and clustering techniques.

The program has been applied to the design of enzyme inhibitors where the constraints can be derived from an enzyme-binding region. The primary constraints are 3D boundary together with target sites within the cavity. These constraints are extracted from the X-ray data of enzymes that have been crystallised, from NMR experiments, or they are inferred by overlaying sets of active and inactive compounds to determine a pharmacophore.

#### **(2.5.2.2) Primary constraints**

##### Target sites

The target sites used as primary constraints are derived from localised regions of the cavity where an atom of the inhibitor must be placed to interact with the receptor. This interaction is usually a hydrogen bond. In skeleton generation, this cavity region is represented by part of the volume, called a target site, which must be occupied by a vertex. Accurate hydrogen bonding geometries can be found if the receptor is assumed to be rigid and the receptor hydrogen positions are known (104, 111-112). In the present work, the target sites are represented by larger volumes to allow for the differences between skeletons and molecules mentioned previously. They are represented by spheres of fixed radius; a value of 0.5 Å is

used for this radius in the present work. A target site becomes *satisfied* when exactly one vertex of the skeleton falls within its associated sphere.

Approximating the target sites to spheres allows some freedom in the position of a skeleton: the position of a partial skeleton can be altered without losing the correspondence between vertices and target sites. For example, if only one target site is satisfied the partial skeleton can be moved by any distance less than the radius of the target site sphere in any direction without losing the correspondence between the anchoring vertex and the target site. Moving a partial skeleton may allow a previously unsatisfied site to become satisfied. Whereas it is impractical to explore all the possible displacements available to a skeleton each time a new template is added it is worthwhile in some situations. These situations are identified by examining the relationship between a skeleton and the target sites each time when a new template is added. A number of relationships can exist between a skeleton and a target site:

- (I) The skeleton satisfies the target site, i.e., one of its vertices lies inside the sphere representing the target site.
- (II) The skeleton does not satisfy the target site but is close enough to it to prevent any successor from satisfying the target site. If a vertex is within half a bond length from the centre of the site it will prevent any successor from satisfying the site. If it is more than half a bond length away adding a new template to the skeleton can result in a new vertex being closer to the site or satisfying the site.
- (III) The distance between the skeleton and the target site is too large for either (I) or (II) to apply. Templates are added until (I) or (II) occurs.

Cases (I) and (III) do not require any special handling. Case (II) can result in the loss of potential solutions. The procedure for avoiding this situation is as follows: an additional sphere is included around each target site of radius equal to half an average bond length. The area inside this sphere and outside the target site sphere is called the *close region*. Wherever a vertex is found in a close region then this skeleton is the best that can be achieved for that arrangement of templates. An

attempt is now made to adjust the position of the skeleton to satisfy this site. If this fails the skeleton is discarded.

Three types of adjustment are possible: displacement of the skeleton as a rigid structure, rotation of the skeleton as a rigid structure about an axis, and internal changes to the skeleton (changing bond lengths and /or torsion and bond angles). Any combination of these transformations could be applied to the skeleton.

However, rotations are not considered because a small rotation can have a large effect on the position of the skeleton and it is difficult to maintain existing correspondences between target site and vertices (and ensure that the boundary is not violated). Internal changes are not considered at this stage. Therefore, the problem is reduced to that of finding a displacement vector that will move the vertex in the close region into the target site sphere but still maintain all existing correspondences

### **(3). Biological functions and chemical synthesis of Cyclothialidine**

#### **(3.1) Topoisomerases**

Topoisomerases form enzyme-bridged breaks in DNA strands which allow other strands to pass through. An example of a topoisomerase mechanism is the nicking of one strand, allowing rotation about the intact strand, after which the nick can be sealed. Each cycle of nicking and sealing releases one superhelical turn.

In addition to relaxing supercoils in DNA, topoisomerases can produce other topological changes, including the catenation (linking) and decatenation (unlinking) of DNA helices. Near the end of a round of replication of a circular piece of DNA, two interlinked circles have been generated. It is therefore expected that topoisomerase action is necessary to bring the replication of circular DNA to a satisfactory conclusion. Topoisomerases can also perform knotting and unknotting of strands. These essential enzymes are divided into two types according to their mechanism.



### **(3.1.1) Topoisomerases I**

Topoisomerases I reversibly break one strand of the double helix. The enzyme remains covalently bound to the 5' end of the broken strand while the 3' end moves aside to allow the unbroken strand in effect to pass through the gap. The hydroxy group of the 3' end then attacks the activated, covalently bound phosphate group at the 5' end, thus resealing the nick with the relief of one supercoil.

This process does not need adenosine triphosphate (ATP) because the topoisomerase enzyme is able to store and reuse the energy involved in cleaving and remaking the phosphodiester bond. Prokaryotic topoisomerase I requires  $Mg^{2+}$  and relaxes only negative supercoils, while the eukaryotic enzyme does not require  $Mg^{2+}$  and can relax both positive and negative supercoils. Both enzymes are monomeric and can bring about the catenation of two circular DNA molecules only if one is already nicked (113).

### **(3.1.2) Topoisomerases II**

Topoisomerase II enzymes create a double-stranded break in one duplex through which a section of unbroken duplex passes before the break is resealed. DNA wraps around the outside of the enzyme, which catalyses the breakage of both strands of one loop. Another loop of DNA crosses through the break, the break is then resealed and the DNA released (114). The result is the introduction of two negative supercoils and a decrease of the linking number ( $L$ ) by two.

Prokaryotic topoisomerase II is able both to relax positive supercoils and to introduce negative ones. The enzyme is a tetrameric complex of two proteins A and B, taking the form  $A_2B_2$ . Eukaryotic topoisomerase II is a homodimer, not a tetramer, and while it can relax both positive and negative supercoils, it has no supercoiling ability. Topoisomerases II from both cell types require ATP and are able to knot/un knot DNA strands and catenate/decatenate DNA circles.

### (3.1.3) DNA gyrase

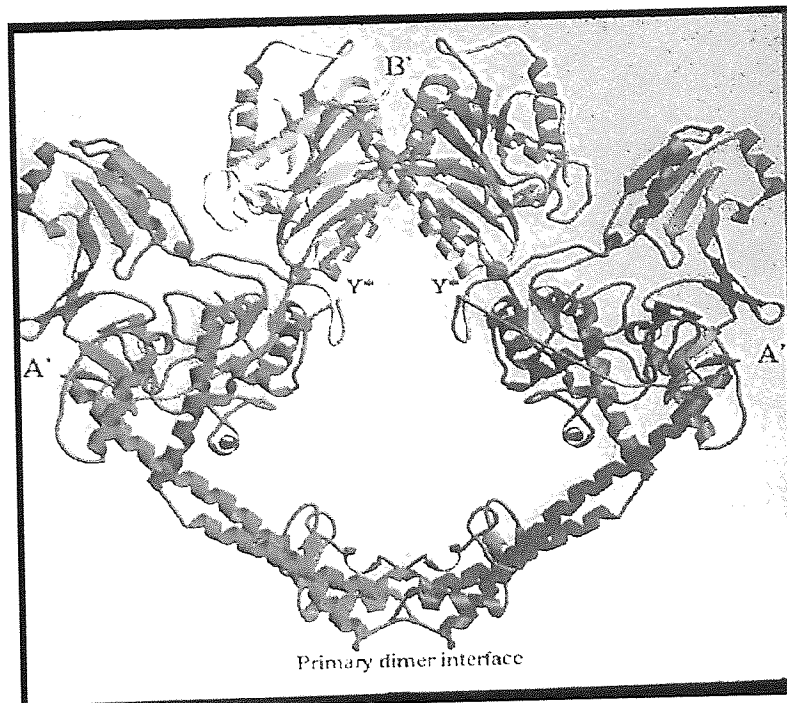
DNA gyrase is a type II topoisomerase found only in prokaryotic cells. It is essential for bacterial growth. It exists as a complex of two proteins, designated A and B, in the form of a tetramer  $A_2B_2$ . Protein A, coded for by the *gyr A* gene, consists of 875 amino acids and has a molar mass of 97,000. Protein B, coded for by the *gyr B* gene, contains 804 amino acids with a molar mass of 90,000 (115).

Gyrase has the unique ability to convert relaxed circular DNA into a negative superhelix. The supercoiling facilitates DNA replication. One DNA gyrase molecule can introduce about 100 supercoils per minute. This conversion requires energy, which is supplied by the hydrolysis of ATP. If ATP is absent, gyrase still interacts with DNA but relaxes negative supercoiling instead of introducing it. However, the rate of supercoil relaxation is more than 10 times slower than that of supercoil introduction (116).

Current models (117) for the mechanism of supercoiling assume that a DNA strand passes through a double-stranded break which is held open by the protein. DNA is believed to pass through a gateway in the protein, and X-ray studies have revealed a hole 20 Å in diameter, which is a reasonable approximation to the diameter of the DNA helix.

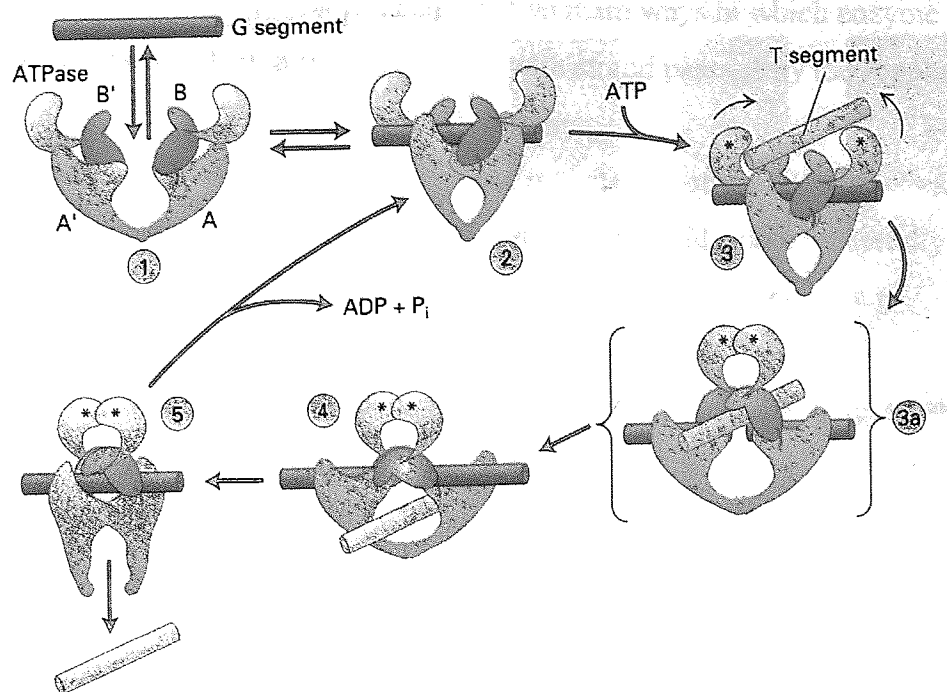
Biological studies show that gyrase subunits A and B work together to introduce supercoils. First, gyrase A binds to DNA, leading to cleavage of the DNA strands. The protein then undergoes a large conformational change to allow another strand to pass through the break into the interior of the protein complex. To stabilise an otherwise unfavourable conformation of the protein, it is thought that binding energy from the association of ATP and the B protein is used. The break is resealed and binding energy is released when ATP undergoes hydrolysis to ADP and inorganic phosphate. These products dissociate from the protein allowing relaxation back to the original conformation at the start of the cycle.

The  $A_2B_2$  heterotetramer structure and the 20 Å hole are clearly visible in **Figure 3** (118).



**Figure 3. DNA gyrase's structure**

It is apparent that the A protein can interact with DNA and is responsible for breaking and rejoining it, while the B subunit processes ATP by itself, with a weak ATPase activity (119). The ATPase-catalysed hydrolysis of ATP is essential for activity. With the aid of this structural knowledge the mechanism conceptually described above can be given a structural interpretation (**Figure 4**) (120).



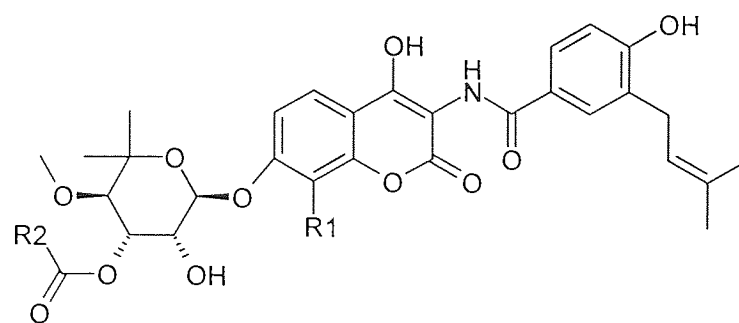
**Figure 4. The activity of *E. coli* DNA gyrase**

The steps in this cartoon representation are:

1. The DNA gyrase firstly binds to the G segment of the DNA. This binding changes the conformation of both subunits.
2. A break is made in the G segment, followed by binding of ATP to its binding domain and movement of the T segment towards the centre of these domains.
3. The ATP-binding domains approach each other, facilitating the movement of the T segment through the break in the G segment.
4. Resealing of the G segment then occurs.
5. The A and A' subunits then re-form to the ATP required state.

Because of their important biological functions, topoisomerases have attracted interest as therapeutic targets for curing bacterial infection and cancer. Topoisomerase poisons interrupt the enzyme mechanism part way through and thereby convert essential enzymes into intracellular toxins which kill proliferating





novobiocin : R1=Me, R2=NH<sub>2</sub>

clorobiocin : R1=Cl, R2 = 5-methylpyrole

**Figure 5. Structures of Coumarine**

Early studies suggested that the coumarins were simple competitive inhibitors. However, recent work indicates a separate, though overlapping, drug binding site which results in the stabilisation of a conformation which cannot bind nucleoside triphosphate, thus preventing ATP hydrolysis. However, the coumarins have found little pharmaceutical use for a variety of reasons, which include poor oral absorption, low activity against some Gram-negative bacteria, and a tendency for initially coumarin-sensitive organisms to acquire novobiocin resistance easily. One mechanism for such resistance is the ability of some prokaryotes to modify novobiocin chemically, transferring the carbamoyl group from position 3' to position 2' of the sugar moiety (121).

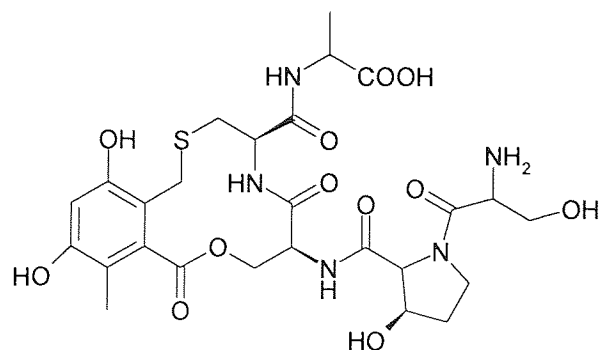
### **(3.2) Cyclothialidine, an inhibitor of DNA gyrase**

Cyclothialidine, a potent DNA gyrase inhibitor, was isolated from *Streptomyces filipinensis* NR 0484 (122). It is the first representative of a unique new class of natural products, containing a 12-membered lactone ring that is fused

to a highly substituted benzene ring, and partly integrated into a pentapeptide chain that includes the rare amino acid cis-3-hydroxy-L-proline (123).

Cyclothialidine exhibited the highest observed inhibitory activity against DNA gyrases from several important bacterial species, including *Escherichia coli* and *Staphylococcus aureus*, with high selectivity in its biological activity. Previous studies (125) indicated that under steady-state conditions cyclothialidine inhibits the ATPase activity of *E. coli* gyrase competitively with a  $K_i$  of action that stands comparison with the coumarin antibiotics novobiocin and coumermycin A<sub>1</sub>. However, cyclothialidine remained active against a DNA gyrase resistant to novobiocin, suggesting that the residues required for novobiocin binding are not required by cyclothialidine. Biological studies (124) indicate that cyclothialidine does not inhibit strand cutting, strand rejoining or DNA binding. Instead it inhibits ATP hydrolysis, a mode of action that is similar to the coumarins. In 1995 Naoki Nakada et al. (125) suggested that cyclothialidine, ATP and the coumarins showed some overlap of their binding sites but also some differences. It was proposed that cyclothialidine binding occurred close to the ATP binding site of the gyrase B subunit, stabilising a conformation of the protein that did not permit ATP binding, and it recognised a site different to that of the coumarins. A year later, the first crystal structures of a topoisomerase / drug complex were published, proving that binding was indeed competitive because of a small degree of overlap between the binding sites. The overlapping regions of the binding sites were found to interact with the resorcinol ring of cyclothialidine and the adenine ring of ATP. This relatively small amount of overlap is enough for cyclothialidine, once bound, to act as a plug preventing ATP binding. This explains how three very different structures can compete with each other.

The structure of cyclothialidine (**Figure 6**) was first identified by a combination of spectroscopic methods and amino-acid analysis.

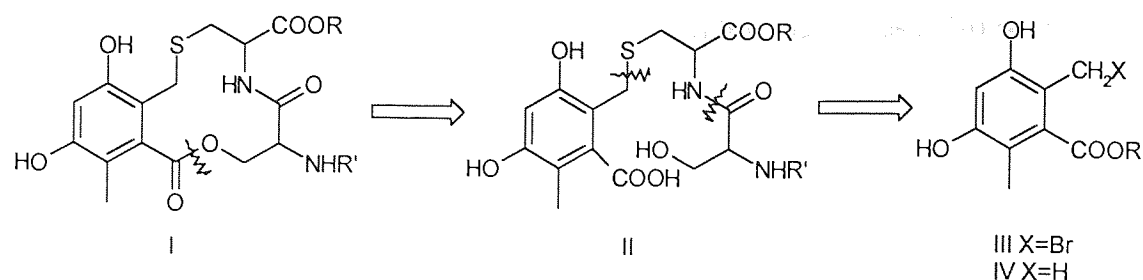


**Figure 6. The Structure of Cyclothialidine**

The total synthesis of cyclothialidine was carried out at Hoffmann-LaRoche Ltd. in 1996 (126). Although it exhibits hardly any antibacterial action, probably due to inadequate penetration through the cytoplasmic membrane, cyclothialidine was considered to be a potential lead compound which could be chemically modified to create a new class of antibacterials. To explore and exploit this potential, the researchers at Hoffmann LaRoche developed an efficient and flexible synthetic route that made it feasible to prepare a great variety of analogues.

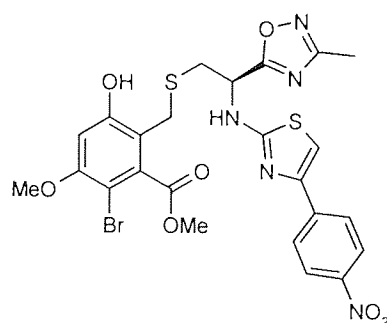
The retrosynthetic analysis (**Figure 7**) suggested that the bicyclo lactone **I** would be a versatile key intermediate. To form the macrocycle, a classical lactonization of an  $\omega$ -hydroxy acid **II** seemed to be the best method. Disconnection of the peptide chain from the seco-acid **II**, either stepwise or all at once, would thereupon lead back to a benzylating agent, such as a bromide **III** which could be derived from a 3,5-dihydroxy-2,6-dimethylbenzoate **IV**. Based on this analysis, they brought into existence a synthetic scheme, which has demonstrated its flexibility in many different syntheses (126).





**Figure 7. Retrosynthetic Analysis**

Very recently, researchers at Bayer (127) synthesized a series of seco-cyclothialidines. Their investigation showed that the lactone ring is not mandatory for activity against the target, encouraging the development of open chain analogues. In fact, these seco-cyclothialidines were able to penetrate bacterial membranes. The lead compound Ro-61-6653, already patented by Hoffmann-LaRoche (**Figure 8**), exerted excellent activity against Gram-positive bacteria but was inactive against Gram-negative strains. These investigations provided the first reports of the *in vivo* antibacterial potential of the seco-cyclothialidines in animal infection models.



**Figure 8. Ro-61-6653**

This analysis points us in the direction to design and synthesize further open chain analogues, which are smaller and simpler for both synthesis and modelling.

In this thesis, we first analysed the 20 Cyclothialidine analogues (126) with the TSAR program to generate an equation between the substituents' parameters and the correlated activity. The synthesis was performed according to this equation to synthesize the new analogues. The MIC was tested. However, the result had failed to prove the equation. Then the structure based drug design program SPROUT, was applied to design the open chained Cyclothialidine analogues. Since the X-ray structure of ligand Cyclothialidine interacting with DNA gyrase was not available to the public, we can only use the ligand unknown module in the SPROUT program, which causes many difficulties in the research. The MIC was tested and from the improved analogues, a 3D pharmacophore was generated to the DNA gyrase.

## **Chapter 2. Methods**

### (1). Cyclothialidine synthesis route

In the organic synthesis scheme (**Figure 9**) we used 3,5 – dihydroxybenzoic acid **1** as the starting material, treated by the Mannich reagent to obtain **2**, which was hydrogenated in the presence of palladium in carbon to afford **3**. The Mannich reaction and Mannich base reduction were repeated and we obtained the 3,5-dihydroxy-2, 6-dimethyl benzoic acid **5**. The acid **5** was then esterified by 1,1,3,3-tetramethylguanidine with 4-nitrobenzyl bromide in DMF. The use of the guanidine as base can provide less of the unwanted phenol ether. The two phenol groups of ester **6** were protected by tert-butyldimethylsilyl chloride, which is abbreviated as TBDMS, then the symmetrical 2,6-dimethylbenzoate **7** was brominated using NBS in refluxing carbon tetrachloride under irradiation with light. This bromination reaction provided a 4:1:1 mixture of the desired monobromide **8**, symmetrical dibromide and the starting material **7** (126). It seems to have no significant effects to carry on the reaction route without purification of this mixture (126). On the other hand, the two amino acids cysteine ethyl ester **9** and Boc-serine **10** were coupled using DCC method without influencing the thiol group. The **8** and **11** were then mixed with triethylamine in DCM and the sulfide was easily isolated by column chromatography to afford **12**. The hydrogenolytic cleavage of **12** provided the  $\omega$ -hydroxy acid **13**. For the lactonization of the **13**, the Mitsunobu reaction provided the best cyclization results **14**. After cleavage of the protecting group, this compound was regarded as the leading compound.

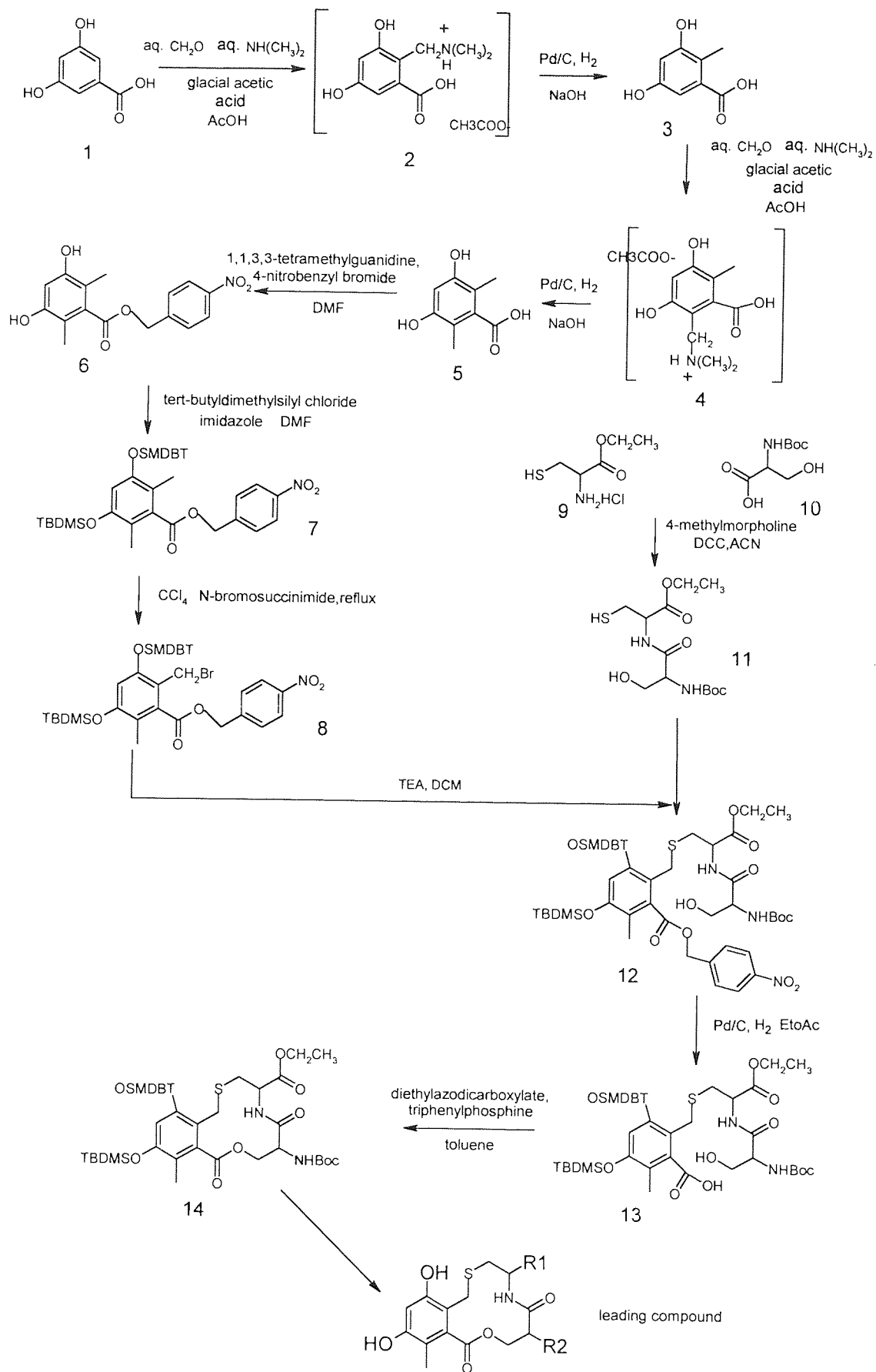
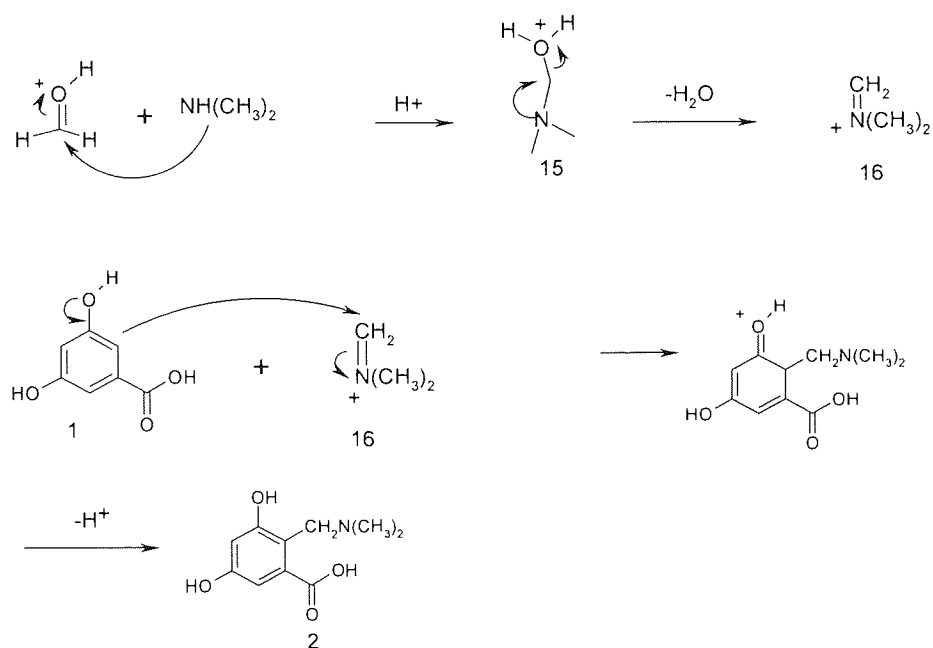


Figure 9. The organic synthesis scheme

## (2). The Mannich reaction and Mannich base reduction

### (2.1) The Mannich reaction

The first reaction is to add two required methyl groups onto the aromatic ring, on which the hydroxyl groups are already present. The activating nature of the hydroxyl groups was expected to ensure efficient electrophilic aromatic substitution. In the acid catalysed Mannich reaction, an aldehyde (usually formaldehyde) is condensed with an amine (eg. dimethylamine) to produce an intermediate such as **15**, which loses water to form iminium ion **16**. A nucleophilic substitution reaction on **16** by a compound containing an active hydrogen, such as 3,5-dihydroxybenzoic acid **1**, affords a 'Mannich base'.

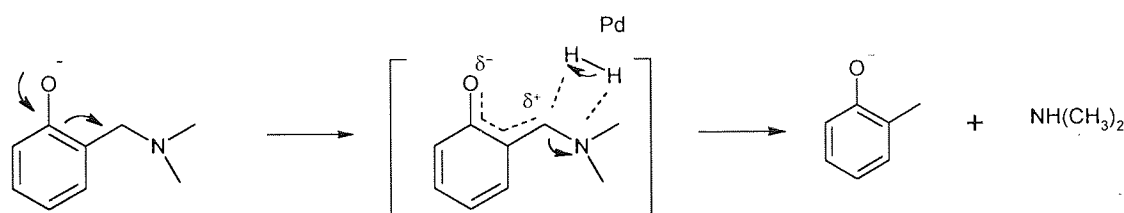


Aqueous dimethylamine was added dropwise, with cooling, to a stirred mixture of aqueous formaldehyde, ethanol and glacial acetic acid. Stirring was continued for 30 minutes, when the 3,5-dihydroxybenzoic acid was added. After stirring overnight at room temperature, the resulting white precipitate was isolated by filtration to afford the desired salt.

## (2.2) The Mannich base reduction

In order to preserve sensitive groups it was decided to perform the reduction by hydrogenation. To ensure good yields, thorough mixing of the salt, catalyst and hydrogen gas was vital. The mixture was saturated with hydrogen via a balloon, and a large flask was used to contain a small volume of methanol with 3M NaOH, ensuring adequate space for mixing was available. Surprisingly, carrying out the reaction under pressure did little to increase yield.

Under basic conditions, the phenolic proton is abstracted and the resulting  $\delta^+$  carbon atom readily attacked by  $\text{H}^-$  to produce a methyl group and liberate the amine.



Treatment of a suspension of acetate salt **2** in methanol with hydrogen and palladium/carbon catalyst was to afford 3,5-dihydroxy-2-methylbenzoic acid **3** as an orange powder

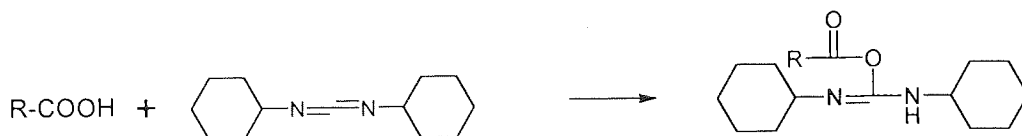
## (3). Peptide synthesis

### (3.1) Coupling Reagents

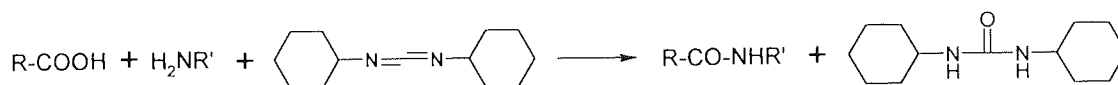
#### (3.1.1) DCC (dicyclohexylcarbodiimide) reagent

In 1955, dicyclohexylcarbodiimide (DCC or DCCI) was proposed as a reagent, which can effect the formation of peptide bonds, by Sheeha and Hess (128). It is widely used as the most successful coupling reagent so far. In this case, activation

of the carboxyl group occurs through its addition to an  $N = C$  double bond in carbodiimides:



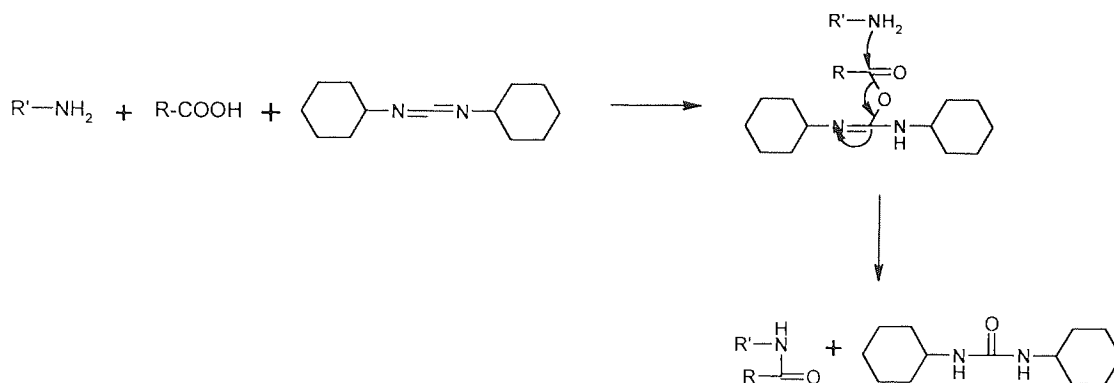
A characteristic feature of this procedure is the application of the carboxyl-activating compound in the presence of the amino-component. Condensing agents which can be added to a mixture of both components are more numerous by now and are called *coupling reagents*. An obvious prerequisite of a coupling reagent is inertness toward primary and secondary amines. This requirement is not completely fulfilled in the case of carbodiimide, which can combine with amines to give guanidine derivatives. But under the usual conditions of peptide synthesis, this reaction is too slow to compete with the rapid addition of the carboxyl group. Therefore, carbodiimides can be used as coupling reagents and are usually added to the mixture of carboxyl and amino-components:



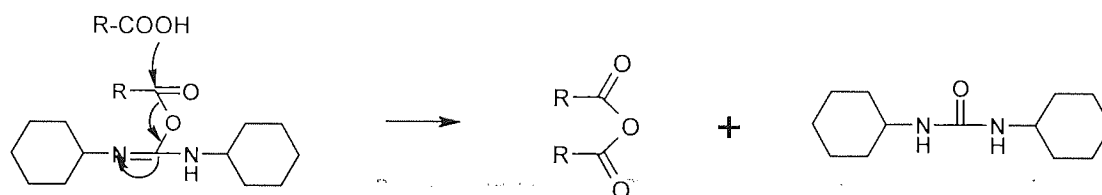
The by-product of the coupling reactions, *N, N'*-dicyclohexylurea is readily removed from the reaction mixtures by extraction and column chromatography.

There are two main mechanisms in this DCC method. One is the nucleophilic attack of the amino-component on the *O*-acylisourea intermediate:





An alternative mechanism, however, is equally important. Reaction of the *O*-acylisourea with unreacted carboxyl-component yields a symmetrical anhydride, a potent acylating agent.



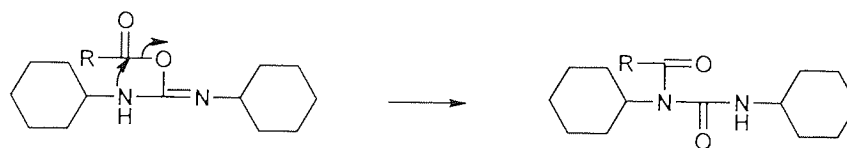
If the carboxyl-component is a protected amino acid, it is possible and often desirable to use two moles of carboxylic acid for one mole of DCC. Such a ratio will, of course, favour the formation of symmetrical anhydrides and diminish the concentration of the *O*-acylisourea derivative. In turn, the extent of *O*  $\rightarrow$  *N* acyl migration and, consequently, the production of ureide is suppressed. For an even more perfect execution of the coupling reaction, symmetrical anhydrides can also be isolated (129), but in this case, DCC is used as an activating agent and not as a coupling reagent. In the practical procedure of peptide synthesis, carbodiimides are mostly applied without isolation of the symmetrical anhydride intermediates. Acylation with a 2:1 mixture of the protected amino acid and carbodiimide is often referred to as acylation with symmetrical anhydride. This designation is not fully warranted, since participation of some *O*-acylisourea in the acylation is still likely. The assumption that symmetrical anhydrides are the reactive intermediates mainly responsible for the coupling is perhaps more justified in the so-called “preactivation” approach, where the protected amino acid (2 moles) and DCC (1

mole) are allowed to react with each other before the mixture is brought into contact with the amino-component. Yet, the reformation and possible loss of one mole of the protected amino acid or protected peptide, which necessarily occurs in this reaction, seemed unacceptable for a long time. When Boc-amino acids were already commercially available and were becoming less and less expensive, the loss of a part of the acylating agent caused no more concern. The successful application of symmetrical anhydrides and particularly their widely accepted use in solid-phase peptide synthesis led to further improvements in their preparation.

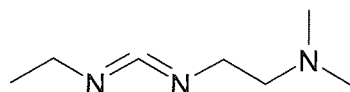
Activation by DCC might cause racemization of the carboxy-terminal residue, but the careful operation can minimise the risk. The DCC reaction, which can be carried out in a large variety of solvents, proceeds rapidly and yields the desired product without fail.

### (3.1.2) 1-(3-Dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC) reagent.

There are several shortcomings of the DCC method. The  $N, N'$ -dicyclohexylurea by-product, while indeed insoluble in most organic solvents (except in alcohols) and thus removable by filtration, is not entirely insoluble, particularly in the presence of other dissolved materials and therefore it frequently contaminates the product of coupling steps. A more disturbing side reaction is the intramolecular rearrangement of the  $O$ -acyl isourea derivative. The attack of the activated carbonyl group by the nearby nucleophile (NH) results in an  $O \rightarrow N$  shift yielding an  $N$ -acylurea derivative as by-product.

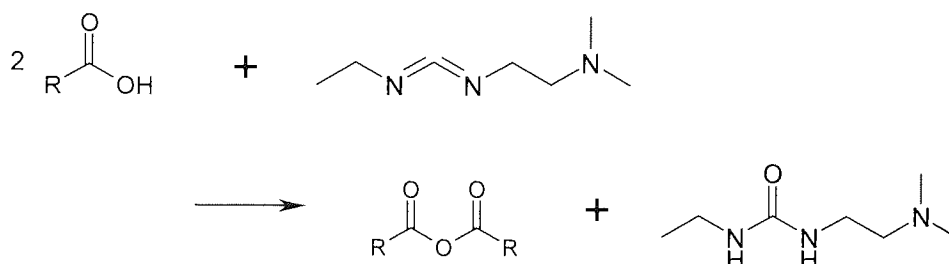


Such ureides are undesirable not only because they represent a loss of valuable carboxyl-component, but also because their separation from the main products of the reaction might be difficult, especially in the coupling of larger peptide segments. A remedy for this imperfection could be in the use of water-soluble carbodiimides, such as (salts of):



The salts of the urea-derivative formed in couplings with such modified reagents are extracted with water.

The agent EDC gives the symmetrical anhydrides, which could be isolated in the crystalline form by extraction of the reaction mixture with water in most cases. The symmetrical anhydride is of major significance to the chain elongation.



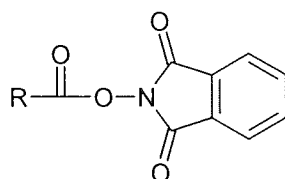
### (3.2) Active esters

Some electron-withdrawing substituents used for the activation of the carboxyl group cannot play the role of an acylating agent, such substituents are: acid chlorides and azides of carboxylic acids. They form a second acylation product since the leaving group cannot produce amides. The same can be said about alcohols if they can be used for activation. In esters the leaving group, an

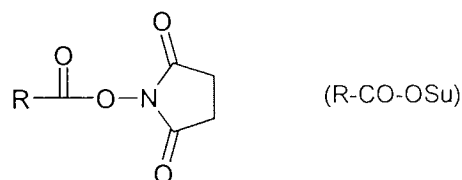
alcoholate or after protonation an alcohol, does not combine with the amino component:



An entirely new class of active esters, o-acyl derivatives of substituted hydroxylamines gained considerable importance. The first representatives of this class esters of N-hydroxyphthalimide, were discovered by Nefkens and Tesser (130):

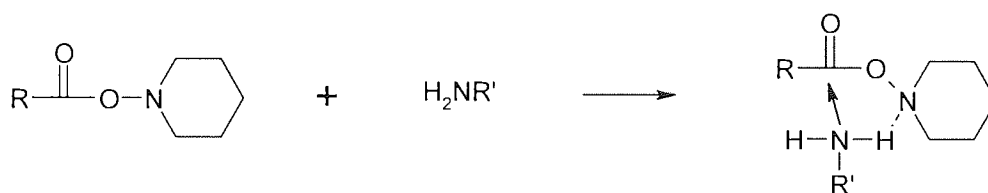


In a sense, these compounds could be considered mixed anhydrides, with the protected amino acid (R-COOH) as one of the acid constituents and a hydroxyamic acid (R'-CO-NHOH) as the other. In the practice of peptide synthesis, however, hydroxyphthalimide esters and their modified successors, the esters of N-hydroxysuccinimide (131),

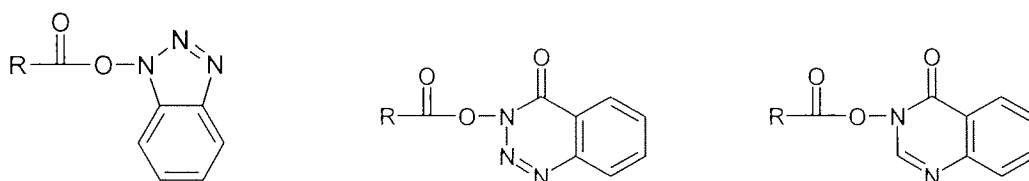


behave like active esters. They do not disproportionate to symmetrical anhydride and only exceptionally produce (via opening of the five-membered ring) a second acylation product.

An entire series of O-acyl hydroxylamines carry no acyl group of the nitrogen atom and are good acylating agents. Thus, the reactivity of hydroxyphthalimide and hydroxysuccinimide esters should not be attributed solely to their anhydride character. The best-known representative is N-hydroxypiperidine esters (132). The reactivity of O-acyl hydroxylamines is usually explained with anchimeric assistance provided by the nitrogen atom next to the ester oxygen (133);



In this connection, we should mention also the O-acyl derivatives of 1-hydroxybenzotriazole (HOBT), 3-hydroxy-3,4-dihydro-benzotriazine-4-one and 3-hydroxy-3,4-dihydro-quinazoline-4-one (133) as the active ester reagent.



Active esters are mostly not prepared in isolated form, but are produced and react in situations when these additives are applied for the suppression of racemization in coupling with carbodiimides (128) or for the catalysis (134) of otherwise only moderately reactive aryl esters.

While many O-acyl hydroxylamine derivatives have been proposed for peptide synthesis, only the esters of N-hydroxysuccinimide are widely used. They are crystalline, stable compounds, which have excellent reactivity in aminolysis reactions. Also the leaving N-hydroxysuccinimide is readily soluble in water and thus easily separated from the usually water insoluble product. In this respect N-hydroxysuccinimide esters look somewhat superior to their predecessors, the ester of N-hydroxyphthalimide that for the removal of the by-product required extraction with an aqueous solution of bicarbonate. A certain ambiguity seems to exist in the reactions of N-hydroxysuccinimide esters. The strained five-membered ring is fairly sensitive to nucleophiles, which can open it. The undesired acylation was recognized by Savrda (135).

#### **(4). Carboxylic acid protection**

Esterification of a carboxylic acid cannot usually be performed directly, but requires activation, often via the more reactive acyl chloride. 4-nitrobenzyl bromide, a fairly cheap and commercially available reagent was used to protect the carboxylic acid. Also, the ester of 4-nitrobenzyl bromide can be easily removed to the carboxylic acid by hydrogenation with palladium on carbon catalyst.

#### **(5). Hydroxyl group protection**

The hydroxyl groups were protected as tert-butyldimethyl silyl ethers due to their ease of synthesis and deprotection. While more expensive than trimethylsilyl chloride, tert-butyldimethyl silyl chloride was used in preference to the former due to the enhanced stability of the resulting ethers. The bulky tert-butyl groups sterically hinder attack on silicon thereby greatly reducing the risk of accidental deprotection during washing and purification stages.

## (6). Bromination

Bromination of one of the methyl groups was performed using the N-bromosuccinimide (NBS) in refluxing carbon tetrachloride via a radical reaction. Addition of only 1.2 equivalents of NBS to the symmetrical intermediate led to predominantly mono bromination, with only a minor amount of the di-brominated product being formed, which is unavoidable. The rate is about 4:1 (mono: di-bromination) according to Erwin Gotschi etc. (126). The resulting solid was used in the next step without purification. The main indications in  $^1\text{H}$  NMR were a new  $\text{CH}_2\text{Br}$  singlet at 4.49 ppm and peaks for the silyl protection groups which changed from singlets to a double-doublet at 0.18-0.28 ppm for  $\text{Si}(\text{CH}_3)_2$  and a doublet at 0.98-1.02 ppm for  $(\text{CH}_3)_3$ . This occurred as a result of the compound no longer being symmetrical.

## (7). The sulfide preparation

Methyl N-[N(tert-butoxycarbonyl)-L-seryl]-S-{4,6-bis[(tert-butyl)dimethylsiloxy]-3-methyl-2-[(4-nitrobenzyloxy)carbonyl]benzyl}-L-cysteinate **12**, the sulfur analogue of ethers, is prepared by a Williamson type  $\text{S}_{\text{N}}2$  reaction between a thiolate anion N-[N-(t-butoxycarbonyl)-L-seryl]-L-cysteine methyl ester **11** and a primary alkyl halide 4-Nitrobenzyl 2-bromomethyl-3,5-bis(tert-butyl)dimethylsilyloxy)-6-methylbenzoate **8** in the presence of the strong base TEA.

## (8). Deprotection of 4-nitrobenzyl ester

Deprotection was performed by hydrogenation of the 4-nitrobenzyl ester. Benzyl ethers and ester are cleaved by reductive hydrogenolysis, a reaction that does not affect other ethers and esters. If the hydrogen pressure is higher, the reaction would be quicker. Only in the hydrogen atmosphere, it could be 2 days to complete the hydrogenation and the yield is not so high.





#### **(10). Broth microdilution method for the determination of MIC values**

The broth microdilution method was carried out in accordance with NCCLS standards (NCCLS 2001). An overnight culture of each organism was grown on MHA plates at 37 °C. A loopful of the organism was suspended in sterile saline and the optical density was adjusted to 1.0 at 470nm while the water was set to 0.0 at 470nm. This gives a concentration of approximately  $10^8$  cfu/ml. This suspension was then diluted 1:100 in sterile MHB to give a cell density of  $10^6$  cfu/ml. Sterile U-bottom microtitre plates (Fisher, UK) were used for the MIC tests. A volume of 100 $\mu$ l of sterile MHB was dispensed into each well on the plates. The compounds were dissolved in DMSO at the concentration 5 mg/ml. It does not need to be sterilised, because no organism can survive in DMSO. The DMSO solution was diluted with broth (Mueller-Hinton broth) to the concentration 1 mg/ml. A volume of 100 $\mu$ l was then dispensed into the wells of the first column and mixed thoroughly. A 100 $\mu$ l broth was then removed from the wells in the first column and dispensed into the wells of the second column. This process was repeated across the wells of the plate. The concentration of the test antibiotic now ranged from 512  $\mu$ g/ml in the first column of wells to 0.25  $\mu$ g/ml in the twelfth well. The microtitre plates were sealed with adhesive plastic film and incubated for twenty-four hours at 37°C. Growth of the test organism was seen as a "button" of cells at the bottom of the well.

#### **(11). The enzyme-ligand design method**

With the fast development of computer technology, more and more programs are available commercially. And PC's are becoming powerful enough to operate the molecular graphics.

SPROUT version 3.4 is an interactive computer system for the structure based molecular design, (developed by ICAMS, School of Chemistry, University of Leeds, Leeds, UK, LS2 9JT). The system consists of five main modules, which incorporate constrained structure generation methods: detection of potential

protein cleft interaction sites, the use of the most appropriate structural units to fit those sites and within the boundary constraints, and generating primary and secondary structures for subsequent analysis. It can therefore generate an exhaustive range of structures that fit a specified protein receptor site. These structures are docked, using several algorithms, to the receptor site by a fast geometric rigid body docking process (75-77). SPROUT was incorporated into the Unix system on a 'Silicon Graphics Octane Workstation'. The computer system also met all the minimum requirements for running this version of SPROUT.

SPROUT is a computer program that is designed to generate molecules for a range of applications in molecular recognition. Structures are generated as primary and secondary structures, which incorporate various template, electrostatic or hydrophobic interactions. It has been successfully used in the primary structure generation of two enzyme inhibitors, namely at the trypsin and HIV protease receptor sites, that resemble the known substrates (139). As a result both time and money could be saved, resulting in more resources being allocated to the synthesis and testing of novel structures. There are five main parts in the SPROUT program.

### (11.1) CAnGAROO



Cleft Analysis by Geometry based Algorithm Regardless Of the Orientation.

The CAnGAROO module is used to detect potential binding pockets of protein structures by detecting clefts in the solvent accessible surface of the protein. This is achieved by determining the enzyme-ligand complex, which can be defined as either cavity or receptor files.

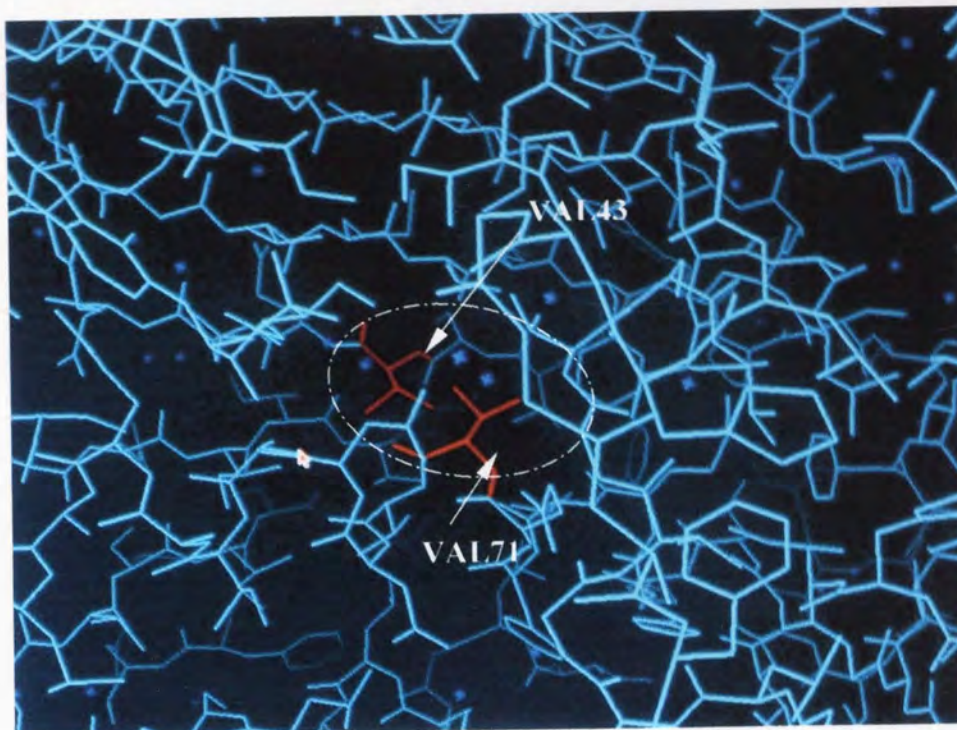
There are two methods that can be used in this module:

. *Choosing a cleft when there is a ligand attached to the protein:* this is the simplest method since if the enzyme-ligand complex is defined in a file, then the program will analyse the cleft and so in turn choose the appropriate target sites. The great difficulty in using this method is that no protein X-ray crystallographic structure is available, in which cyclothialidine is bound to the B subunit of DNA gyrase. As a consequence this method had to be disregarded.

. *Choosing a cleft when there is no ligand attached to the protein:* this is a much more difficult and lengthy method. It relies on use of all the extensive data available for the binding of cyclothialidine to the B subunit. Therefore, work was started by looking at all the research carried by Lewis *et al*, (140) and Boehm *et al*, (141) on the binding of cyclothialidine, at a molecular level. This allowed a molecular representation of the various amino acids, which were believed to be essential for the binding of cyclothialidine to be produced.

Initially the solvent accessible surface of the B subunit of DNA gyrase was computed from the Protein Data Bank as the PDB file. A 'new current set' was created, which enabled the two valine residues (VAL43, VAL71) to be defined as the cavity file. This is because these two residues are proved to interact with the resorcinol ring, which contributed the majority of the interaction between cyclothialidine and the protein (140). These amino acids were then used to define the receptor files, with the 'distance from current' being confirmed at 10Å, which means all the amino acid residues within 10Å radius around the cavity are selected for analysis.

The two amino acids, making up the binding cleft, are clearly visible in red from the remaining amino acids around the binding pockets in light blue (**Figure 10**).



**Figure 10: The two amino acids (in red) within the binding pocket of the B subunit.**

In the figure above, the binding site can be rotated along the  $x, y, z$  -axis. This enables the two amino acids to be clearly distinguished from all the remaining amino acids. The dashed oval shape clearly indicates the possible binding pocket of the B subunit.

### (11.2) HIPPO

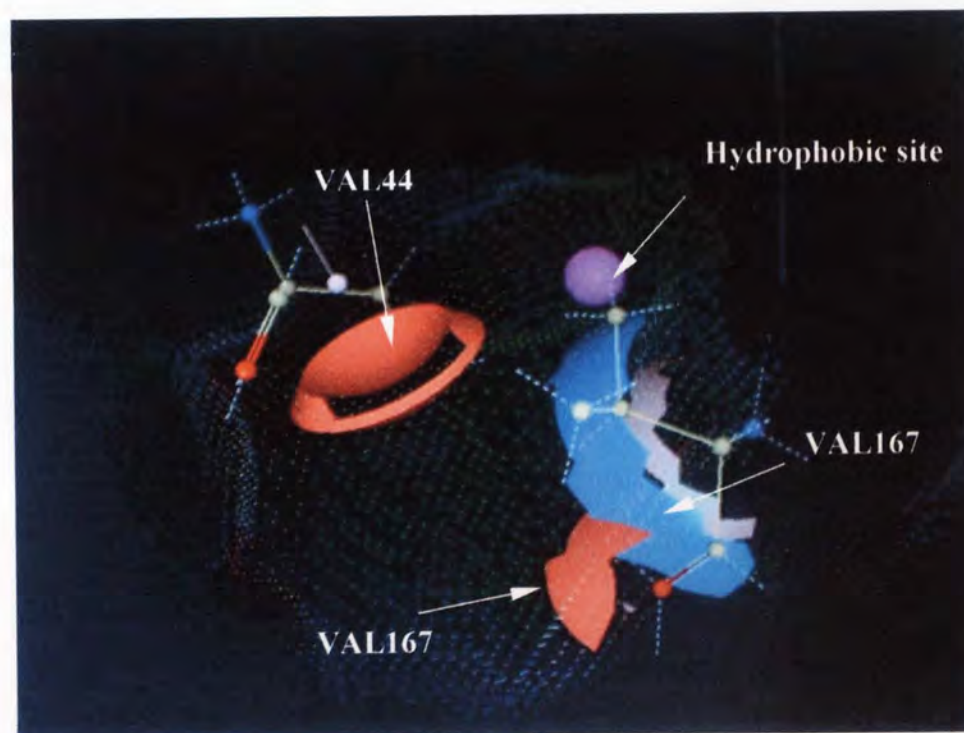


Hydrogen Bonding Interaction Site Prediction as Positions with Orientations

The HIPPO module identifies favourable hydrogen bonding, covalent bonding, and hydrophobic regions within a binding cleft. In turn it selects the interaction

sites, known as *target sites*, within the receptor site that are the starting points for structure generation.

If an enzyme-ligand complex was defined, then HIPPO would simply analyse the ligand and binding cleft, and highlight the various target sites automatically. Unfortunately, since there was no ligand bound to the cleft, the computer used the cavity (VAL 43 and VAL 71) as the ligand alternatively. After exploring the hydrogen bond in the receptor site as well as the ligand site, the hydrogen bond donor or acceptor for both receptor and ligand are highlighted. It is also necessary to highlight the hydrophobic site, which shows as a pink ball. So there are altogether 4 sites selected for the docking module. They are VAL 44 (donor), VAL 167 (donor), VAL 167 (acceptor) and the hydrophobic site (pink ball), (Figure 11).



**Figure 11: The four target sites making up the binding cleft.**

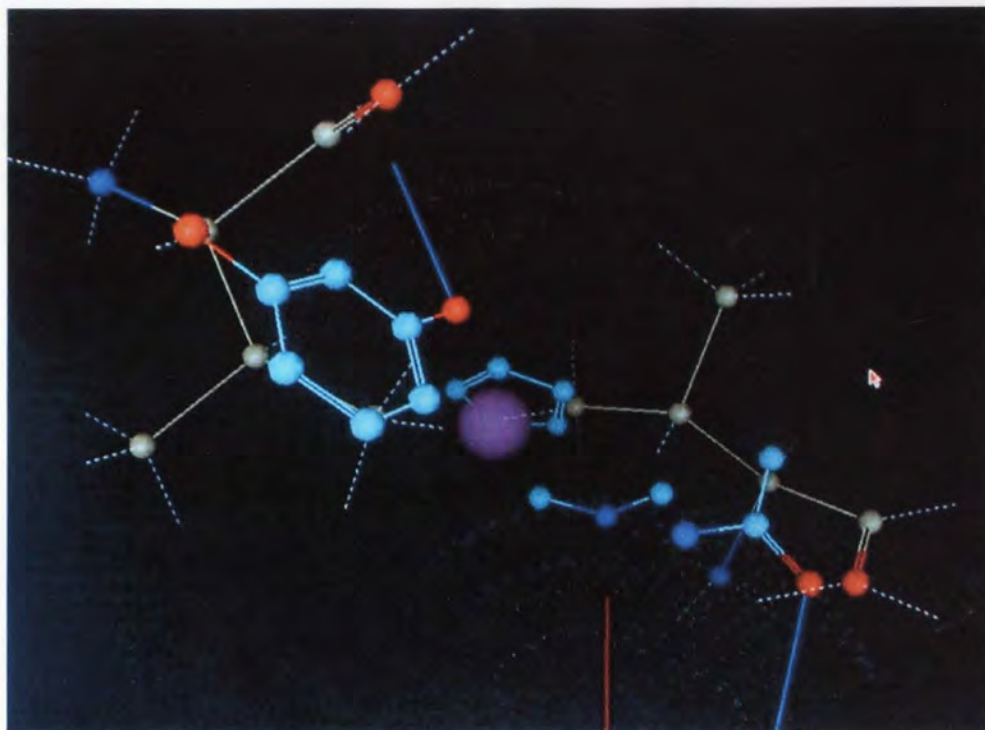
Once these amino acids have been computed, tolerances are applied to the distance and direction so that geometric regions define the target site. The solid coloured regions represent the target sites that satisfy the geometric parameters. And the 'points' represents the boundary surface, within which the potential compounds must lie (**Figure 11**).

### (11.3) ELeFAnT



**E**lection of Functional Groups and **A**nchoring them to **T**arget Sites

The EleFAnT module is designed to select the most favourable start templates or molecular fragments to dock to the selected target sites. There are only few templates in the SPROUT program database. A 'template library manager' mode can be used in both EleFAnT and SPIDeR, which enables a wider variety of desired templates to be incorporated into the target sites. The different structure fragment was docked in to fulfil the active site (**Figure 12**).



**Figure 12. The fulfilled templates to the active sites**

Usually, we select as many groups as we have. The program can automatically find the best fitted one according to the distances, angles when the structure is growing. In the **Figure 12**, the screen only displays one template for each active site. The oxygen is the hydrogen bond acceptor. The five or six-membered aromatic ring is for the hydrophobic site; the amino hydrogen is the hydrogen bond donor.

#### (11.4) SPIDeR

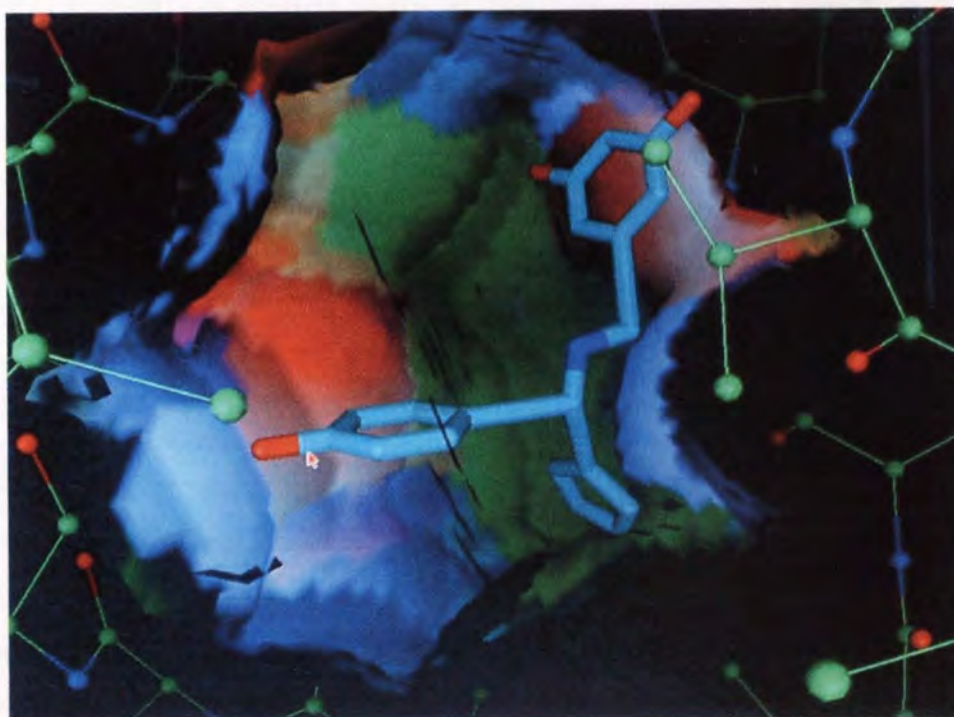


Structure Production with **Interactive Design of Results**

The SPIDeR module connects the start templates generated in EleFAnT, with spacer templates, to give complete molecular structures (skeletons) or molecular graphs that satisfy the steric constraints of the binding pocket.

SPIDeR automatically selects the tree pair to be connected. This is the pair of trees that are most sterically hindered or are the closest together.

- The first tree pair connection: was made between the hydrophobic site and VAL 167 acceptor sites. All of the spacer templates were selected, since there would be a high degree of steric flexibility and this option would yield structures as many as the computer can. The number of partial skeletons produced at this stage, was 469.
- The second tree pair connection: was made between the 469 partial skeletons produced and VAL 167 donor site. The number of complete skeletons produced by geometric docking and that satisfied all the geometric constraints of the target sites, was 369.
- The third tree pair connection: was made between the 369 partial skeletons and VAL 44 donor site. The final generated structure number is 17.



**Figure 13.** A complete skeleton produced at the end of SPIDeR



**Figure 13** shows one of the 17 skeletons produced, and its binding to the four respective interaction sites. The colourful pocket is the binding pocket, which the ligand should lie in.

### (11.5) ALLigaTOR



Analyse Lots of **L**igands, Test and Order **R**esults

The ALLigaTOR module enables the clustering, scoring and display of solutions, which thereby allows a more efficient means of evaluating the results. It also adds heteroatoms into the skeleton to generate molecules with similar properties to the receptor site. This would be very helpful if we have generated thousands of structures, so that we can cluster them into several different groups, then select to see the samples in each group instead of overviewing all of them. While we only have 17 structures to study, it is wise to read them just from the screen.

## (12). Quantitative Structure-Activity Relationships (QSARs) methods

### (12.1) TSAR Program

TSAR is an integrated analysis package for interactive investigation of Quantitative Structure-Activity Relationships (QSARs). It is intended to provide all the functions required to carry out any QSAR investigation, whether in pharmaceutical, agrochemical, toxicological, or any other area of application. TSAR uses an integrated approach to provide all components together. It uses a chemically aware spreadsheet to store and manipulate different types of data, including:

- Molecular descriptions
- 3D structures
- Activity data
- Computed data

All the structures, properties, and associated data are held in a chemical spreadsheet called a project. Each row stores information about one molecule (or substituent) and each column stores one type of information, for example, a 3D structure, a molecular descriptor, activity data, or computed data. TSAR does not limit the amount of data one can store, so one can use it to spot the highlights from large screening runs. TSAR stores full 3D structural information, not just an image of a structure. Therefore, one can use TSAR to work interactively with these data, and calculate and predict structural properties. TSAR derives a wide range of properties from the structural information. For example, there are simple calculations such as 2D topology indices, and more complex calculations, including determining molecular similarity values; building 3D molecular structures, and determining quantum mechanics derived properties (via interfaces to the companion programs Asp, Corina, and Vamp, respectively). 3D structures can be seen using a sophisticated display facility that makes optimum use of the graphics facilities of the hardware. The user has full interactive control of the structures on display, such as 3D manipulation, including hardware depth cueing, where available. One can generate and manipulate shaded transparent molecular surfaces, and display more than one molecule to compare alignment methods.

Because all data can be stored together in a project, it is easy to combine external data, computed data, and structural data, and use any of them for analysis and prediction. TSAR provides facilities to analyze continuous, discrete, and classification data. TSAR provides the following predictive techniques:

- Neural network analysis (good for non-linear relationships)
- Regression analysis
- Partial least squares analysis

TSAR provides the following data reduction techniques:

- Principal components analysis

Results of analysis are displayed as one or more tabbed views showing details appropriate to the analysis technique. Each view highlights one aspect of the analysis, e.g., a summary list of equations and statistical tests, a spreadsheet of predicted values, or a correlation matrix. There are a variety of graphical views of the results to aid further analysis. One can add the results of the analysis (for example, predicted results, principal components, or partial least squares dimensions) to the project. TSAR can store the calculation equations with the project. Results that are added to the project can be manipulated in the normal way, for example, to use the data to plot a graph, or perform mathematical operations on it to explore non-linear functions.

All the structures are inputted in the Chem-X program with 3D sketch, and the energy was optimized with the MOPAC package. The output was saved as the MDL SD file. For a file in MDL SD format, one can import both structural data and numeric (or text) data into a project at the same time. TSAR enables the user to consider structures as whole molecules and also to define areas of interest or variability as substituents (an area of variability or substitution in a group of structures). The part of a molecule not defined as a substituent is considered to be the generic (the part of a structure or group of structures that does not vary) or template part of the molecule. So the two side chains: the C-terminal in Cysteine and the N-terminal in Serine were defined as the two substituents; the 12-member ring becomes the template part of the molecule. When these molecules are imported into the project, they are considered only as a whole molecule. We must then define any substituents that we want to work with. To do this, we must select all the structures for which we want to define substituents, but define the substituents for only one structure in the selection. The substituents that we define for this one structure are used by TSAR to define substituents for all of the other structures that are currently selected, as long as they are all based on a common

generic, or template. Then to identify the substituents, TSAR searches the current database for matching structures and adds the names to the substituent name columns in the project. Most of the properties reported in TSAR are calculated for either whole molecules or for substituents. For each of the property columns, there is a label specifying whether the information in the column is associated with the whole molecule, or with one of the substituents. The column that contains the data produced by a calculation on a substituent is labeled as (Subst.  $n$ ), where  $n$  is the number of the substituent.

First of all, we should calculate the partial atomic charges for a molecule using Charge-2. Deriving partial charges is a prerequisite for several structure manipulations, such as aligning structures by molecular weighted extent and optimizing a 3D model, and for many property calculations. All the structures are aligned with each other so that it is easier to identify the differences between them. They are aligned by the shape and weight of the whole molecule. The parameters can include the size, charge, log P value, and molar refractivity of the molecule, which is called Molecular Weighted Extent.

## **(12.2) Data reduction**

Principal components analysis (PCA) is a technique to reduce a large number of variables to a smaller number without losing useful information. PCA concentrates the variance of the dataset in the first few principal components. This means that we can view much of the variance of the dataset by plotting the first two or three principal components. We can use the most significant principal components calculated as a starting point for further analysis. Unfortunately, there is no simple way to define which principal components are significant and which are insignificant. Possible strategies include:

- Keeping a fixed number of components, for example, the first three
- Discarding any components that explain less than a chosen fraction of the variance.
- Keeping as many components as are required to explain a fixed percentage of the variance

- Discarding any components whose associated eigenvalue is less than 1.0, that is, any principal components that explain less of the variance than one of the original variables.

PCA will give a matrix. This matrix lists the Principal Components with their composition in terms of the original variables. The larger the magnitude of each entry, the more significant the contribution of that variable to that component. (Positive and negative values are equally significant - only the sense of the contribution is reversed.)

### **(12.3) Multiple Regression Analysis**

Using the data generated from the data reduction to develop a linear relationship between the parameters and activity, regression analysis calculates an equation describing the relationship between a single dependent  $y$  variable, and several explanatory  $x$  variables. This equation can then be used to predict values of  $y$ , and these can be added to the project. The results of the regression analysis are displayed as:

**Summary** Lists the regression equation, calculation details, variance analysis information, and a series of statistical test results

**Confidence** Lists the variables used in the regression equation and the standardization used. It includes the number of times each variable is used in the cross validation model, with coefficients, standard errors.

**Stepping** Shows details of each step performed during the stepping process. It includes details of each variable movement with information about the current fractions of points that are well classified

**Correlation** A correlation matrix for the input variables

**Data** Lists the actual, predicted, residual, and residual variance values

#### **(12.4) The Cross Validation**

In the Cross Validation of Multiple Regression Analysis, the project rows are sorted by  $y$  value and then data is deleted or excluded from the calculations in a systematic way (defined by the user when the calculation is set up). There are two options available.

##### **Leave out one row**

Each row is left out in turn, so that the value of each row is predicted from all others. Alternatively, leave out a single row, chosen at random, a specific number of times.

##### **Leave out groups of rows**

Groups of rows are left out, excluding a third of the data from each model in a fixed pattern. Alternatively, leave out groups of rows chosen at random.

The cross-validation  $R$  squared is the key measurement to the reliability of the generated equations.

## **Chapter 3. Experimental**

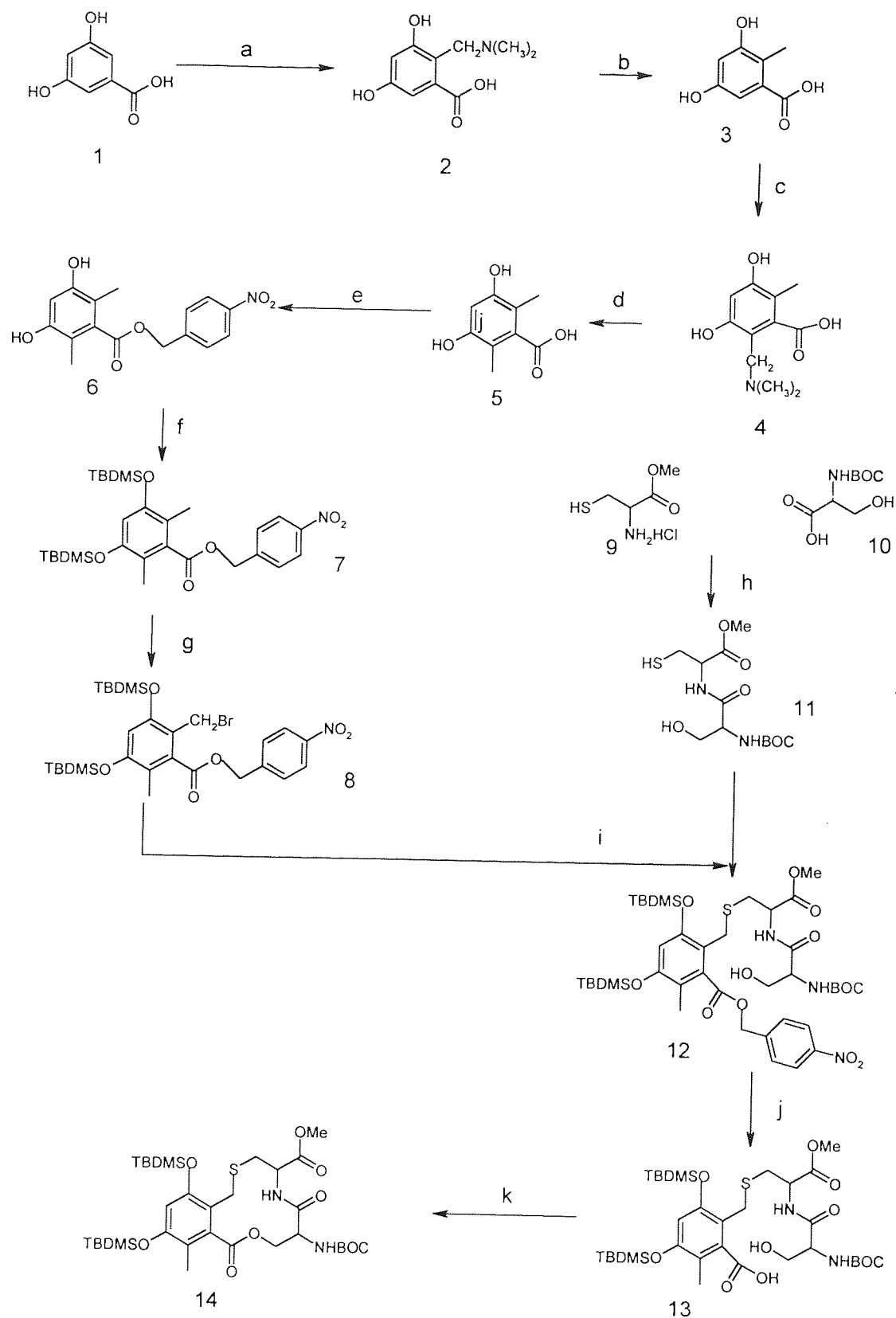


Figure 9. The organic synthesis scheme



- a) aq. CH<sub>2</sub>O, aq. NH(CH<sub>3</sub>)<sub>2</sub>, AcOH, r.t, 18h;
- b) Pd/C, H<sub>2</sub>, 8-9 bar, 30h;
- c) aq. CH<sub>2</sub>O, aq. NH(CH<sub>3</sub>)<sub>2</sub>, AcOH, r.t, 18h;
- d) Pd/C, H<sub>2</sub>, 8-9 bar, 30h;
- e) 1,1,3,3-tetramethylguanidine, 4-nitrobenzyl bromide, DMF, r.t, 18 hours;
- f) TBDMSCl, imidazole, DMF, 0°C, 3 hours;
- g) NBS, CCl<sub>4</sub>, reflux, irradiation, 3 hours;
- h) DCC, Acetonitrile, 4-methylmorpholine, 0°C, 3 hours;
- i) DCM, TEA, 2 hours at 0°C, then r.t 18 hours;
- j) Pd/C EtOAc, H<sub>2</sub>, 6-7 bar, r.t, 30 hours;
- k) DEAD, Ph<sub>3</sub>P, 0°C, 40 mins, then 5 hours at r.t;

### 1. Synthesis of Cyclothialidine analogues:

h) Synthesis of the dipeptides **11**. (N-BOC-Ser-L-Cys methyl ester.)

A suspension of L-Cys methyl ester.HCl (1.86g, 10mmol) and N-Boc-Ser-OH (2g, 10mmol) in dry acetonitrile 25 ml was treated at 0°C with 4-methylmorpholine (1.3ml, 12 mmol). To the stirred solution was added dropwise at 0°C a solution of DCC (2.47g, 12 mmol) in dry acetonitrile 5ml. After stirring for 3 hours at 0°C, the resulting white precipitate was removed by filtration and the filtrate was evaporated in vacuo. The residue was dissolved in 150ml EtOAc, washed with 2M HCl (30ml x 2), saturated NaHCO<sub>3</sub> (30ml x 2) and brine (30ml x 2), then dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed in vacuo to yield the white solid. Purified by column chromatography using EtOAc: Hexane (1:1) as the eluant afforded the dipeptides. Yield 60%. M.p. 74-76°C. lit (74-76°C).

<sup>1</sup> H NMR (CDCl<sub>3</sub>): δ 1.44 (9H, s, Boc), 2.97-3.04 (2H, m, CH<sub>2</sub>SH), 3.61-3.72 (1H, m, CHCH<sub>2</sub>OH), 3.77 (3H, s, OCH<sub>3</sub>), 4.03-4.12 (2H, m, CH<sub>2</sub>OH), 4.18-4.28

(1H, m, SH), 4.78-4.84 (1H, m, CHCH<sub>2</sub>SH), 5.55 (1H, d, NHBOC), 7.35 (1H, m, NH).

<sup>13</sup>C NMR (CDCl<sub>3</sub>): δ 26.5 (CH<sub>2</sub>SH), 28.9 (CH<sub>3</sub>), 53.0 (CHNH), 53.7 (CHNH), 54.9 (OCH<sub>3</sub>), 62.7 (CH2OH), 81.1 (C(CH<sub>3</sub>)<sub>3</sub>), 156.7(COO), 170.2 (NHCO), 171.2 (NHCO);

IR (KBr): ν<sub>max</sub> 1161, 1222, 1248, 1367, 1519, 1670, 1685, 1741cm<sup>-1</sup>.

m/z 323 [M + H]<sup>+</sup>

a) Synthesis of α-(dimethylamino)-3,5-dihydroxy-2-methylbenzoic acid acetate **2**:

To a stirred mixture of 37% aqueous formaldehyde (1ml, 13mmol), ethanol (3ml) and glacial acetic acid (3ml) was added 40% aqueous dimethylamine (1.6ml, 13mmol) at 0°C. Stirring was continued for 30 minutes at room temperature. (2g, 13mmol) 3,5-dihydroxybenzoic acid was added at 0°C. Then the cooling bath was removed and the stirring was continued overnight. The resulting white precipitate was isolated by filtration and washed with ethanol (20 ml) and diethyl ether (20 ml) to afford **2**. Yield 68.5%. M.p. >300 °C

<sup>1</sup>H NMR (CDCl<sub>3</sub>): δ 2.55 (6H, s, N (CH<sub>3</sub>)<sub>2</sub>), 3.92 (2H, s, CH<sub>2</sub>N), 6.34 (1H, d, Ar-CH, J=4), 6.65 (1H, d, Ar-CH), 9.54 (1H, b, COOH);

<sup>13</sup>C NMR (CDCl<sub>3</sub>): δ 40.3 (N (CH<sub>3</sub>)<sub>2</sub>), 52.6 (CH<sub>2</sub>N), 103.1 (Ar-CH), 107.1 (CCH<sub>3</sub>), 109.2 (Ar-CH), 143.5 (CCOOH), 157.2 (COH), 158.2 (COH), 171.1 (COOH);

IR (KBr disc): ν<sub>max</sub> 764, 1150, 1294, 1356, 1471, 1558, 1614, 3130cm<sup>-1</sup>.

m/z 212 [M + H]<sup>+</sup>

b). Synthesis of 3,5-dihydroxy-2-methylbenzoic acid **3**:

(1g, 3.7mmol) **2** in methanol 15ml was treated with a suspension of 10 % Pd/ C in 3M NaOH solution 1.5ml under 7-8 bar hydrogen pressure at room temperature. Stirring was continued for 36 hours. The pH of the mixture was adjusted to 1 by addition of 2MHCl. The catalyst was removed by filtration through celite and the filtrate concentrated in vacuo. The residue was extracted with EtOAc (15 ml), washed with 2M HCl (10 ml x 2) and brine (10 ml x 2). The solvent was dried over Na<sub>2</sub>SO<sub>4</sub> and evaporated to afford an organic solid. TLC R<sub>f</sub> = 0.25 as eluent MeOH:EtOAc (1:1). Yield 87.6% M.p. 245-248°C. lit (245°C, dec)

<sup>1</sup>H NMR (DMSO): δ 2.11 (3H, s, CH<sub>3</sub>), 6.34-6.44 (1H, d, Ar-CH), 6.59-6.60 (1H, d, Ar-CH), 9.23 (1H, s, Ar-OH), 9.42 (1H, s, Ar-OH);

<sup>13</sup>C NMR (DMSO): δ 17.2 (CH<sub>3</sub>), 110.3 (Ar-CH), 112.4 (Ar-CH), 120.3 (CCH<sub>3</sub>), 137.3 (CCOOH), 160.1 (COH), 161.3 (COH); 174.4 (COOH);

IR (KBr disc): ν<sub>max</sub> 1009, 1159, 1306, 1328, 1411, 1610, 1689, 3240cm<sup>-1</sup>.

m/z 167 [M + H]<sup>+</sup>

c). Synthesis of α-(dimethylamino)-3,5-dihydroxy-2,6-dimethylbenzoic acid acetate **4**:

To a stirred mixture of 37% aqueous formaldehyde (5ml, 60 mmol), ethanol 25ml and glacial acetic acid 25 ml was added 40% dimethylamine (7ml, 60 mmol) at 0°C. Stirring was continued for 30 minutes at room temperature. (10g, 60mmol) **3** was added at 0°C. The ice bath was removed and stirring was continued overnight. The resulting white precipitate was isolated by filtration and washed with ethanol (40 ml) and diethyl ether (40 ml). Yield 52%. M.p.>300°C

<sup>1</sup>H NMR (DMSO): δ 1.95 (3H, s, CH<sub>3</sub>), 2.53 (6H, s, N(CH<sub>3</sub>)<sub>2</sub>), 3.80 (2H, s, CH<sub>2</sub>N), 6.31 (1H, s, Ar-CH);

$^{13}\text{C}$  NMR (DMSO):  $\delta$  12.7 ( $\underline{\text{C}}\text{H}_3$ ), 42.2 ( $\text{N}(\underline{\text{C}}\text{H}_3)_2$ ), 53.8 ( $\underline{\text{C}}\text{H}_2\text{N}$ ), 100.5 ( $\text{Ar}-\underline{\text{C}}\text{H}$ ), 104.5 ( $\text{Ar}-\underline{\text{C}}\text{CH}_3$ ), 111.5 ( $\text{Ar}-\underline{\text{C}}\text{CH}_2\text{NH}$ ), 145.9 ( $\text{Ar}-\underline{\text{C}}\text{COOH}$ ), 154.5 ( $\text{Ar}-\underline{\text{C}}\text{OH}$ ), 156.8 ( $\text{Ar}-\underline{\text{C}}\text{OH}$ );

IR (KBr disc):  $\nu_{\text{max}}$  1301, 1369, 1398, 1465, 1568, 1596, 2730, 2873  $\text{cm}^{-1}$ .

$m/z$  226  $[\text{M} + \text{H}]^+$

d). Synthesis of 3,5-dihydroxy-2,6-dimethylbenzoic acid **5**:

1g, 3.5mmol  $\alpha$ -dimethylamino-3,5-dihydroxy-2,6-dimethylbenzoic acid acetate **4** in methanol 10 ml was treated with a suspension of 100mg Pd/C in 3M NaOH solution 0.9ml. The resulting mixture was stirred at 8-9 bar under hydrogen at room temperature for 30 hours. The pH of the mixture was adjusted to 1 by 2M HCl. The catalyst was removed by filtration through celite and the yellow filtrate was evaporated in vacuo. The residue was extracted with EtOAc 100ml, washed with 2M HCl (2 x 30ml), brine (2 x 30ml), dried over  $\text{Na}_2\text{SO}_4$  and the solvent evaporated to afford an orange solid. TLC  $R_f=0.8$  with eluent EtOAc:  $\text{CH}_3\text{OH}$  (1:1). Yield 51%. M.p. 166-172°C. lit (178-179°C).

$^1\text{H}$  NMR (DMSO):  $\delta$  1.91 (6H, s,  $\underline{\text{C}}\text{H}_3$ ), 6.37 (1H, s,  $\text{Ar}-\underline{\text{C}}\text{H}$ ), 9.15 (2H, s,  $\text{Ar}-\underline{\text{O}}\text{H}$ );

$^{13}\text{C}$  NMR (DMSO):  $\delta$  12.3 ( $\underline{\text{C}}\text{H}_3$ ), 102.4 ( $\text{Ar}-\underline{\text{C}}\text{H}$ ), 109.7 ( $\text{Ar}-\underline{\text{C}}\text{CH}_3$ ), 138.0 ( $\text{Ar}-\underline{\text{C}}\text{COOH}$ ), 153.5 ( $\text{Ar}-\underline{\text{C}}\text{OH}$ ), 171.2 ( $\underline{\text{C}}\text{OOH}$ );

IR (KBr disc):  $\nu_{\text{max}}$  1120, 1261, 1340, 1375, 1605, 1668, 3236, 3384  $\text{cm}^{-1}$ .

$m/z$  181  $[\text{M} + \text{H}]^+$

e). Synthesis of 4-nitrobenzyl-3,5-dihydroxy-2,6-dimethylbenzoate **6**:

To a solution of 3,5-dihydroxy-2,6-dimethylbenzoic acid ( 5g, 27.5mmol) in DMF 20ml was added at 0°C 1,1,3,3-tetramethylguanidine (3.5ml, 27.5mmol). After

stirring for 15 minutes at room temperature (pink precipitate formed), 4-nitrobenzyl bromide (5.87g, 27.5mmol) was added and stirring continued at room temperature overnight. The mixture was diluted with EtOAc 100ml, washed with NaHCO<sub>3</sub> (1 x 30ml), 2M HCl (2 x 30ml) and brine (2 x 30ml), dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was evaporated in vacuo to afford the yellow solid. Yield 79%.

M.p. 161-164°C. lit (168-170°C)

<sup>1</sup>H NMR (DMSO): δ 1.86 (6H, s, CH<sub>3</sub>), 5.44 (2H, s, OCH<sub>2</sub>), 6.42 (1H, s, Ar-CH), 7.68 (2H, d, Ar-CHNO<sub>2</sub>), 8.24 (2H, d, Ar-CHNO<sub>2</sub>), 9.26 (1H, s, Ar-OH);

<sup>13</sup>C NMR (DMSO): δ 12.3 (CH<sub>3</sub>), 65.1 (OCH<sub>2</sub>-Ar), 103.2 (Ar-CH), 110.6 (Ar-CCH<sub>3</sub>), 123.8 (Ar-CHCNO<sub>2</sub>), 129.4 (Ar-CHCNO<sub>2</sub>), 135.2 (Ar-CCOO), 143.6 (OCH<sub>2</sub>C-Ar), 147.3 (Ar-CNO<sub>2</sub>), 153.7 (Ar-COH), 169.3 (COOH);

IR (KBr disc): ν<sub>max</sub> 1103, 1245, 1261, 1344, 1515, 1602, 1711, 3473cm<sup>-1</sup>.

m/z 317 [M + H]<sup>+</sup>

f). Synthesis of 4-nitrobenzyl 3,5-bis(tert-butyldimethylsilyloxy)-2,6-dimethylbenzoate 7:

To a stirred mixture of 4-nitrobenzyl-3,5-dihydroxy-2,6-dimethylbenzoate **6** (1g, 3.2mmol) and TBDMSCl (1g, 6.6mmol) in DMF 5ml was added at 0°C imidazole (430mg, 6.4mmol). The mixture was stirred at room temperature for 3 hours, then diluted with EtOAc 50ml, washed with water (2 x 15ml), 2M HCl (2 x 15ml) and brine (2 x 15ml), dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was evaporated. The product was purified with column chromatography using eluant EtOAc:Hexane (1:15). R<sub>f</sub>=0.2 to afford a pale yellow solid. Yield 94%. M.p. 116-118°C. lit (121-122°C)

<sup>1</sup>H NMR (CDCl<sub>3</sub>): δ 0.18 (12H, d, 2 x Si(CH<sub>3</sub>)<sub>2</sub>), 0.98 (18H, d, 2 x (CH<sub>3</sub>)<sub>3</sub>), 2.00 (6H, s, 2 x CH<sub>3</sub>), 5.41 (2H, s, OCH<sub>2</sub>), 6.32 (1H, s, Ar-CH), 7.61 (2H, d, Ar-CHNO<sub>2</sub>), 8.21 (2H, d, Ar-CHNO<sub>2</sub>);

$^{13}\text{C}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  -4.3 ( $\text{Si}(\underline{\text{C}}\text{H}_3)_2$ ), 13.0 ( $\underline{\text{C}}\text{H}_3$ ), 18.1 ( $\text{Si}\underline{\text{C}}(\text{CH}_3)_3$ ), 25.6 ( $\underline{\text{C}}(\text{CH}_3)_3$ ), 65.1 ( $\text{O}\underline{\text{C}}\text{H}_2\text{-Ar}$ ), 110.7 ( $\text{Ar-}\underline{\text{C}}\text{H}$ ), 117.8 ( $\text{Ar-}\underline{\text{C}}\text{CH}_3$ ), 123.6 ( $\text{Ar-}\underline{\text{C}}\text{HNO}_2$ ), 128.9 ( $\text{Ar-}\underline{\text{C}}\text{HNO}_2$ ), 135.2 ( $\text{Ar-}\underline{\text{C}}\text{COO}$ ), 142.7 ( $\text{OCH}_2\underline{\text{C}}\text{-Ar}$ ), 147.4 ( $\text{Ar-}\underline{\text{C}}\text{NO}_2$ ), 151.8 ( $\text{Ar-}\underline{\text{C}}\text{OSi}$ ), 169.5 ( $\underline{\text{C}}\text{OO}$ );  
IR (KBr disc):  $\nu_{\text{max}}$  783, 837, 1031, 1257, 1340, 1469, 1527, 1723  $\text{cm}^{-1}$ .  
 $m/z$  546  $[\text{M} + \text{H}]^+$

g). Synthesis of 4-nitrobenzyl 2-bromomethyl-3,5-bis(tert-butyldimethylsilyloxy)-6-methylbenzoate **8**:

A stirred mixture of 4-nitrobenzyl 3,5-bis(tert-butyldimethylsilyloxy)-2,6-dimethylbenzoate **7** (22.44g, 41.11mmol) and NBS (8.78g, 49.43mmol) in  $\text{CCl}_4$  140ml was refluxed with 60W light irradiation for 2 hours. On cooling, the solvent was evaporated in vacuo to afford yellow oil. The oil was used directly to the next step without further purification.

$^1\text{H}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  0.18-0.28 (12H, d, 2 x  $\text{Si}(\underline{\text{C}}\text{H}_3)_2$ ), 0.98 (18H, d, 2 x  $(\underline{\text{C}}\text{H}_3)_3$ ), 2.02 (3H, s,  $\underline{\text{C}}\text{H}_3$ ), 4.49 (2H, s,  $\underline{\text{C}}\text{H}_2\text{Br}$ ), 5.46 (2H, s,  $\text{O}\underline{\text{C}}\text{H}_2$ ), 6.34 (1H, s,  $\text{Ar-}\underline{\text{C}}\text{H}$ ), 7.64 (2H, d,  $\text{Ar-}\underline{\text{C}}\text{HNO}_2$ ), 8.21 (2H, d,  $\text{Ar-}\underline{\text{C}}\text{HNO}_2$ );  
 $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  -4.5 ( $\text{Si}(\underline{\text{C}}\text{H}_3)_3$ ), -4.2 ( $\text{Si}(\underline{\text{C}}\text{H}_3)_3$ ), 13.1 ( $\underline{\text{C}}\text{H}_3$ ), 18.2 ( $\text{Si}\underline{\text{C}}(\text{CH}_3)_3$ ), 25.6 ( $\text{Si}\underline{\text{C}}(\text{CH}_3)_3$ ), 25.7 ( $\text{Si}\underline{\text{C}}(\text{CH}_3)_3$ ), 30.8 ( $\underline{\text{C}}\text{H}_2\text{Br}$ ), 64.9 ( $\text{O}\underline{\text{C}}\text{H}_2\text{-Ar}$ ), 114.6 ( $\text{Ar-}\underline{\text{C}}\text{H}$ ), 122.9 ( $\text{Ar-}\underline{\text{C}}\text{CH}_3$ ), 123.6 ( $\text{Ar-}\underline{\text{C}}\text{HNO}_2$ ), 128.1 ( $\text{Ar-}\underline{\text{C}}\text{HNO}_2$ ), 135.2 ( $\text{Ar-}\underline{\text{C}}\text{COO}$ ), 143.2 ( $\text{OCH}_2\underline{\text{C}}\text{-Ar}$ ), 152.5 ( $\text{Ar-}\underline{\text{C}}\text{OSi}$ ), 155.1 ( $\text{Ar-}\underline{\text{C}}\text{OSi}$ ), 167.4 ( $\underline{\text{C}}\text{OO}$ );  
IR (KBr disc):  $\nu_{\text{max}}$  1119, 1281, 1313, 1348, 1521, 1781, 2952, 3249  $\text{cm}^{-1}$ .  
 $m/z$  624  $[\text{M} + \text{H}]^+$

i). Synthesis of methyl N-[N(tert-butoxycarbonyl)-L-seryl]-S-{4,6-bis[(tert-butyl)dimethylsiloxy]-3-methyl-2-[(4-nitrobenzyloxy)carbonyl]benzyl}-L-cysteinate **12**:

To a solution of 4-nitrobenzyl 2-bromomethyl-3,5-bis(tert-butyl dimethylsilyloxy)-6-methylbenzoate **8** (5.73g, 9.17mmol) and N-[N(tert-butoxycarbonyl)-L-seryl]-L-cysteine methyl ester **11** (2.98g, 9.17mmol) in anhydrous DCM (75ml) was added dropwise at 0°C, TEA (1.27ml, 9.17mmol) in DCM (5ml). After stirring for 2 hours at 0°C and overnight at room temperature, the reaction mixture was washed with saturated NaHCO<sub>3</sub> (2 x 20ml) and brine (2 x 20ml), dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed in vacuo. The product was purified by column chromatography with the eluent EtOAc:Hexane (1:2) R<sub>f</sub>=0.3 to afford the yellow oil. Yield 82%

<sup>1</sup>H NMR (CDCl<sub>3</sub>): δ 0.22-0.24 (12H, d, 2 x Si(CH<sub>3</sub>)<sub>2</sub>), 0.97-0.99 (18H, d, 2 x C(CH<sub>3</sub>)<sub>3</sub>), 1.42 (9H, s, BOC), 2.00 (3H, s, Ar-CH<sub>3</sub>), 2.87-2.90 (2H, m, SCH<sub>2</sub>), 3.07 (1H, br, CH<sub>2</sub>OH), 3.68-3.76 (6H, m, OCH<sub>3</sub>, Ar-CH<sub>2</sub>S, CHOH), 4.06 (1H, br, NHCH(Ser)), 4.26 (1H, br, CHOH), 4.68-4.71 (1H, m, NHCH(Cys)), 5.47 (2H, s, COOCH<sub>2</sub>), 5.60 (1H, d, NH), 6.34 (1H, s, Ar-CH), 7.10 (1H, d, NH), 7.63-7.67 (2H, d, Ar-CHNO<sub>2</sub>), 8.22-8.25 (2H, d, Ar-CHNO<sub>2</sub>);

<sup>13</sup>C NMR (CDCl<sub>3</sub>): δ -4.3 (Si(CH<sub>3</sub>)<sub>2</sub>), -4.2 (Si(CH<sub>3</sub>)<sub>2</sub>), 13.2 (CH<sub>3</sub>), 18.1 (SiC(CH<sub>3</sub>)<sub>3</sub>), 18.2 (SiC(CH<sub>3</sub>)<sub>3</sub>), 25.5 (SiC(CH<sub>3</sub>)<sub>3</sub>), 25.6 (SiC(CH<sub>3</sub>)<sub>3</sub>), 28.2 (C(CH<sub>3</sub>)<sub>3</sub>), 33.7 (CH<sub>2</sub>S + PH-CH<sub>2</sub>S), 52.0 (NHCH), 52.6 (OCH<sub>3</sub>), 55.3 (NHCH), 63.2 (CH<sub>2</sub>OH), 65.9 (OCH<sub>2</sub>-Ar), 80.1 (C(CH<sub>3</sub>)<sub>3</sub>), 110.5 (Ar-CH), 117.7 (Ar-CCH<sub>2</sub>S), 119.2 (Ar-CCH<sub>3</sub>), 123.7 (Ar-CHNO<sub>2</sub>), 129.2 (Ar-CHNO<sub>2</sub>), 134.6 (CCOO), 142.4 (OCH<sub>2</sub>C-Ar), 147.7 (Ar-CNO<sub>2</sub>), 152.2 (COSi), 153.5 (COSi), 155.6 (COO), 169.1 (COO), 171.0 (NHC=O);

IR (KBr disc): ν<sub>max</sub> 837, 1155, 1257, 1342, 1465, 1523, 1724, 1731cm<sup>-1</sup>.

m/z 866 [M + H]<sup>+</sup>

j). Synthesis of methyl N-[N(tert-butoxycarbonyl)-L-seryl]-S-{4,6-bis[(tert-butyl)dimethylsiloxy]-2-carboxy-3-methylbenzyl-L-cysteinate **13**:

A mixture of methyl N-[N(tert-butoxycarbonyl)-L-seryl]-S-{4,6-bis[(tert-butyl)dimethylsiloxy]-3-methyl-2-[(4-nitrobenzyloxy)carbonyl]benzyl}-L-cysteinate **12** (630mg, 0.73mmol) and 10% Pd/c in EtOAc 10ml was shaken under 6-7 bar hydrogen at room temperature for 30 hours. The catalyst was removed by filtration through celite and the filtrate was evaporated in vacuo to afford a brown solid. Yield 90%. M.p. 164-167°C

$^1\text{H NMR}$  ( $\text{CDCl}_3$ ):  $\delta$  0.15 (12H, d, 2 x  $\text{Si}(\text{CH}_3)_2$ ), 0.95 (18H, d, 2 x  $\text{Si}(\text{CH}_3)_3$ ), 1.38 (9H, s, BOC), 2.03 (3H, s, Ar- $\text{CH}_3$ ), 2.93 (2H, br,  $\text{SCH}_2$ ), 3.64-3.67 (6H, m,  $\text{CH}_2\text{S}$ ,  $\text{OCH}_3$ ,  $\text{CHOH}$ ), 3.82 (1H, d,  $\text{CHOH}$ ), 4.08 (1H, m,  $\text{NHCH}$  (Ser)), 4.36 (1H, m,  $\text{CHOH}$ ), 4.75 (1H, br,  $\text{NHCH}$  (Cys)), 6.11 (1H, d,  $\text{NH}$ ), 6.19 (1H, s, Ar- $\text{CH}$ ), 8.01 (1H, d,  $\text{NH}$ );

$^{13}\text{C NMR}$  ( $\text{CDCl}_3$ ):  $\delta$  -4.4 ( $\text{Si}(\text{CH}_3)_2$ ), -4.1 ( $\text{Si}(\text{CH}_3)_2$ ), 13.3 ( $\text{CH}_3$ ), 18.1 ( $\text{Si}(\text{CH}_3)_3$ ), 25.6 ( $\text{Si}(\text{CH}_3)_3$ ), 25.7 ( $\text{Si}(\text{CH}_3)_3$ ), 28.1 ( $\text{C}(\text{CH}_3)_3$ ), 29.1 ( $\text{CH}_2\text{S}$ ), 29.6 ( $\text{CH}_2\text{S}$ ), 52.7 ( $\text{OCH}_3$ ), 54.2 ( $\text{NHCH}$ ), 62.1 ( $\text{CH}_2\text{OH}$ ), 77.1 ( $\text{COCH}$ ), 80.5 ( $\text{C}(\text{CH}_3)_3$ ), 108.4 (Ar- $\text{CH}$ ), 115.5 (Ar- $\text{CCH}_3$ ), 117.0 (Ar- $\text{CCH}_2\text{S}$ ), 152.0 (Ar- $\text{COSi}$ ), 153.2 (Ar- $\text{COSi}$ ), 156.2 ( $\text{COO}$ ), 166.5 ( $\text{COO}$ ), 166.6 ( $\text{COO}$ ), 171.2 ( $\text{C=O}$ );  
IR (KBr disc):  $\nu_{\text{max}}$  738, 835, 1521, 1683, 1716, 1747, 2929, 2954 $\text{cm}^{-1}$ .  
 $m/z$  731 [ $\text{M} + \text{H}$ ] $^+$

k). Synthesis of methyl (4R,7S)-7-[(tert-butoxycarbonyl)amino]-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **14**:

To a solution of methyl N-[N(tert-butoxycarbonyl)-L-seryl]-S-{4,6-bis[(tert-butyl)dimethylsiloxy]-2-carboxy-3-methylbenzyl-L-cysteinate **13** (1g, 1.4mmol) in toluene 20ml was added at 0°C  $\text{Ph}_3\text{P}$  (704mg, 2.7mmol) and DEAD (0.4ml, 2.7mmol). The mixture was stirred at 0°C for 40 minutes, followed by 5 hours at



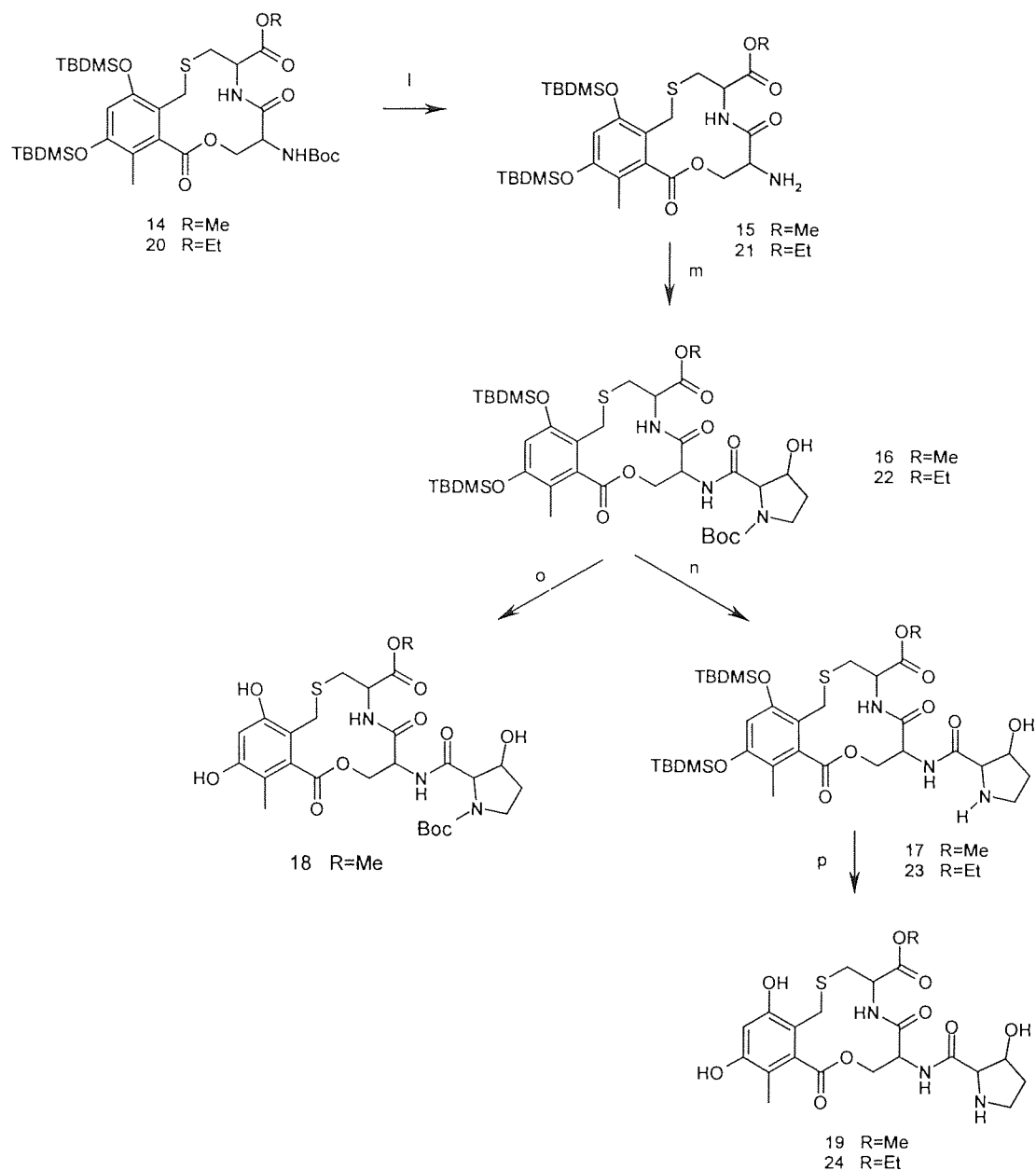
room temperature. The solvent was removed in vacuo. The residue was treated with DCM 20ml and stirred at 0°C for an hour. The yellow precipitate was removed by filtration. Purification with column chromatography with the eluant EtOAc: Hexane (1:3)  $R_f=0.25$  afforded the yellow solid. Yield 54%. M.p. 66-72°C

$^1\text{H NMR}$  ( $\text{CDCl}_3$ ):  $\delta$  0.19 (12H, d, 2 x  $\text{Si}(\underline{\text{CH}_3})_2$ ), 0.98 (18H, d, 2 x  $(\underline{\text{CH}_3})_3$ ), 1.48 (9H, s, BOC), 2.02 (3H, s,  $\underline{\text{CH}_3}$ ), 3.05 (2H, dd,  $\text{SCH}_2$ ), 3.29-3.34 (1H, d, Ph- $\underline{\text{CHS}}$ ), 3.74 (3H, s,  $\text{OCH}_3$ ), 3.85 (1H, d, Ph- $\underline{\text{CHS}}$ ), 4.24 (1H, d,  $\text{COOCH}$ ), 4.60 (1H, m,  $\text{CHNH}_2$ ), 4.83 (1H, m,  $\underline{\text{CHCOOCH}_3}$ ), 5.33 (1H, d,  $\text{COOCH}$ ), 5.74 (1H, d,  $\underline{\text{NH}}$ ), 6.32 (1H, s, Ar- $\underline{\text{CH}}$ ), 7.11 (1H, d,  $\underline{\text{NH}}$ );

$^{13}\text{C NMR}$  ( $\text{CDCl}_3$ ):  $\delta$  -4.3 ( $\text{Si}(\underline{\text{CH}_3})_2$ ), -4.2 ( $\text{Si}(\underline{\text{CH}_3})_2$ ), 13.2 ( $\underline{\text{CH}_3}$ ), 18.2 ( $\text{Si}(\underline{\text{C}}(\text{CH}_3)_3$ ), 25.5 ( $\text{Si}(\underline{\text{C}}(\text{CH}_3)_3$ ), 25.6 ( $\text{Si}(\underline{\text{C}}(\text{CH}_3)_3$ ), 28.1 ( $\text{C}(\underline{\text{CH}_3})_3$ ), 31.4 (Ph- $\underline{\text{CH}_2\text{S}}$ ), 34.6 ( $\underline{\text{CH}_2\text{S}}$ ), 51.7 ( $\underline{\text{NHCH}}$ ), 52.7 ( $\text{OCH}_3$ ), 65.8 ( $\underline{\text{CH}_2\text{OH}}$ ), 80.5 ( $\text{OC}(\underline{\text{CH}_3})_3$ ), 110.4 (Ar- $\underline{\text{CH}}$ ), 116.7 ( $\underline{\text{CCH}_2\text{S}}$ ), 118.9 ( $\underline{\text{CCH}_3}$ ), 135.0 ( $\underline{\text{CCOO}}$ ), 152.6 ( $\underline{\text{COSi}}$ ), 153.7 ( $\underline{\text{COSi}}$ ), 168.8 ( $\underline{\text{COO}}$ ), 169.0 ( $\underline{\text{COO}}$ ), 170.8 ( $\underline{\text{C=O}}$ );

IR (KBr disc):  $\nu_{\text{max}}$  1155, 1257, 1338, 1473, 1683, 1724, 2935, 2954 $\text{cm}^{-1}$ .

$m/z$  713  $[\text{M} + \text{H}]^+$



- l) TFA, anhydrous DCM, r.t, 40 mins.
- m) Boc-Hyp-OH, EDC, dried MeCN, 0°C, 5h.
- n) TFA, anhydrous DCM, r.t, 40 mins.
- o) 1M solution of TBAF in THF, THF, r.t, 1h.
- p) 1M solution of TBAF in THF, THF, r.t, 1h.

l) Synthesis of methyl (4R,7S)-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **15**:

To a solution of methyl (4R,7S)-7-[(tert-butoxycarbonyl)amino]-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **14** (100mg, 0.14 mmol) in anhydrous DCM 2ml was added at room temperature TFA 2ml. Stirring was continued at room temperature for 40 minutes. The solvent was removed in vacuo and the residue was dissolved in EtOAc 10ml, washed with saturated NaHCO<sub>3</sub> (2 x 5ml), brine (2 x 5ml), dried over NaSO<sub>4</sub>. The EtOAc was removed in vacuo to afford the yellow solid. Yield 71%. M.p. 74-77°C. lit (154-155°C, dec)

<sup>1</sup>H NMR (CDCl<sub>3</sub>): δ 0.18-0.24 (12H, d, 2 x Si(CH<sub>3</sub>)<sub>2</sub>), 0.97-1.00 (18H, d, 2 x (CH<sub>3</sub>)<sub>3</sub>), 2.10 (3H, s, CH<sub>3</sub>), 2.75-2.79 (1H, dd, CHS), 3.12-3.19 (1H, dd, CHS), 3.45-3.49 (1H, d, Ph-CHS), 3.74 (3H, s, OCH<sub>3</sub>), 3.75 (1H, m, CHNH<sub>2</sub>) 3.83-3.87 (1H, d, PhCHS), 4.30-4.41 (1H, m, PhCOOCH), 4.87 (1H, m, CHCOOCH<sub>3</sub>), 5.25-5.31 (1H, m, CHOCH<sub>2</sub>), 6.31 (1H, s, Ar-CH), 8.25 (1H, d, NH);

<sup>13</sup>C NMR (CDCl<sub>3</sub>): δ -4.3 (Si(CH<sub>3</sub>)<sub>2</sub>), -4.2 (Si(CH<sub>3</sub>)<sub>2</sub>), 13.1 (CH<sub>3</sub>), 18.2 (SiC(CH<sub>3</sub>)<sub>3</sub>), 25.5 (SiC(CH<sub>3</sub>)<sub>3</sub>), 25.7 (SiC(CH<sub>3</sub>)<sub>3</sub>), 30.8 (Ph-CH<sub>2</sub>S), 34.4 (CH<sub>2</sub>S), 51.8 (NHCH), 52.5 (OCH<sub>3</sub>), 67.5 (CH<sub>2</sub>O), 110.4 (Ar-CH), 116.5 (CCH<sub>2</sub>S), 118.8 (CCH<sub>3</sub>), 135.5 (CCOO), 152.6 (COSi), 153.7 (COSi), 168.9 (COO), 171.4 (COO), 171.8 (C=O);

IR (KBr disc):  $\nu_{\max}$  841, 1259, 1338, 1467, 1733, 2358, 2856, 2929cm<sup>-1</sup>.

m/z 613 [M + H]<sup>+</sup>

m) Synthesis of methyl (4R,7S)-7-{(3R)-1-(tert-Butoxycarbonyl)-3-hydroxy-L-prolyl}amino}-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **16**:

To an ice-cold mixture of methyl (4R,7S)-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **15** (500mg, 0.8 mmol) and Boc-Hyp-OH (207 mg, 0.9 mmol) in dried MeCN was added EDC (172 mg, 0.9 mmol). The mixture was stirred at 0°C for 5 hours. The solvent was removed in vacuo. The residue was diluted with EtOAc 20 ml, washed with 2M HCl (2 x 10ml), saturated NaHCO<sub>3</sub> (2 x 10ml) and brine (2 x 10ml). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>. Afterwards it was purified with column chromatography with the eluent EtOAc : Hexane (3:1) to afford the white solid. R<sub>f</sub>=0.3 in EtOAc. Yield 62%. M.p. 155-158°C

<sup>1</sup>H NMR (DMSO): δ 0.20-0.22 (12H, d, 2 x Si (CH<sub>3</sub>)<sub>2</sub>), 0.96-0.99 (18H, d, 2 x (CH<sub>3</sub>)<sub>3</sub>), 1.39 (9H, s, Boc), 1.42 (2H, m, Hyp CH<sub>2</sub>CHOH), 2.47 (3H, s, Ph-CH<sub>3</sub>), 2.80-2.87 (2H, m, Hyp NHCH<sub>2</sub>), 3.05-3.09 (2H, dd, SCH<sub>2</sub>), 3.49-3.69 (4H, m, Ph-CH<sub>2</sub>, Hyp COCH, Hyp CHOH), 3.62 (3H, s, OCH<sub>3</sub>) 3.96-4.21 (1H, m, PhCOOCH), 4.47 (PhCOOCH), 4.67 (1H, m, CHCOOEt), 5.11 (1H, dd, CHNH-Hyp), 6.48 (1H, s, Ph-H), 8.34 (1H, d, NHCO), 8.66 (1H, d, NHCO), 9.55 (2H, br, 2 x Ph-OH).

<sup>13</sup>C NMR (DMSO): δ -4.3 (Si (CH<sub>3</sub>)<sub>2</sub>), -4.2 (Si (CH<sub>3</sub>)<sub>3</sub>), 12.28 (Ar-CH<sub>3</sub>), 19.73 (Ar-CH<sub>2</sub>S), 25.5 (Si C(CH<sub>3</sub>)<sub>2</sub>), 25.6 (Si C(CH<sub>3</sub>)<sub>3</sub>), 26.34 (Hyp CH<sub>2</sub>CHOH), 28.9 (BOC CH<sub>3</sub>) 29.35 (SCH<sub>2</sub>CH), 32.87 (Hyp CH<sub>2</sub>NH), 52.09 (CHNHCO), 52.42 (CHCOO), 52.5 (OCH<sub>3</sub>), 59.01 (Hyp COCH), 60.92 (OCH<sub>2</sub>CH), 70.64 (Hyp CHOH), 81.1 (C(CH<sub>3</sub>)<sub>3</sub>) 103.33 (Ar-C), 110.45 (ArC-CH<sub>3</sub>), 112.18 (ArC-CH<sub>2</sub>), 135.01 (ArC-COO), 154.18 (ArC-OH), 155.22 (ArC-OH), 168.09 (Ar-COO), 168.76 (COOEt), 170.19 (Hyp, NHCO), 171.2 (Boc NHCO) 173.12 (NHCO).

IR (KBr disc): ν<sub>max</sub> (3348, 2958, 2847, 2368, 1739, 1668, 1588, 1522, 1464, 1389, 1335, 1251, 1163, 1039, 834, 777, 684, 573cm<sup>-1</sup>)

m/z 827 [M + H]<sup>+</sup>

n) Synthesis of methyl (4R,7S)-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-7-[[{(3R)-3-hydroxy-L-prolyl]amino}-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **17**:

To a solution of methyl (4R,7S)-7-[[{(3R)-1-(tert-Butoxycarbonyl)-3-hydroxy-L-prolyl]amino}-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **16** (100mg, 0.12 mmol) in anhydrous DCM 2ml was added at room temperature TFA 2ml. Stirring was continued at room temperature for 40 minutes. The solvent was removed in vacuum and the residue was dissolved in EtOAc 10ml, washed with saturated NaHCO<sub>3</sub> (2 x 5ml), brine (2 x 5ml), dried over NaSO<sub>4</sub>. The EtOAc was removed in vacuo to afford the white solid. Yield 67%. M.p. >140°C

<sup>1</sup>H NMR (DMSO): δ 0.20-0.22 (12H, d, 2 x Si (CH<sub>3</sub>)<sub>2</sub>), 0.96-0.99 (18H, d, 2 x (CH<sub>3</sub>)<sub>3</sub>), 1.42 (2H, m, Hyp CH<sub>2</sub>CHOH), 2.47 (3H, s, Ph-CH<sub>3</sub>), 2.80-2.87 (2H, m, Hyp NHCH<sub>2</sub>), 3.05-3.09 (2H, dd, SCH<sub>2</sub>), 3.49-3.69 (4H, m, Ph-CH<sub>2</sub>, Hyp COCH, Hyp CHOH), 3.62 (3H, s, OCH<sub>3</sub>) 3.96-4.21 (1H, m, PhCOOCH), 4.47 (PhCOOCH), 4.67 (1H, m, CHCOOEt), 5.11 (1H, dd, CHNH-Hyp), 6.48 (1H, s, Ph-H), 8.34 (1H, d, NHCO), 8.66 (1H, d, NHCO), 9.55 (2H, br, 2 x Ph-OH).

<sup>13</sup>C NMR (DMSO): δ -4.3 (Si (CH<sub>3</sub>)<sub>2</sub>), -4.2 (Si (CH<sub>3</sub>)<sub>3</sub>), 12.28 (Ar-CH<sub>3</sub>), 19.73 (Ar-CH<sub>2</sub>S), 25.5 (Si C(CH<sub>3</sub>)<sub>2</sub>), 25.6 (Si C(CH<sub>3</sub>)<sub>3</sub>), 26.34 (Hyp CH<sub>2</sub>CHOH), 29.35 (SCH<sub>2</sub>CH), 32.87 (Hyp CH<sub>2</sub>NH), 52.09 (CHNHCO), 52.42 (CHCOO), 52.5 (OCH<sub>3</sub>), 59.01 (Hyp COCH), 60.92 (OCH<sub>2</sub>CH), 70.64 (Hyp CHOH), 103.33 (Ar-C), 110.45 (ArC-CH<sub>3</sub>), 112.18 (ArC-CH<sub>2</sub>), 135.01 (ArC-COO), 154.18 (ArC-OH), 155.22 (ArC-OH), 168.09 (Ar-COO), 168.76 (COOEt), 170.19 (Hyp, NHCO), 173.12 (NHCO).

IR (KBr disc): ν<sub>max</sub> (3340, 2950, 2361, 1734, 1724, 1585, 1522, 1428, 1230, 1024, 853, 624cm<sup>-1</sup>)

m/z 727 [M + H]<sup>+</sup>

o) Synthesis of methyl (4R,7S)-7-{[(3R)-1-(tert-Butoxycarbonyl)-3-hydroxy-L-prolyl]amino}-12,14-dihydroxy-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **18**:

To a solution of methyl (4R,7S)-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-7-{[(3R)-3-hydroxy-L-prolyl]amino}-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **17** (200 mg, 0.24ml) in THF 2 ml at room temperature was added 1M solution of TBAF in THF (0.48 ml, 0.48 mmol). Stirring was continued for 1 hour at room temperature. The product was purified with column chromatography with the eluent CH<sub>2</sub>Cl<sub>2</sub>: CH<sub>3</sub>OH (10:1) to afford the white solid. R<sub>f</sub>=0.2. Yield = 70%. M.p. 129-131°C

<sup>1</sup>H NMR (DMSO): δ 1.39 (9H, s, Boc), 1.42 (2H, m, Hyp CH<sub>2</sub>CHOH), 2.47 (3H, s, Ph-CH<sub>3</sub>), 2.80-2.87 (2H, m, Hyp NHCH<sub>2</sub>), 3.05-3.09 (2H, dd, SCH<sub>2</sub>), 3.49-3.69 (4H, m, Ph-CH<sub>2</sub>, Hyp COCH, Hyp CHOH), 3.62 (3H, s, OCH<sub>3</sub>) 3.96-4.21 (1H, m, PhCOOCH), 4.47 (PhCOOCH), 4.67 (1H, m, CHCOOEt), 5.11 (1H, dd, CHNH-Hyp), 6.48 (1H, s, Ph-H), 8.34 (1H, d, NHCO), 8.66 (1H, d, NHCO), 9.55 (2H, br, 2 x Ph-OH).

<sup>13</sup>C NMR (DMSO): δ 12.28 (Ar-CH<sub>3</sub>), 19.73 (Ar-CH<sub>2</sub>S), 26.34 (Hyp CH<sub>2</sub>CHOH), 28.9 (BOC CH<sub>3</sub>) 29.35 (SCH<sub>2</sub>CH), 32.87 (Hyp CH<sub>2</sub>NH), 52.09 (CHNHCO), 52.42 (CHCOO), 52.5 (OCH<sub>3</sub>), 59.01 (Hyp COCH), 60.92 (OCH<sub>2</sub>CH), 70.64 (Hyp CHOH), 81.1 (C(CH<sub>3</sub>)<sub>3</sub>) 103.33 (Ar-C), 110.45 (ArC-CH<sub>3</sub>), 112.18 (ArC-CH<sub>2</sub>), 135.01 (ArC-COO), 154.18 (ArC-OH), 155.22 (ArC-OH), 168.09 (Ar-COO), 168.76 (COOEt), 170.19 (Hyp, NHCO), 171.2 (Boc NHCO) 173.12 (NHCO).

IR (KBr disc): ν<sub>max</sub> (3332, 2959, 2864, 2354, 1727, 1677, 1600, 1518, 1400, 1223, 1168, 1032, 854, 582cm<sup>-1</sup>)

m/z 598 [M + H]<sup>+</sup>

HRMS (ES<sup>+</sup>): m/z 598.2070 (M+H<sup>+</sup>, C<sub>26</sub>H<sub>35</sub>N<sub>3</sub>O<sub>11</sub>S requires 598.2072).

p) Synthesis of methyl (4R,7S)-12,14-dihydroxy-1,3,4,5,6,7,8,10-octahydro-7-  
-{[(3R)-3-hydroxy-L-prolyl]amino}-11-methyl-6,10-dioxo-9,2,5-  
benzoxathiazacyclododecine-4-carboxylate **19**:

To a solution of methyl (4R,7S)-12,14-bis[(tert-butyl)dimethylsiloxy]-  
1,3,4,5,6,7,8,10-octahydro-7-{[(3R)-3-hydroxy-L-prolyl]amino}-11-methyl-6,10-  
dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **17** (300 mg, 0.41 mmol)  
in THF was added 1M solution of TBAF in THF (0.84 ml, 0.84 mmol) at room  
temperature. Stirring was continued for 1 hour. The residue was purified with  
column chromatography with the eluent EtOAc:MeOH (2:1).  $R_f$  = 0.25. Yield =  
60%. M.p. 191-194°C

$^1\text{H}$  NMR (DMSO):  $\delta$  1.42 (2H, m, Hyp  $\text{CH}_2\text{CHOH}$ ), 2.47 (3H, s, Ph- $\text{CH}_3$ ), 2.80-  
2.87 (2H, m, Hyp  $\text{NHCH}_2$ ), 3.05-3.09 (2H, dd,  $\text{SCH}_2$ ), 3.49-3.69 (4H, m, Ph- $\text{CH}_2$ ,  
Hyp  $\text{COCH}$ , Hyp  $\text{CHOH}$ ), 3.62 (3H, s,  $\text{OCH}_3$ ) 3.96-4.21 (1H, m,  $\text{PhCOOCH}$ ),  
4.47 ( $\text{PhCOOCH}$ ), 4.67 (1H, m,  $\text{CHCOOEt}$ ), 5.11 (1H, dd,  $\text{CHNH-Hyp}$ ), 6.48  
(1H, s, Ph- $\text{H}$ ), 8.34 (1H, d,  $\text{NHCO}$ ), 8.66 (1H, d,  $\text{NHCO}$ ), 9.55 (2H, br, 2 x Ph-  
 $\text{OH}$ ).

$^{13}\text{C}$  NMR (DMSO):  $\delta$  12.28 (Ar- $\text{CH}_3$ ), 19.73 (Ar- $\text{CH}_2\text{S}$ ), 26.34 (Hyp  $\text{CH}_2\text{CHOH}$ ),  
29.35 ( $\text{SCH}_2\text{CH}$ ), 32.87 (Hyp  $\text{CH}_2\text{NH}$ ), 52.09 ( $\text{CHNHCO}$ ), 52.42 ( $\text{CHCOO}$ ), 52.5  
( $\text{OCH}_3$ ), 59.01 (Hyp  $\text{COCH}$ ), 60.92 ( $\text{OCH}_2\text{CH}$ ), 70.64 (Hyp  $\text{CHOH}$ ), 103.33 (Ar-  
 $\text{C}$ ), 110.45 (Ar $\text{C-CH}_3$ ), 112.18 (Ar $\text{C-CH}_2$ ), 135.01 (Ar $\text{C-COO}$ ), 154.18 (Ar $\text{C-OH}$ ),  
155.22 (Ar $\text{C-OH}$ ), 168.09 (Ar- $\text{COO}$ ), 168.76 ( $\text{COOEt}$ ), 170.19 (Hyp,  $\text{NHCO}$ ),  
173.12 ( $\text{NHCO}$ ).

$m/z$  498  $[\text{M} + \text{H}]^+$

HRMS ( $\text{ES}^+$ ):  $m/z$  498.1546 ( $\text{M} + \text{H}^+$ ,  $\text{C}_{21}\text{H}_{27}\text{N}_3\text{O}_9\text{S}$  requires 498.1547).

k) Synthesis of ethyl (4R,7S)-7-[(tert-butoxycarbonyl)amino]-12,14-bis[(tert-  
butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-  
benzoxathiazacyclododecine-4-carboxylate **20** :

As described for **14**. Yield 42%. M p is not available.

$^1\text{H NMR}$  ( $\text{CDCl}_3$ ):  $\delta$  0.18 (12H, d, 2 x  $\text{Si}(\text{CH}_3)_2$ ), 0.97 (18H, d, 2 x  $(\text{CH}_3)_3$ ), 1.16 (3H, t,  $\text{CH}_2\text{CH}_3$ ), 1.39 (9H, s, Boc), 2.01 (3H, s,  $\text{CH}_3$ ) 2.75-2.79 (1H, dd,  $\text{CHS}$ ), 3.12-3.19 (1H, dd,  $\text{CHS}$ ), 3.45-3.49 (1H, d,  $\text{Ph-CHS}$ ), 3.74 (3H, s,  $\text{OCH}_3$ ), 3.83-3.87 (1H, d,  $\text{PhCHS}$ ), 4.00 ( $\text{COOCH}_2$ ), 4.30-4.41 (1H, m,  $\text{COOCH}$ ), 4.68 (1H, m,  $\text{CHCOOEt}$ ), 4.87 (1H, dd,  $\text{CHNHBOc}$ ), 5.25-5.31 (1H, dd,  $\text{CHOCH}_2$ ), 6.31 (1H, s,  $\text{Ar-CH}$ ), 8.25 (1H, d,  $\text{NH}$ );

$^{13}\text{C NMR}$  ( $\text{CDCl}_3$ ):  $\delta$  -4.3 ( $\text{Si}(\text{CH}_3)_2$ ), -4.2 ( $\text{Si}(\text{CH}_3)_2$ ), 13.1 ( $\text{CH}_3$ ), 13.74 ( $\text{CH}_2\text{CH}_3$ ), 18.2 ( $\text{SiC}(\text{CH}_3)_3$ ), 25.5 ( $\text{SiC}(\text{CH}_3)_3$ ), 25.7 ( $\text{SiC}(\text{CH}_3)_3$ ), 28.9 (Boc  $\text{CH}_3$ ), 30.8 ( $\text{Ph-CH}_2\text{S}$ ), 34.4 ( $\text{CH}_2\text{S}$ ), 51.8 ( $\text{NHCH}$ ), 54.50 ( $\text{OCH}_2\text{CH}_3$ ), 67.5 ( $\text{CH}_2\text{O}$ ), 81.1 ( $\text{C}(\text{CH}_3)_3$ ), 110.4 ( $\text{Ar-CH}$ ), 116.5 ( $\text{CCH}_2\text{S}$ ), 118.8 ( $\text{CCH}_3$ ), 135.5 ( $\text{CCOO}$ ), 152.6 ( $\text{COSi}$ ), 153.7 ( $\text{COSi}$ ), 168.9 ( $\text{COO}$ ), 171.2 (Boc  $\text{NHCO}$ ) 171.4 ( $\text{COO}$ ), 171.8 ( $\text{C=O}$ );

IR (KBr disc):  $\nu_{\text{max}}$  (2954, 2935, 1724, 1683, 1473, 1338, 1257, 1155  $\text{cm}^{-1}$ ).

$m/z$  727  $[\text{M} + \text{H}]^+$

l) Synthesis of ethyl (4R,7S)-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **21**:

As described for **15**. Yield and m p are not available.

$^1\text{H NMR}$  ( $\text{CDCl}_3$ ):  $\delta$  0.18 (12H, d, 2 x  $\text{Si}(\text{CH}_3)_2$ ), 0.97 (18H, d, 2 x  $(\text{CH}_3)_3$ ), 1.16 (3H, t,  $\text{CH}_2\text{CH}_3$ ), 2.01 (3H, s,  $\text{Ar-CH}_3$ ) 2.75-2.79 (1H, dd,  $\text{CHS}$ ), 3.12-3.19 (1H, dd,  $\text{CHS}$ ), 3.45-3.49 (1H, d,  $\text{Ph-CHS}$ ), 3.83-3.87 (1H, d,  $\text{PhCHS}$ ), 4.00 ( $\text{COOCH}_2$ ), 4.30-4.41 (1H, dd,  $\text{PhCOOCH}$ ), 4.68 (1H, m,  $\text{CHCOOEt}$ ) 4.87 (1H, dd,  $\text{CHNH}$ ), 5.25-5.31 (1H, dd,  $\text{PhCOOCH}$ ), 6.31 (1H, s,  $\text{Ar-CH}$ ), 8.25 (1H, d,  $\text{NH}$ );

$^{13}\text{C NMR}$  ( $\text{CDCl}_3$ ):  $\delta$  -4.3 ( $\text{Si}(\text{CH}_3)_2$ ), -4.2 ( $\text{Si}(\text{CH}_3)_2$ ), 13.1 ( $\text{CH}_3$ ), 13.74 ( $\text{CH}_2\text{CH}_3$ ), 18.2 ( $\text{SiC}(\text{CH}_3)_3$ ), 25.5 ( $\text{SiC}(\text{CH}_3)_3$ ), 25.7 ( $\text{SiC}(\text{CH}_3)_3$ ), 30.8 ( $\text{Ph-CH}_2\text{S}$ ), 34.4 ( $\text{CH}_2\text{S}$ ), 51.8 ( $\text{NHCH}$ ), 54.50 ( $\text{OCH}_2\text{CH}_3$ ), 67.5 ( $\text{CH}_2\text{O}$ ), 110.4 ( $\text{Ar-CH}$ ), 116.5 ( $\text{CCH}_2\text{S}$ ), 118.8 ( $\text{CCH}_3$ ), 135.5 ( $\text{CCOO}$ ), 152.6 ( $\text{COSi}$ ), 153.7 ( $\text{COSi}$ ), 168.9 ( $\text{COO}$ ), 171.4 ( $\text{COO}$ ), 171.8 ( $\text{C=O}$ );



IR (KBr disc):  $\nu_{\max}$  (2929, 2856, 2358, 1733, 1683, 1476, 1338, 1259, 841  $\text{cm}^{-1}$ ).  
 $m/z$  627  $[\text{M} + \text{H}]^+$

m) Synthesis of methyl (4R,7S)-7- $\{[(3R)$ -1-(tert-Butoxycarbonyl)-3-hydroxy-L-prolyl]amino $\}$ -12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **22**:

As described for **16**. Yield and m p are not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  0.20-0.22 (12H, d, 2 x Si ( $\text{CH}_3$ )<sub>2</sub>), 0.96-0.99 (18H, d, 2 x ( $\text{CH}_3$ )<sub>3</sub>), 1.16 (3H, t,  $\text{CH}_2\text{CH}_3$ ), 1.39 (9H, s, Boc), 1.48 (2H, m, Hyp  $\text{CH}_2\text{CHOH}$ ), 2.48 (3H, s, Ph- $\text{CH}_3$ ), 2.71-2.87 (2H, m, Hyp  $\text{NHCH}_2$ ), 3.05-3.09 (2H, dd,  $\text{SCH}_2$ ), 3.49-3.69 (4H, m, Ph- $\text{CH}_2$ , Hyp  $\text{COCH}$ , Hyp  $\text{CHOH}$ ), 3.97-4.21 (3H, m,  $\text{OCH}_2\text{CH}_3$ , Ph $\text{COOCH}$ ), 4.38 (Ph $\text{COOCH}$ ), 4.68 (1H, m,  $\text{CHCOOEt}$ ), 5.07 (1H, dd,  $\text{CHNH}$ -Hyp), 6.48 (1H, s, Ph- $\text{H}$ ), 8.34 (1H, d,  $\text{NHCO}$ ), 8.66 (1H, d,  $\text{NHCO}$ ), 9.63 (2H, br, 2 x Ph- $\text{OH}$ ).

$^{13}\text{C}$  NMR (DMSO):  $\delta$  -4.3 (Si ( $\text{CH}_3$ )<sub>2</sub>), -4.2 (Si ( $\text{CH}_3$ )<sub>3</sub>), 12.28 (Ar- $\text{CH}_3$ ), 13.74 ( $\text{CH}_2\text{CH}_3$ ), 19.73 (Ar- $\text{CH}_2\text{S}$ ), 25.5 (Si C( $\text{CH}_3$ )<sub>2</sub>), 25.6 (Si C( $\text{CH}_3$ )<sub>3</sub>), 26.34 (Hyp  $\text{CH}_2\text{CHOH}$ ), 28.9 (Boc  $\text{CH}_3$ ), 29.35 ( $\text{SCH}_2\text{CH}$ ), 32.87 (Hyp  $\text{CH}_2\text{NH}$ ), 52.09 ( $\text{CHNHCO}$ ), 52.42 ( $\text{CHCOOEt}$ ), 54.50 ( $\text{OCH}_2\text{CH}_3$ ), 59.01 (Hyp  $\text{COCH}$ ), 60.92 ( $\text{OCH}_2\text{CH}$ ), 70.64 (Hyp  $\text{CHOH}$ ), 81.1 ( $\text{C}(\text{CH}_3)_3$ ), 103.33 (Ar- $\text{C}$ ), 110.45 (Ar $\text{C}-\text{CH}_3$ ), 112.18 (Ar $\text{C}-\text{CH}_2$ ), 135.01 (Ar $\text{C}-\text{COO}$ ), 154.18 (Ar $\text{C}-\text{OH}$ ), 155.22 (Ar $\text{C}-\text{OH}$ ), 168.09 (Ar- $\text{COO}$ ), 168.76 ( $\text{COOEt}$ ), 170.19 (Hyp,  $\text{NHCO}$ ), 171.2 (Boc  $\text{NHCO}$ ), 173.12 ( $\text{NHCO}$ ).

IR (KBr disc):  $\nu_{\max}$  (3355, 2954, 2859, 2704, 1736, 1586, 1514, 1473, 1382, 1332, 1254, 1150, 1032, 832, 782, 682, 564, 536  $\text{cm}^{-1}$ )  
 $m/z$  841  $[\text{M} + \text{H}]^+$

n) Synthesis of methyl (4R,7S)-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-7- $\{[(3R)$ -3-hydroxy-L-prolyl]amino $\}$ -11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **23**:

As described for **17**. Yield and m p are not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  0.20-0.22 (12H, d, 2 x Si ( $\text{CH}_3$ )<sub>2</sub>), 0.96-0.99 (18H, d, 2 x ( $\text{CH}_3$ )<sub>3</sub>), 1.16 (3H, t,  $\text{CH}_2\text{CH}_3$ ), 1.48 (2H, m, Hyp  $\text{CH}_2\text{CHOH}$ ), 2.48 (3H, s, Ph- $\text{CH}_3$ ), 2.71-2.87 (2H, m, Hyp  $\text{NHCH}_2$ ), 3.05-3.09 (2H, dd,  $\text{SCH}_2$ ), 3.49-3.69 (4H, m, Ph- $\text{CH}_2$ , Hyp  $\text{COCH}$ , Hyp  $\text{CHOH}$ ), 3.97-4.21 (3H, m,  $\text{OCH}_2\text{CH}_3$ ,  $\text{PhCOOCH}$ ), 4.38 ( $\text{PhCOOCH}$ ), 4.68 (1H, m,  $\text{CHCOOEt}$ ), 5.07 (1H, dd,  $\text{CHNH-Hyp}$ ), 6.48 (1H, s, Ph- $\text{H}$ ), 8.34 (1H, d,  $\text{NHCO}$ ), 8.66 (1H, d,  $\text{NHCO}$ ), 9.63 (2H, br, 2 x Ph- $\text{OH}$ ).

$^{13}\text{C}$  NMR (DMSO):  $\delta$  -4.3 (Si ( $\text{CH}_3$ )<sub>2</sub>), -4.2 (Si ( $\text{CH}_3$ )<sub>3</sub>), 12.28 (Ar- $\text{CH}_3$ ), 13.74 ( $\text{CH}_2\text{CH}_3$ ), 19.73 (Ar- $\text{CH}_2\text{S}$ ), 25.5 (Si C( $\text{CH}_3$ )<sub>2</sub>), 25.6 (Si C( $\text{CH}_3$ )<sub>3</sub>), 26.34 (Hyp  $\text{CH}_2\text{CHOH}$ ), 29.35 ( $\text{SCH}_2\text{CH}$ ), 32.87 (Hyp  $\text{CH}_2\text{NH}$ ), 52.09 ( $\text{CHNHCO}$ ), 52.42 ( $\text{CHCOOEt}$ ), 54.50 ( $\text{OCH}_2\text{CH}_3$ ), 59.01 (Hyp  $\text{COCH}$ ), 60.92 ( $\text{OCH}_2\text{CH}$ ), 70.64 (Hyp  $\text{CHOH}$ ), 103.33 (Ar- $\text{C}$ ), 110.45 (Ar- $\text{C-CH}_3$ ), 112.18 (Ar- $\text{C-CH}_2$ ), 135.01 (Ar- $\text{C-COO}$ ), 154.18 (Ar- $\text{C-OH}$ ), 155.22 (Ar- $\text{C-OH}$ ), 168.09 (Ar- $\text{C-COO}$ ), 168.76 ( $\text{COOEt}$ ), 170.19 (Hyp,  $\text{NHCO}$ ), 173.12 ( $\text{NHCO}$ ).

$m/z$  741 [ $\text{M} + \text{H}$ ]<sup>+</sup>

p) Synthesis of ethyl (4R,7S)-12,14-dihydroxy-1,3,4,5,6,7,8,10-octahydro-7-  
{[(3R)-3-hydroxy-L-prolyl]amino}-11-methyl-6,10-dioxo-9,2,5-  
benzoxathiazacyclododecine-4-carboxylate **24**:

As described for **19**. Yield and m p are not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  1.16 (3H, t,  $\text{CH}_2\text{CH}_3$ ), 1.48 (2H, m, Hyp  $\text{CH}_2\text{CHOH}$ ), 2.48 (3H, s, Ph- $\text{CH}_3$ ), 2.71-2.87 (2H, m, Hyp  $\text{NHCH}_2$ ), 3.05-3.09 (2H, dd,  $\text{SCH}_2$ ), 3.49-3.69 (4H, m, Ph- $\text{CH}_2$ , Hyp  $\text{COCH}$ , Hyp  $\text{CHOH}$ ), 3.97-4.21 (3H, m,  $\text{OCH}_2\text{CH}_3$ ,  $\text{PhCOOCH}$ ), 4.38 ( $\text{PhCOOCH}$ ), 4.68 (1H, m,  $\text{CHCOOEt}$ ), 5.07 (1H, dd,  $\text{CHNH-Hyp}$ ), 6.48 (1H, s, Ph- $\text{H}$ ), 8.34 (1H, d,  $\text{NHCO}$ ), 8.66 (1H, d,  $\text{NHCO}$ ), 9.63 (2H, br, 2 x Ph- $\text{OH}$ ).

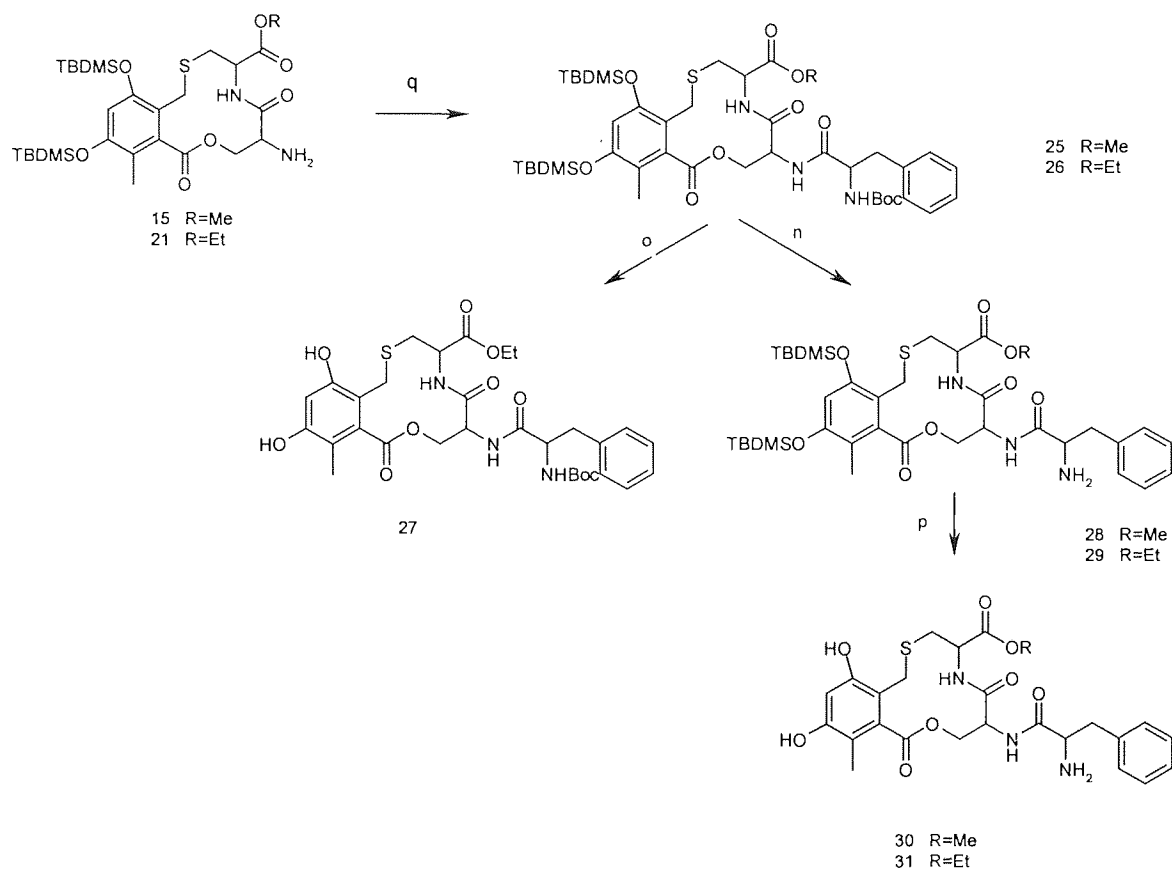
$^{13}\text{C}$  NMR (DMSO):  $\delta$  12.28 (Ar- $\text{CH}_3$ ), 13.74 ( $\text{CH}_2\text{CH}_3$ ), 19.73 (Ar- $\text{CH}_2\text{S}$ ), 26.34 (Hyp  $\text{CH}_2\text{CHOH}$ ), 29.35 ( $\text{SCH}_2\text{CH}$ ), 32.87 (Hyp  $\text{CH}_2\text{NH}$ ), 52.09 ( $\text{CHNHCO}$ ), 52.42 ( $\text{CHCOOEt}$ ), 54.50 ( $\text{OCH}_2\text{CH}_3$ ), 59.01 (Hyp  $\text{COCH}$ ), 60.92 ( $\text{OCH}_2\text{CH}$ ),

70.64 (Hyp  $\underline{\text{C}}\text{HOH}$ ), 103.33 (Ar- $\underline{\text{C}}$ ), 110.45 (Ar $\underline{\text{C}}$ -CH<sub>3</sub>), 112.18 (Ar $\underline{\text{C}}$ -CH<sub>2</sub>),  
135.01 (Ar $\underline{\text{C}}$ -COO), 154.18 (Ar $\underline{\text{C}}$ -OH), 155.22 (Ar $\underline{\text{C}}$ -OH), 168.09 (Ar- $\underline{\text{C}}\text{OO}$ ),  
168.76 ( $\underline{\text{C}}\text{OOEt}$ ), 170.19 (Hyp, NH $\underline{\text{C}}\text{O}$ ), 173.12 (NH $\underline{\text{C}}\text{O}$ ).

IR (KBr disc):  $\nu_{\text{max}}$  (3309, 2963, 2366, 1733, 1648, 1594, 1549, 1235, 1105, 1019,  
975, 853, 709 $\text{cm}^{-1}$ )

$m/z$  512 [M + H]<sup>+</sup>

HRMS (ES<sup>+</sup>):  $m/z$  521.1703 (M+H<sup>+</sup>, C<sub>22</sub>H<sub>29</sub>N<sub>3</sub>O<sub>9</sub>S requires 521.1035).



n) TFA, anhydrous DCM, r.t, 40 mins.

o) 1M solution of TBAF in THF, THF, r.t, 1h.

p) 1M solution of TBAF in THF, THF, r.t, 1h.

q) Boc-D-Phe-OH, EDC, dried MeCN, 0°C, 5h.

q) Synthesis of methyl (4R,7S)-7-[[[(3R)-1-(tert-butoxycarbonyl)-D-phenylalanine]-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **25**:

To an ice-cold mixture of methyl (4R,7S)-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **15** (450mg, 0.72mmol) and Boc-D-Phe-OH (210mg, 0.8mmol) in dried MeCN was added EDC (153mg, 0.8mmol). The mixture was stirred at 0°C for 5 hours. The solvent was removed in vacuo. The residue was diluted with EtOAc 20 ml, washed with 2M HCl (2 x 10ml), saturated NaHCO<sub>3</sub> (2 x 10ml) and brine (2 x 10ml). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and purified by column chromatography with the eluent EtOAc: hexane (1: 2). R<sub>f</sub>=0.2 to afford the white solid. Yield 67%. M.p. 105-107°C

<sup>1</sup>H NMR (CDCl<sub>3</sub>): δ 0.22 (12H, d, 2 x Si(CH<sub>3</sub>)<sub>2</sub>), 0.97 (18H, d, 2 x (CH<sub>3</sub>)<sub>3</sub>), 1.23 (9H, s, Boc), 2.48 (3H, s, PhCH<sub>3</sub>), 2.75-2.79 (1H, dd, CHS), 3.09 (2H, m, CH<sub>2</sub>Ph in Phe), 3.12-3.19 (1H, dd, CHS), 3.45-3.49 (1H, d, Ph-CHS), 3.62 (3H, s, OCH<sub>3</sub>), 3.83-3.87 (1H, d, PhCHS), 4.30-4.41 (1H, dd, PhCOOCH), 4.49 (1H, m, CHCH<sub>2</sub>Ph), 4.64 (1H, dd, CHNH), 4.68 (1H, m, CHCOO), 5.25-5.31 (1H, dd, CHOCH<sub>2</sub>), 6.31 (1H, s, Ar-CH), 7.27 (5H, m, 5 x Ph), 8.16 (1H, d, NHCO), 8.25 (1H, d, NHCO);

m/z 861 [M + H]<sup>+</sup>

q) Synthesis of ethyl (4R,7S)-7-[[[(3R)-1-(tert-butoxycarbonyl)-D-phenylalanine]-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **26**:

As described for **25**. Yield and m p are not available.

<sup>1</sup>H NMR (CDCl<sub>3</sub>): δ 0.22-0.24 (12H, d, 2 x Si(CH<sub>3</sub>)<sub>2</sub>), 0.97-1.00 (18H, d, 2 x (CH<sub>3</sub>)<sub>3</sub>), 1.16 (3H, t, CH<sub>2</sub>CH<sub>3</sub>), 1.23 (9H, s, Boc), 2.48 (3H, s, PhCH<sub>3</sub>), 2.75-2.79

(1H, dd, CHS), 3.09 (2H, m, CH<sub>2</sub>Ph in Phe), 3.12-3.19 (1H, dd, CHS), 3.45-3.49 (1H, d, Ph-CHS), 3.83-3.87 (1H, d, PhCHS), 4.00 (2H, m, COOCH<sub>2</sub>), 4.30-4.41 (1H, dd, PhCOOCH), 4.49 (1H, m, CHCH<sub>2</sub>Ph), 4.64 (1H, dd, CHNH), 4.68 (1H, m, CHCOO), 5.25-5.31 (1H, dd, CHOCH<sub>2</sub>), 6.31 (1H, s, Ar-CH), 7.27 (5H, m, 5 x Ph), 8.16 (1H, d, NHCO), 8.25 (1H, d, NHCO);

IR (KBr disc):  $\nu_{\max}$  (3335, 2958, 2932, 2851, 2366, 1733, 1589, 1527, 1459, 1338, 1253, 1150, 1033, 840, 773, 595, 512cm<sup>-1</sup>)

m/z 875 [M + H]<sup>+</sup>

o) Synthesis of ethyl (4R,7S)-7-{[(3R)-1-(tert-butoxycarbonyl)-D-phenylalanine]-12,14-dihydroxy -1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **27**:

To a solution of methyl (4R,7S)-7-{[(3R)-1-(tert-butoxycarbonyl)-D-phenylalanine]-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **25** (100mg, 0.11mmol) in THF 2 ml at room temperature was added 1M solution of TBAF in THF ( 0.22ml, 0.22mmol). Stirring was continued for 1 hour at room temperature. The product was purified by column chromatography with the eluent CH<sub>2</sub>Cl<sub>2</sub>: CH<sub>3</sub>OH (10: 1) R<sub>f</sub>=0.33 to afford the white solid. Yield 53%. M.p. 128-130°C

<sup>1</sup>H NMR (CDCl<sub>3</sub>):  $\delta$  1.16 (3H, t, CH<sub>2</sub>CH<sub>3</sub>), 1.23 (9H, s, Boc), 2.48 (3H, s, PhCH<sub>3</sub>), 2.75-2.79 (1H, dd, CHS), 3.09 (2H, m, CH<sub>2</sub>Ph in Phe), 3.12-3.19 (1H, dd, CHS), 3.45-3.49 (1H, d, Ph-CHS), 3.83-3.87 (1H, d, PhCHS), 4.00 (2H, m, COOCH<sub>2</sub>), 4.30-4.41 (1H, dd, PhCOOCH), 4.49 (1H, m, CHCH<sub>2</sub>Ph), 4.64 (1H, dd, CHNH), 4.68 (1H, m, CHCOO), 5.25-5.31 (1H, dd, CHOCH<sub>2</sub>), 6.31 (1H, s, Ar-CH), 7.27 (5H, m, 5 x Ph), 8.16 (1H, d, NHCO), 8.25 (1H, d, NHCO), 9.54 (2H, br, 2 x OH);  
IR (KBr disc):  $\nu_{\max}$  (3353, 2981, 2932, 2361, 1742, 1661, 1612, 1509, 1365, 1244, 1159, 1096, 1019, 849, 737, 683cm<sup>-1</sup>)  
m/z 646 [M + H]<sup>+</sup>

HRMS (ES<sup>+</sup>): *m/z* 646.2434 (M+H<sup>+</sup>, C<sub>31</sub>H<sub>39</sub>N<sub>3</sub>O<sub>10</sub>S requires 646.2436).

n) Synthesis of methyl (4R,7S)-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-7-[(3R)-D-phenylalanine]-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **28**:

To a solution of methyl (4R,7S)-7-{[(3R)-1-(tert-butoxycarbonyl)-D-phenylalanine]-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **25** (100mg, 0.11 mmol) in anhydrous DCM 2ml was added at room temperature TFA 2ml. Stirring was continued at room temperature for 40 minutes. The solvent was removed in vacuo and the residue was dissolved in EtOAc 10ml, washed with saturated NaHCO<sub>3</sub> (2 x 5ml), brine (2 x 5ml), dried over NaSO<sub>4</sub>. The EtOAc was removed in vacuo to afford the white solid. R<sub>f</sub>=0.2 in EtOAc. Yield 72%. M.p. 120-122°C

<sup>1</sup>H NMR (CDCl<sub>3</sub>): δ 0.22 (12H, d, 2 x Si(CH<sub>3</sub>)<sub>2</sub>), 0.97 (18H, d, 2 x (CH<sub>3</sub>)<sub>3</sub>), 2.48 (3H, s, PhCH<sub>3</sub>), 2.75-2.79 (1H, dd, CHS), 3.09 (2H, m, CH<sub>2</sub>Ph in Phe), 3.12-3.19 (1H, dd, CHS), 3.45-3.49 (1H, d, Ph-CHS), 3.62 (3H, s, OCH<sub>3</sub>), 3.83-3.87 (1H, d, PhCHS), 4.30-4.41 (1H, dd, PhCOOCH), 4.49 (1H, m, CHCH<sub>2</sub>Ph), 4.64 (1H, dd, CHNH), 4.68 (1H, m, CHCOO), 5.25-5.31 (1H, dd, PhCOOCH), 6.31 (1H, s, Ar-CH), 7.27 (5H, m, 5 x Ph), 8.16 (1H, d, NHCO), 8.25 (1H, d, NHCO);  
IR (KBr disc): ν<sub>max</sub> (3349, 2958, 2927, 2860, 2366, 1747, 1675, 1589, 1527, 1468, 1347, 1244, 1150, 1033, 912, 840, 768, 692cm<sup>-1</sup>)

n) Synthesis of ethyl (4R,7S)-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-7-[(3R)-D-phenylalanine]-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **29**:

As described for **28**. Yield and m p are not available.

$^1\text{H}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  0.22 (12H, d, 2 x  $\text{Si}(\text{CH}_3)_2$ ), 0.97 (18H, d, 2 x  $(\text{CH}_3)_3$ ), 1.16 (3H, t,  $\text{CH}_2\text{CH}_3$ ), 2.75-2.79 (1H, dd,  $\text{CHS}$ ), 3.09 (2H, m,  $\text{CH}_2\text{Ph}$  in Phe), 3.12-3.19 (1H, dd,  $\text{CHS}$ ), 3.45-3.49 (1H, d, Ph- $\text{CHS}$ ), 3.83-3.87 (1H, d, Ph $\text{CHS}$ ), 4.00 (2H, m,  $\text{COOCH}_2$ ), 4.30-4.41 (1H, dd, Ph $\text{COOCH}$ ), 4.49 (1H, m,  $\text{CHCH}_2\text{Ph}$ ), 4.64 (1H, dd,  $\text{CHNH}$ ), 4.68 (1H, m,  $\text{CHCOO}$ ), 5.25-5.31 (1H, dd, Ph $\text{COOCH}$ ), 6.31 (1H, s, Ar- $\text{CH}$ ), 7.27 (5H, m, 5 x Ph), 8.16 (1H, d,  $\text{NHCO}$ ), 8.25 (1H, d,  $\text{NHCO}$ );  
IR (KBr disc):  $\nu_{\text{max}}$  (3348, 2958, 2925, 2843, 2361, 1736, 1676, 1603, 1521, 1497, 1332, 1245, 1195, 1029, 822, 772, 694 $\text{cm}^{-1}$ )

p) Synthesis of methyl (4R,7S)-12,14-dihydroxy-1,3,4,5,6,7,8,10-octahydro-7-[(3R)-D-phenylalanine]-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **30**:

To a solution of methyl (4R,7S)-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-7-[(3R)-D-phenylalanine]-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **28** (170mg, 0.22mmol) in THF 2ml was added 1M solution of TBAF in THF (0.44ml, 0.44mmol) at room temperature. Stirring was continued for 1 hour. The product was purified by column chromatography with the eluent EtOAc:  $\text{CH}_3\text{OH}$  (5: 1)  $R_f=0.2$  to afford the white solid. Yield 44 %. M.p. 112-115 $^\circ\text{C}$

$^1\text{H}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  2.48 (3H, s, Ph $\text{CH}_3$ ), 2.75-2.79 (1H, dd,  $\text{CHS}$ ), 3.09 (2H, m,  $\text{CH}_2\text{Ph}$  in Phe), 3.12-3.19 (1H, dd,  $\text{CHS}$ ), 3.45-3.49 (1H, d, Ph- $\text{CHS}$ ), 3.62 (3H, s,  $\text{OCH}_3$ ), 3.83-3.87 (1H, d, Ph $\text{CHS}$ ), 4.30-4.41 (1H, dd, Ph $\text{COOCH}$ ), 4.49 (1H, m,  $\text{CHCH}_2\text{Ph}$ ), 4.64 (1H, dd,  $\text{CHNH}$ ), 4.68 (1H, m,  $\text{CHCOO}$ ), 5.25-5.31 (1H, dd, Ph $\text{COOCH}$ ), 6.31 (1H, s, Ar- $\text{CH}$ ), 7.27 (5H, m, 5 x Ph), 8.16 (1H, d,  $\text{NHCO}$ ), 8.25 (1H, d,  $\text{NHCO}$ );  
IR (KBr disc):  $\nu_{\text{max}}$  (3340, 2958, 2368, 1718, 1611, 1538, 1436, 1236, 1090, 1044, 830, 777, 689, 537 $\text{cm}^{-1}$ )  
m/z 532  $[\text{M} + \text{H}]^+$



p) Synthesis of ethyl (4R,7S)-12,14-dihydroxy-1,3,4,5,6,7,8,10-octahydro-7-[(3R)-D-phenylalanine]-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **31**:

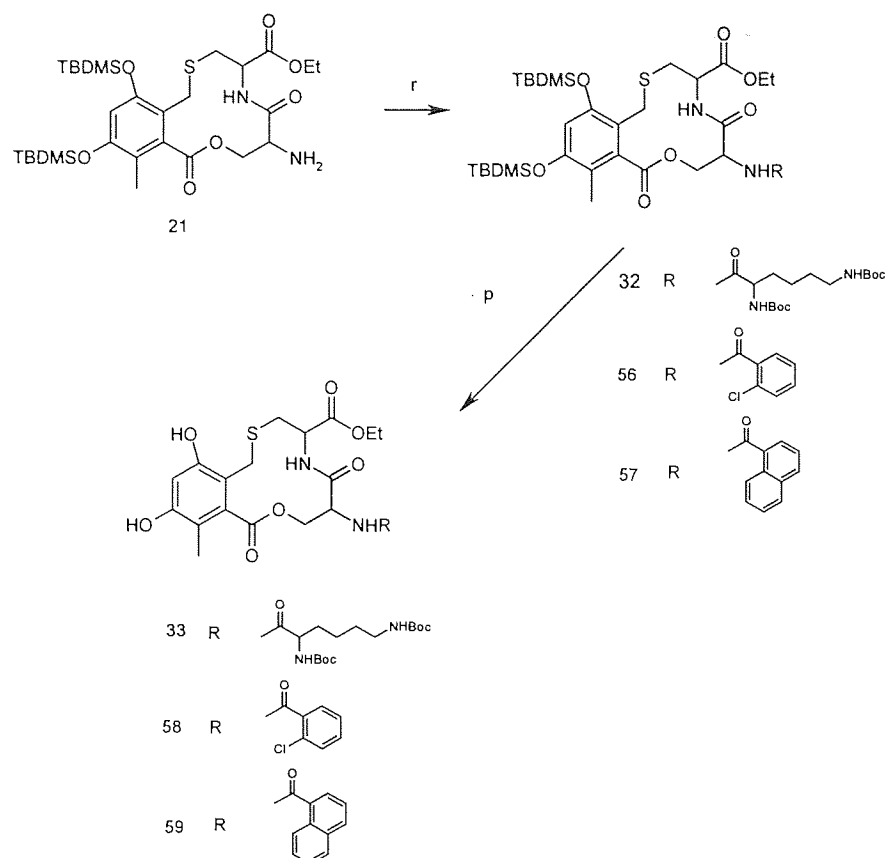
As described for **30**. Yield and m p are not available.

$^1\text{H NMR}$  ( $\text{CDCl}_3$ ):  $\delta$  1.16 (3H, t,  $\text{CH}_2\text{CH}_3$ ), 2.48 (3H, s,  $\text{PhCH}_3$ ), 2.75-2.79 (1H, dd,  $\text{CHS}$ ), 3.09 (2H, m,  $\text{CH}_2\text{Ph}$  in Phe), 3.12-3.19 (1H, dd,  $\text{CHS}$ ), 3.45-3.49 (1H, d,  $\text{Ph-CHS}$ ), 3.83-3.87 (1H, d,  $\text{PhCHS}$ ), 4.00 (2H, m,  $\text{COOCH}_2$ ), 4.30-4.41 (1H, dd,  $\text{PhCOOCH}$ ), 4.49 (1H, m,  $\text{CHCH}_2\text{Ph}$ ), 4.64 (1H, dd,  $\text{CHNH}$ ), 5.25-5.31 (1H, dd,  $\text{PhCOOCH}$ ), 6.31 (1H, s,  $\text{Ar-CH}$ ), 7.27 (5H, m, 5 x Ph), 8.16 (1H, d,  $\text{NHCO}$ ), 8.25 (1H, d,  $\text{NHCO}$ );

IR (KBr disc):  $\nu_{\text{max}}$  (3336, 2923, 2350, 1718, 1600, 1538, 1436, 1236, 1100, 1032, 850, 764, 709 $\text{cm}^{-1}$ )

$m/z$  546  $[\text{M} + \text{H}]^+$

HRMS ( $\text{ES}^+$ ):  $m/z$  546.1910 ( $\text{M} + \text{H}^+$ ,  $\text{C}_{26}\text{H}_{31}\text{N}_3\text{O}_8\text{S}$  requires 546.1906).



r) EDC, dried DCM, 0°C, 5h.

p) 1M solution of TBAF in THF, THF, r.t, 1h.

r) Synthesis of ethyl (4R,7S)-7-{{(3R)-1-N-di-(tert-butoxycarbonyl)-L-lysine}-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **32**:

To an ice-cold mixture of ethyl (4R,7S)-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **21** (500 mg, 0.81 mmol) and Boc-Lys(Boc)-OH.DCHA (290mg, 0.81mol) in dried DCM was added EDC (172mg, 0.89mmol). The mixture was stirred at 0°C for 5 hours. The solvent was removed in vacuo. The residue was diluted with EtOAc 20 ml, washed with 2M HCl (2 x 10ml), saturated NaHCO<sub>3</sub> (2 x 10ml) and brine (2 x 10ml). The organic layer was

dried over  $\text{Na}_2\text{SO}_4$ . The product was purified by column chromatography with the eluent EtOAc: Hexane (1:2) to afford white solid.  $R_f=0.33$ . Yield 36% M.p. 111-114°C

$^1\text{H}$  NMR (DMSO):  $\delta$  0.20-0.22 (12H, d, 2 x Si ( $\text{CH}_3$ )<sub>2</sub>), 0.96-0.99 (18H, d, 2 x ( $\text{CH}_3$ )<sub>3</sub>), 1.14 (3H, t,  $\text{OCH}_2\text{CH}_3$ ), 1.2-1.4 (4H, m, Lys  $\text{CH}_2\text{-CH}_2$ ), 1.35 (18H, s, 2 x Boc), 1.57 (2H, m, Lys  $\text{COCHCH}_2$ ), 1.88 (3H, s, Ph- $\text{CH}_3$ ), 2.83 (2H, m,  $\text{SCH}_2\text{CH}$ ), 3.60 (2H, m, Ph- $\text{CH}_2\text{S}$ ), 3.91 (1H, m, Ph- $\text{COOCH}$ ), 4.06 (2H, m,  $\text{OCH}_2\text{CH}_3$ ), 4.28 (1H, m, Ph- $\text{COOCH}$ ), 4.45 (1H, m,  $\text{CHNH-lys}$ ), 4.57 (1H, m, Lys  $\text{COCH-NHBoc}$ ), 4.84 (1H, m,  $\text{CHCOO}$ ), 6.44 (1H, s, Ph- $\text{H}$ ), 6.73 (1H, d, Lys  $\text{NH-Boc}$ ), 6.86 (1H, d, Lys  $\text{NH-Boc}$ ), 8.10 (1H, d,  $\text{NHCO}$ ), 8.45 ( $\text{NHCO}$ ), 9.49 (1H, s, Ph- $\text{OH}$ ), 9.52 (1H, s, Ph- $\text{OH}$ ).

$^{13}\text{C}$  NMR (DMSO):  $\delta$  -4.3 (Si ( $\text{CH}_3$ )<sub>2</sub>), -4.2 (Si ( $\text{CH}_3$ )<sub>3</sub>), 12.27 (Ar- $\text{CH}_3$ ), 13.58 ( $\text{OCH}_2\text{CH}_3$ ), 22.83 (Lys  $\text{COCHCH}_2\text{CH}_2$ ), 25.25 (Ph- $\text{CH}_2\text{S}$ ), 25.5 (Si C( $\text{CH}_3$ )<sub>2</sub>), 25.6 (Si C( $\text{CH}_3$ )<sub>3</sub>), 27.23 (Lys  $\text{COCHCH}_2\text{CH}_2\text{CH}_2\text{CH}_2$ ), 28.17 (Boc C( $\text{CH}_3$ )<sub>3</sub>), 28.30 (Boc C( $\text{CH}_3$ )<sub>3</sub>), 29.25 (Lys  $\text{COCHCH}_2\text{CH}_2\text{CH}_2$ ), 31.78 (Lys  $\text{COCHCH}_2$ ), 33.24 ( $\text{SCH}_2\text{CH}$ ), 51.82 (Lys  $\text{COCH}$ ), 52.25 ( $\text{COCHNH}$ ), 54.34 ( $\text{CHCOO}$ ), 60.94 ( $\text{OCH}_2\text{CH}_3$ ), 64.46 ( $\text{CH}_2\text{OCO}$ ), 77.36 (Boc C( $\text{CH}_3$ )<sub>3</sub>), 78.09 (Boc C( $\text{CH}_3$ )<sub>3</sub>), 103.18 (Ar- $\text{CH}$ ), 112.06 (ArC- $\text{CH}_2$ ), 134.78 (ArC- $\text{COO}$ ), 154.86 (Boc  $\text{COO}$ ), 154.96 (Boc  $\text{COO}$ ), 155.32 (ArC- $\text{O}$ ), 155.57 (ArC- $\text{O}$ ), 168.56 (Ar- $\text{COO}$ ), 168.63 ( $\text{COOEt}$ ), 170.24 ( $\text{NHCO}$ ), 172.54 ( $\text{NHCO}$ ).

IR (KBr disc):  $\nu_{\text{max}}$  (3354, 2927, 2851, 1697, 1589, 1513, 1468, 1365, 1244, 1172, 1037, 925, 835, 786, 665 $\text{cm}^{-1}$ )

Synthesis of ethyl (4R,7S)-7- $\{[1-(2\text{-chlorophenyl})\text{ethanone}]-12,14\text{-bis}[(\text{tert-butyl})\text{dimethylsiloxy}]-1,3,4,5,6,7,8,10\text{-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate } \mathbf{56}$ :

To an ice-cold mixture of ethyl (4R,7S)-12,14-bis $\{[(\text{tert-butyl})\text{dimethylsiloxy}]-1,3,4,5,6,7,8,10\text{-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate } \mathbf{21}$  (627 mg, 1 mmol) and 2-

chlorobenzoic acid (172.2mg, 1.1 mmol) in dried DCM 30ml was added EDC (210 mg, 1.1 mmol). The mixture was stirred at 0°C for 5 hours. The solvent was removed in vacuo. The residue was diluted with EtOAc 50 ml, washed with 2M HCl (2 x 20ml), saturated NaHCO<sub>3</sub> (2 x 20ml) and brine (2 x 20ml). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and evaporated in vacuo to give white solid. R<sub>f</sub> = 0.8 in EtOAc : Hexane (1:1). Yield 68 %. M.p. 142-144°C

<sup>1</sup>H NMR (CDCl<sub>3</sub>): δ 0.18-0.21 (12H, d, 2 x Si(CH<sub>3</sub>)<sub>2</sub>), 0.97-1.00 (18H, d, 2 x (CH<sub>3</sub>)<sub>3</sub>), 1.24 (3H, t, CH<sub>2</sub>CH<sub>3</sub>), 2.03 (3H, s, Ar-CH<sub>3</sub>) 2.89 (1H, dd, CHS), 3.18-3.19 (1H, dd, CHS), 3.45 (1H, d, Ph-CHS), 3.83-3.87 (1H, d, PhCHS), 4.00 (2H, m, COOCH<sub>2</sub>), 4.30-4.41 (1H, dd, CHOCH<sub>2</sub>), 4.81 (1H, dd, CHNH), 4.86 (1H, m, CHCOOEt), 5.22-5.31 (1H, dd, CHOCH<sub>2</sub>), 6.31 (1H, s, Ar-CH), 7.39 (4H, m, 4 x ArH), 7.53 (1H, d, CONH), 7.72 (1H, d, CONH);

<sup>13</sup>C NMR (CDCl<sub>3</sub>): δ -4.24 (Si (CH<sub>3</sub>)<sub>2</sub>), -0.1 (Si (CH<sub>3</sub>)<sub>3</sub>), 13.17 (Ar-CH<sub>3</sub>), 13.97 (OCH<sub>2</sub>CH<sub>3</sub>), 18.12 (Ph-CH<sub>2</sub>S), 25.54 (Si C(CH<sub>3</sub>)<sub>2</sub>), 25.63 (Si C(CH<sub>3</sub>)<sub>3</sub>), 34.56 (SCH<sub>2</sub>CH), 52.07 (COCHNH), 54.16 (CHCOO), 61.91 (OCH<sub>2</sub>CH<sub>3</sub>), 65.76 (CH<sub>2</sub>OCO), 110.51 (Ar-CH), 116.79 (ArC-CH<sub>2</sub>), 118.18 (ArC), 127.32 (ArC), 130.16 (ArC), 130.34 (ArC), 131.86 (ArC), 134.03 (ArC-CO), 134.94 (ArC-COO), 152.58 (ArC-O), 153.70 (ArC-O), 166.53 (CO), 168.11 (CO), 169.21 (CO), 170.20 (CO).

IR (KBr disc): ν<sub>max</sub> (3394, 2958, 2922, 2856, 2364, 1743, 1681, 1588, 1508, 1464, 1344, 1251, 1198, 1154, 928, 843, 772, 746, 679, 520cm<sup>-1</sup>)

Synthesis of ethyl (4R,7S)-7-{[1-(1-naphthyl)ethanone]-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **57**:

To an ice-cold mixture of ethyl (4R,7S)-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **21** (400 mg, 0.64 mmol) and 1-naphthoic acid (131.2mg, 0.76 mmol) in dried DCM 20ml was added EDC (145

mg, 0.76 mmol). The mixture was stirred at 0°C for 5 hours. The solvent was removed in vacuo. The residue was diluted with EtOAc 30 ml, washed with 2M HCl (2 x 10ml), saturated NaHCO<sub>3</sub> (2 x 10ml) and brine (2 x 10ml). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and evaporated in vacuo to give the white solid.

Yield 75%. M.p. 156-158°C

<sup>1</sup>H NMR (CDCl<sub>3</sub>): δ 0.18-0.21 (12H, d, 2 x Si(CH<sub>3</sub>)<sub>2</sub>), 0.99-1.00 (18H, d, 2 x (CH<sub>3</sub>)<sub>3</sub>), 1.24 (3H, t, CH<sub>2</sub>CH<sub>3</sub>), 2.03 (3H, s, Ar-CH<sub>3</sub>) 3.00 (1H, dd, CHS), 3.18-3.19 (1H, dd, CHS), 3.45 (1H, d, Ph-CHS), 3.83-3.87 (1H, d, PhCHS), 4.00 (2H, m, COOCH<sub>2</sub>), 4.30-4.41 (1H, dd, CHOCH<sub>2</sub>), 4.81 (1H, dd, CHNH), 4.86 (1H, m, CHCOOEt), 5.22-5.31 (1H, dd, CHOCH<sub>2</sub>), 6.31 (1H, s, Ar-CH), 7.45-7.85 (6H, m, 6 x ArH), 7.94 (1H, d, CONH), 8.41 (1H, d, CONH);

p) Synthesis of ethyl (4R,7S)-12,14-dihydroxy-1,3,4,5,6,7,8,10-octahydro-7-[[[(3R)-R-1-N-di-(tert-butoxycarbonyl)-L-lysine]-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **33**:

To a solution of ethyl (4R,7S)-7-[[[(3R)-1-N-di-(tert-butoxycarbonyl)-L-lysine]-12,14-bis[(tert-butyl)dimethylsiloxy]-1,3,4,5,6,7,8,10-octahydro-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **32** (300 mg, 0.3 mmol) in THF 2ml was added 1M solution of TBAF in THF (0.6 ml, 0.6 mmol) at room temperature. Stirring was continued for 1 hour. The product was purified by column chromatography with the eluent CH<sub>2</sub>Cl<sub>2</sub>: CH<sub>3</sub>OH (10: 1). R<sub>f</sub>=0.45 to afford the white solid. M.p. 164-166°C. Yield is not available.

<sup>1</sup>H NMR (DMSO): δ 1.14 (3H, t, OCH<sub>2</sub>CH<sub>3</sub>), 1.2-1.4 (4H, m, Lys CH<sub>2</sub>-CH<sub>2</sub>), 1.35 (18H, s, 2 x Boc), 1.57 (2H, m, Lys COCHCH<sub>2</sub>), 1.88 (3H, s, Ph-CH<sub>3</sub>), 2.83 (2H, m, SCH<sub>2</sub>CH), 3.60 (2H, m, Ph-CH<sub>2</sub>S), 3.91 (1H, m, Ph-COOCH), 4.06 (2H, m, OCH<sub>2</sub>CH<sub>3</sub>), 4.28 (1H, m, Ph-COOCH), 4.45 (1H, m, CHNH-lys), 4.57 (1H, m, Lys COCH-NHBoc), 4.84 (1H, m, CHCOO), 6.44 (1H, s, Ph-H), 6.73 (1H, d, Lys

NH-Boc), 6.86 (1H, d, Lys NH-Boc), 8.10 (1H, d, NHCO), 8.45 (NHCO), 9.49 (1H, s, Ph-OH), 9.52 (1H, s, Ph-OH).

<sup>13</sup>C NMR (DMSO):  $\delta$  12.27 (Ar-CH<sub>3</sub>), 13.58 (OCH<sub>2</sub>CH<sub>3</sub>), 22.83 (Lys COCHCH<sub>2</sub>CH<sub>2</sub>), 25.25 (Ph-CH<sub>2</sub>S), 27.23 (Lys COCHCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>), 28.17 (Boc C(CH<sub>3</sub>)<sub>3</sub>), 28.30 (Boc C(CH<sub>3</sub>)<sub>3</sub>), 29.25 (Lys COCHCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>), 31.78 (Lys COCHCH<sub>2</sub>), 33.24 (SCH<sub>2</sub>CH), 51.82 (Lys COCH), 52.25 (COCHNH), 54.34 (CHCOO), 60.94 (OCH<sub>2</sub>CH<sub>3</sub>), 64.46 (CH<sub>2</sub>OCO), 77.36 (Boc C(CH<sub>3</sub>)<sub>3</sub>), 78.09 (Boc C(CH<sub>3</sub>)<sub>3</sub>), 103.18 (Ar-CH), 112.06 (ArC-CH<sub>2</sub>), 134.78 (ArC-COO), 154.86 (Boc COO), 154.96 (Boc COO), 155.32 (ArC-OH), 155.57 (ArC-OH), 168.56 (Ar-COO), 168.63 (COOEt), 170.24 (NHCO), 172.54 (NHCO).

IR (KBr disc):  $\nu_{\max}$  (3365, 2971, 2943, 2359, 1688, 1610, 1515, 1366, 1252, 1171, 1107, 1032, 871, 781, 604cm<sup>-1</sup>)

*m/z* 727 [M + H]<sup>+</sup>

HRMS (ES<sup>+</sup>): *m/z* 727.3224

Synthesis of ethyl (4R,7S)-12,14-dihydroxy-1,3,4,5,6,7,8,10-octahydro-7-[[1-(2-chlorophenyl)ethanone]-11-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate **58**:

As described for **33**. M.p. 123-125°C. Yield is not available.

<sup>1</sup>H NMR (DMSO):  $\delta$  1.24 (3H, t, CH<sub>2</sub>CH<sub>3</sub>), 2.03 (3H, s, Ar-CH<sub>3</sub>), 2.89 (1H, dd, CHS), 3.18-3.19 (1H, dd, CHS), 3.45 (1H, d, Ph-CHS), 3.83-3.87 (1H, d, PhCHS), 4.00 (2H, m, COOCH<sub>2</sub>), 4.30-4.41 (1H, dd, CHOCH<sub>2</sub>), 4.81 (1H, dd, CHNH), 4.86 (1H, m, CHCOOEt), 5.31 (1H, dd, CHOCH<sub>2</sub>), 6.31 (1H, s, Ar-CH), 7.39 (4H, m, 4 x ArH), 7.53 (1H, d, CONH), 7.72 (1H, d, CONH);

<sup>13</sup>C NMR (DMSO):  $\delta$  13.07 (Ar-CH<sub>3</sub>), 14.17 (OCH<sub>2</sub>CH<sub>3</sub>), 18.02 (Ph-CH<sub>2</sub>S), 34.60 (SCH<sub>2</sub>CH), 52.17 (COCHNH), 53.86 (CHCOO), 62.11 (OCH<sub>2</sub>CH<sub>3</sub>), 66.01 (CH<sub>2</sub>OCO), 110.51 (Ar-CH), 116.83 (ArC-CH<sub>2</sub>), 118.3 (ArC), 127.42 (ArC), 130.16 (ArC), 130.34 (ArC), 131.86 (ArC), 133.93 (ArC-CO), 134.92 (ArC-COO), 152.55 (ArC-OH), 153.73 (ArC-OH), 167.03 (CO), 168.11 (CO), 169.41 (CO), 171.21 (CO).

IR (KBr disc):  $\nu_{\max}$  (3380, 2958, 2930, 2857, 2356, 1736, 1658, 1603, 1507, 1461, 1254, 1102, 1029, 854, 749, 588 $\text{cm}^{-1}$ )

Synthesis of ethyl (4R,7S)-12,14-dihydroxy-1,3,4,5,6,7,8,10-octahydro-7- $\{[1-(1\text{-naphthyl})\text{ethanone}]-11\text{-methyl-6,10-dioxo-9,2,5-benzoxathiazacyclododecine-4-carboxylate}$  **59**:

As described for **33**. M.p. 156-158°C. Yield is not available.

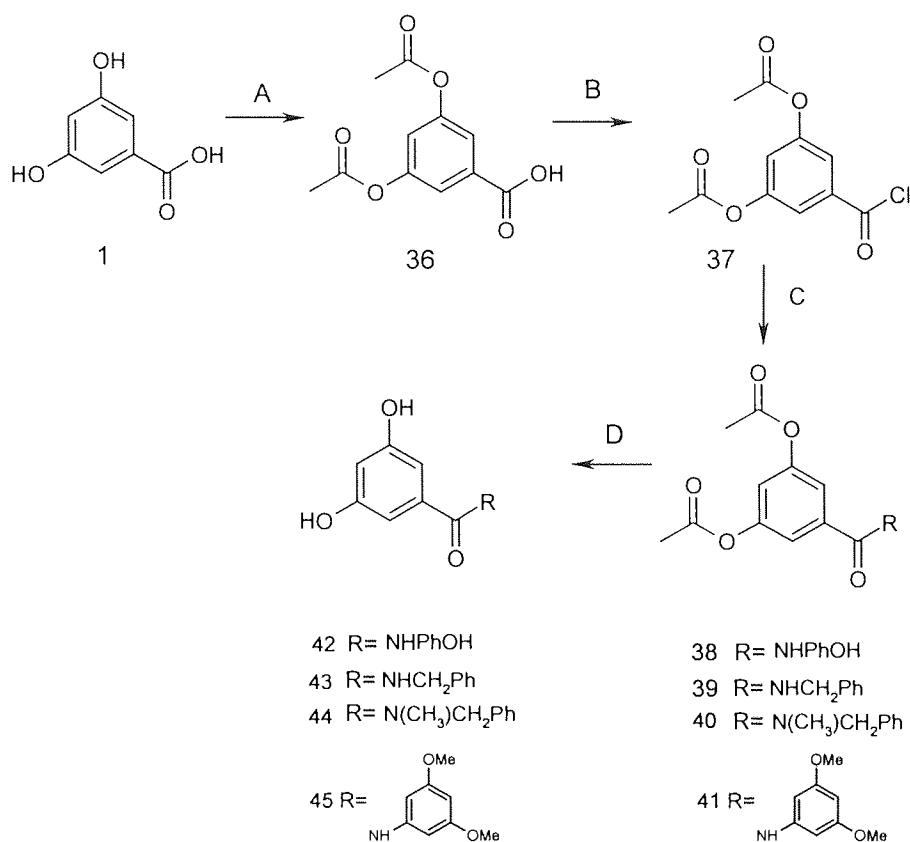
$^1\text{H NMR}$  (DMSO):  $\delta$  1.16 (3H, t,  $\text{CH}_2\text{CH}_3$ ), 1.91 (3H, s, Ar- $\text{CH}_3$ ) 2.87 (1H, dd,  $\text{CHS}$ ), 3.08 (1H, dd,  $\text{CHS}$ ), 3.45 (1H, d, Ph- $\text{CHS}$ ), 3.83-3.87 (1H, d, Ph $\text{CHS}$ ), 4.00 (2H, m,  $\text{COOCH}_2$ ), 4.30-4.41 (1H, dd,  $\text{CHOCH}_2$ ), 4.81 (1H, dd,  $\text{CHNH}$ ), 4.86 (1H, m,  $\text{CHCOOEt}$ ), 5.00 (1H, dd,  $\text{CHOCH}_2$ ), 6.41 (1H, s, Ar- $\text{CH}$ ), 7.45-7.85 (6H, m, 6 x ArH), 7.94 (1H, d,  $\text{CONH}$ ), 8.41 (1H, d,  $\text{CONH}$ );

IR (KBr disc):  $\nu_{\max}$  (2931, 1723, 1663, 1599, 1520, 1231, 1102, 1034, 855, 777, 722, 538 $\text{cm}^{-1}$ )

$m/z$  553  $[\text{M} + \text{H}]^+$

HRMS ( $\text{ES}^+$ ):  $m/z$  553.1644 ( $\text{M} + \text{H}^+$ ,  $\text{C}_{28}\text{H}_{28}\text{N}_2\text{O}_8\text{S}$  requires 553.1650).

## 2. Synthesis of open chained gyrase inhibitors:



A). Acetic anhydride, pyridine, 0 °C for 30 minutes, then r.t. for 3 hours.

B). SOCl<sub>2</sub>, refluxing for 4 hours.

C). R-NH<sub>2</sub>, pyridine in DMF, 0 °C for 30 minutes, then r.t. for 4-6 hours.

D). 2 M NaOH for 2 hours.

### A) Synthesis of 3,5-bis (acetyloxy) benzoic acid **36**:

To an ice-cold solution of 3,5-dihydroxybenzoic acid **1** (1g, 6.5 mmol) in ethyl acetate 15 ml was added acetic anhydride (1.59ml, 16.9 mmol) and pyridine (1.05 ml, 13 mmol). The mixture was stirred at 0 °C for 30 minutes, then at room



temperature for 3 hours. Extracted with ethyl acetate 30 ml and 2 M HCl 15 ml, it was washed with 2M HCl (2 x 15 ml) and brine (2 x 15 ml), dried over Na<sub>2</sub>SO<sub>4</sub>. The ethyl acetate was removed in vacuum and the white solid was crystallized from hexane. R<sub>f</sub> = 0.34 in DCM: MeOH (10:1). Yield: 98%. M.p. 154-157°C

<sup>1</sup>H NMR (DMSO): δ 2.28 (6H, s, 2 x CH<sub>3</sub>), 7.28 (1H, s, Ar-H), 7.58 (2H, d, 2 x Ar-H).

<sup>13</sup>C NMR (DMSO): δ 20.92 (2 x COCH<sub>3</sub>), 120.29 (3 x ArC), 133.04 (ArC-COOH), 150.92 (2 x ArC-OCO), 165.86 (COOH), 168.93 (2 x COCH<sub>3</sub>).

IR (KBr disc): ν<sub>max</sub> (2967, 2591, 1774, 1695, 1601, 1433, 1366, 1322, 1127, 1021, 910, 834, 750, 702, 657, 591, 453cm<sup>-1</sup>)

m/z (APCI) 239 [M + H]<sup>+</sup>

#### B) Synthesis of 3,5-bis (acetyloxy) benzoic acid chloride **37**:

A solution of 3,5-bis (acetyloxy) benzoic acid **36** (5g, 22 mmol) in thionyl chloride (SOCl<sub>2</sub>) 20 ml was refluxed for 4 hours. The thionyl chloride was removed in vacuum. The yellow solid was used directly in the next step. R<sub>f</sub> = 0.9 in DCM: MeOH (10:1). Yield: 66 %. M.p. 160-162°C

<sup>1</sup>H NMR (DMSO): δ 2.28 (6H, s, 2 x CH<sub>3</sub>), 7.28 (1H, t, Ar-H), 7.58 (2H, d, 2 x Ar-H).

<sup>13</sup>C NMR (DMSO): δ 21.0 (2 x COCH<sub>3</sub>), 120.21 (3 x ArC), 132.87 (ArC-CO), 151.10 (2 x ArC-OCO), 165.71 (COCl), 168.91 (2 x COCH<sub>3</sub>).

IR (KBr disc): ν<sub>max</sub> (3445, 3089, 2936, 1804, 1754, 1609, 1441, 1368, 1300, 1195, 1136, 1077, 1027, 941, 741, 700, 659 cm<sup>-1</sup>).

m/z (APCI) 257 [M + H]<sup>+</sup>

#### C) Synthesis of 3,5-bis (acetyloxy)-N-hydroxyphenyl- benzamide **38**:

To an ice-cold solution of 3,5-bis (acetyloxy) benzoic acid chloride **37** (150 mg, 0.61 mmol) and 4-aminophenol (80 mg, 0.73 mmol) in DMF 5 ml was added pyridine (0.25ml, 3.05 mmol). Stirring was continued at 0 °C for 6 hours. The product was extracted with ethyl acetate 20ml and 2 M HCl 10 ml, washed with 2 M HCl (2 x 10 ml), saturated NaHCO<sub>3</sub> (2 x 10 ml) and brine (2 x 10 ml), dried over Na<sub>2</sub>SO<sub>4</sub>. The ethyl acetate was removed in vacuum to afford the clear oil. R<sub>f</sub>=0.5 in DCM: MeOH (10:1). Yield 83 %.

<sup>1</sup>H NMR (DMSO): δ 2.30 (6H, s, 2 x CH<sub>3</sub>), 6.38 (1H, s, Ar-H), 6.80 (4H, m, 2 x Ar-H + 2 x Ar-H), 7.52 (2H, s, 2 x Ar-H), 9.23 (1H, s, NHCO), 9.85 (1H, s, OH).

<sup>13</sup>C NMR (CDCl<sub>3</sub>): δ 20.85 (2 x CH<sub>3</sub>), 115.61(ArC), 118.08 (2 x ArC), 120.75 (2 x ArC, in amino phenol), 121.69 (2 x ArC, in amino phenol), 131.76 (ArC-NH in amino phenol), 135.79 (ArC-CO), 136.79 (2 x ArC-OH in amino phenol), 151.00 (2 x ArC-O), 163.91 (2 x CO), 168.90 (NHCO).

IR (KBr disc): ν<sub>max</sub> (3323, 3072, 1768, 1673, 1595, 1545, 1514, 1432, 1373, 1291, 1191, 1114, 1023, 904, 745, 695, 514 cm<sup>-1</sup>).

m/z (APCI) 330 [M + H]<sup>+</sup>

### C) Synthesis of 3,5-bis (acetyloxy)-N-benzyl-benzamide **39**:

To an ice-cold solution of 3,5-bis (acetyloxy) benzoic acid chloride (**37**) (150 mg, 0.61 mmol) and 1-phenylmethanamine hydrochloride (105 mg, 0.73 mmol) in DMF 5 ml was added pyridine (0.25 ml, 3.05 mmol). Stirring was continued at 0 °C for 6 hours. The product was extracted with ethyl acetate 20ml and 2 M HCl 10 ml, washed with 2 M HCl (2 x 10 ml), saturated NaHCO<sub>3</sub> (2 x 10 ml) and brine (2 x 10 ml), dried over Na<sub>2</sub>SO<sub>4</sub>. The ethyl acetate was removed in vacuum to afford the colourless oil. R<sub>f</sub>=0.4 in DCM: MeOH (10:1). Yield 78%.

<sup>1</sup>H NMR (CDCl<sub>3</sub>): δ 2.17 (6H, s, 2 x CH<sub>3</sub>), 4.41 (2H, d, Ph-CH<sub>2</sub>-), 6.98 (1H, t, Ar-H), 7.21 (5H, m, 5 x Ar-H), 7.30 (2H, m, 2 x Ar-H), 7.41 (1H, d, NHCO)

$^{13}\text{C}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  20.80 (2 x  $\underline{\text{C}}\text{H}_3$ ), 43.9 ( $\underline{\text{C}}\text{H}_2\text{Ph}$ ), 117.89 (2 x  $\text{Ar}\underline{\text{C}}$ ), 118.42 ( $\text{Ar}\underline{\text{C}}$ ), 127.32 ( $\text{Ar}\underline{\text{C}}$ ), 127.69 (2 x  $\text{Ar}\underline{\text{C}}$ ), 128.6 (2 x  $\text{Ar}\underline{\text{C}}$ ), 136.4 ( $\text{Ar}\underline{\text{C}}\text{-CO}$ ), 137.92 ( $\text{Ar}\underline{\text{C}}\text{-CH}_2$ ), 150.90 (2 x  $\text{Ar}\underline{\text{C}}$ ), 165.39 (2 x  $\underline{\text{C}}\text{O}$ ), 168.79 ( $\text{NH}\underline{\text{C}}\text{O}$ ).

IR (KBr disc):  $\nu_{\text{max}}$  (3324, 2939, 2366, 1770, 1645, 1582, 1537, 1452, 1363, 1309, 1219, 1116, 1018, 901, 696  $\text{cm}^{-1}$ ).

$m/z$  (APCI) 328  $[\text{M} + \text{H}]^+$

### C) Synthesis of 3,5-bis (acetyloxy)-N-benzyl(methyl)-benzamide **40**:

To an ice-cold solution of 3,5-bis (acetyloxy) benzoic acid chloride **37** (150 mg, 0.61 mmol) and N-benyl-N-methylamino (88mg, 0.73 mmol) in DMF 5 ml was added pyridine (0.25 ml, 3.05 mmol). Stirring was continued at 0 °C for 6 hours. The product was extracted with ethyl acetate 20ml and 2 M HCl 10 ml, washed with 2 M HCl (2 x 10 ml), saturated  $\text{NaHCO}_3$  (2 x 10 ml) and brine (2 x 10 ml), dried over  $\text{Na}_2\text{SO}_4$ . The ethyl acetate was removed in vacuum to afford the white solid. Yield 60 %. M p is not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  2.28 (6H, s, 2 x  $\underline{\text{C}}\text{H}_3$ ), 2.80 (3H, s,  $\underline{\text{C}}\text{H}_3\text{-N-}$ ), 4.62 (2H, d,  $\text{Ph-CH}_2\text{-}$ ), 6.26 (2H, s, 2 x  $\text{Ar-}\underline{\text{H}}$ ), 7.34 (6H, m, 5 x  $\text{Ar-}\underline{\text{H}}$  and  $\text{Ar-}\underline{\text{H}}$ ).

$^{13}\text{C}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  20.80 (2 x  $\underline{\text{C}}\text{H}_3$ ), 36.90 ( $\text{N}\underline{\text{C}}\text{H}_3$ ), 50.85 ( $\underline{\text{C}}\text{H}_2\text{Ph}$ ), 116.55 (2 x  $\text{Ar}\underline{\text{C}}$ ), 117.79 ( $\text{Ar}\underline{\text{C}}$ ), 126.78 ( $\text{Ar}\underline{\text{C}}$ ), 126.84 ( $\text{Ar}\underline{\text{C}}$ ), 127.56 ( $\text{Ar}\underline{\text{C}}$ ), 128.14 ( $\text{Ar}\underline{\text{C}}$ ), 128.69 ( $\text{Ar}\underline{\text{C}}$ ), 136.56 ( $\text{Ar}\underline{\text{C}}\text{-CO}$ ), 137.70 ( $\text{Ar}\underline{\text{C}}\text{-CH}_2$ ), 150.83 (2 x  $\text{Ar}\underline{\text{C}}$ ), 165.39 (2 x  $\underline{\text{C}}\text{O}$ ), 168.79 ( $\text{NH}\underline{\text{C}}\text{O}$ ).

IR (KBr disc):  $\nu_{\text{max}}$  (3065, 2927, 1779, 1637, 1495, 1437, 1397, 1300, 1290, 1205, 1112, 1010, 912, 707  $\text{cm}^{-1}$ ).

$m/z$  (APCI) 342  $[\text{M} + \text{H}]^+$

C) Synthesis of 3,5-bis (acetyloxy)-N-(5-dimethoxyphenyl)-benzamide **41**:

To an ice-cold solution of 3,5-bis (acetyloxy) benzoic acid chloride **37** (600 mg, 2.4 mmol) and 3,5-dimethoxyaniline (397mg, 2.6mmol) in DMF 5 ml was added pyridine (1ml, 12mmol). Stirring was continued at 0 °C for 6 hours. The product was extracted with ethyl acetate 20ml and 2 M HCl 10 ml, washed with 2 M HCl (2 x 10 ml), saturated NaHCO<sub>3</sub> (2 x 10 ml) and brine (2 x 10 ml), dried over Na<sub>2</sub>SO<sub>4</sub>. The ethyl acetate was removed in vacuum to afford the white solid. M.p. 110-113 °C. Yield 55%.

<sup>1</sup>H NMR (CDCl<sub>3</sub>): δ 2.27 (6H, s, CH<sub>3</sub>CO), 3.83 (6H, s, OCH<sub>3</sub>), 6.80 (1H, d, ArH, between two methoxyl groups), 6.94-6.98 (2H, dd, 2 x ArH between amide and methoxyl), 7.32 (1H, d, ArH between two esters), 7.45 (2H, d, 2 x ArH between amide and ester), 7.96 (1H, s, NHCO);

<sup>13</sup>C NMR (CDCl<sub>3</sub>): δ 20.92 (2 x CH<sub>3</sub>CO), 55.79 (2 x OCH<sub>3</sub>), 105.20 (ArC, between two methoxyl groups), 111.17 (2 x ArC between amide and methoxyl), 112.47 (2 x ArC between amide and ester), 117.74 (ArC, between two esters), 130.98 (ArC-CONH), 137.04 (ArC-NHCO), 146.14 (2 x ArC-O-CO), 148.92 (2 x ArC-OCH<sub>3</sub>), 151.09 (CONH), 168.77 (2 x CH<sub>3</sub>CO);

IR (KBr disc): ν<sub>max</sub> (3410, 2959, 2842, 2369, 1607, 1522, 1346, 1233, 1152, 1004, 842, 801, 756, 670, 531 cm<sup>-1</sup>).

m/z (APCI) 374 [M + H]<sup>+</sup>

D) Synthesis of 3,5-dihydroxy-N-(4-hydroxyphenyl)benzamide **42**:

A solution of 3-(acetyloxy)-5-[[[(4-hydroxyphenyl)amino]carbonyl]phenyl acetate **38** in 2M NaOH 5 ml was stirred at room temperature for 2 hours. The mixture's pH value was adjusted to 1 with the 2 M HCl. It was extracted with ethyl acetate 20 ml, washed with 2M HCl (2 x 10ml) and brine (2 x 10 ml), dried over Na<sub>2</sub>SO<sub>4</sub>. The organic phase was removed in vacuum to afford the white solid. R<sub>f</sub>=0.3 in DCM: MeOH (10:1). M.p. >160 °C. Yield is not available.

$^1\text{H}$  NMR (DMSO): 6.38 (1H, s, Ar-H), 6.80 (4H, m, 2 x Ar-H + 2 x Ar-H), 7.52 (2H, s, 2 x Ar-H), 9.23 (1H, s, NHCO), 9.53 (2H, s, 2 x OH), 9.85 (1H, s, OH).

$^{13}\text{C}$  NMR (DMSO):  $\delta$  105.18 (ArC), 105.72 (2 x ArC), 114.94 (2 x ArC, in amino phenol), 122.10 (2 x ArC, in amino phenol), 130.89 (ArC-NH in amino phenol), 137.45 (ArC-CO), 153.57 (2 x ArC-OH in amino phenol), 158.42 (2 x ArC-OH), 165.23 (Ar-CO).

IR (KBr disc):  $\nu_{\text{max}}$  (3395, 1701, 1657, 1609, 1543, 1451, 1328, 1232, 1158, 1004, 816, 755, 667, 501  $\text{cm}^{-1}$ ).

m/z (APCI) 246  $[\text{M} + \text{H}]^+$

D) Synthesis of N-benzyl-3,5-dihydroxybenzamide **43**:

As described for **42**. M.p. 105-109 °C. Yield is not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  4.41 (2H, d, Ph-CH<sub>2</sub>-), 6.36 (1H, t, Ar-H), 6.72 (2H, d, 2 x Ar-H), 7.29 (5H, m, 5 x Ar-H), 8.84 (1H, t, NHCO), 9.48 (2H, s, 2 x Ar-OH).

$^{13}\text{C}$  NMR (DMSO):  $\delta$  42.52 (CH<sub>2</sub>Ph), 105.12 (ArC), 105.48 (2 x ArC), 126.68 (ArC), 127.29 (2 x ArC), 128.30 (2 x ArC), 136.66 (ArC-CO), 139.88 (ArC-CH<sub>2</sub>), 158.31 (2 x ArC-OH), 166.48 (CO).

IR (KBr disc):  $\nu_{\text{max}}$  (3300, 2366, 1594, 1545, 1450, 1338, 1163, 1001, 853, 759, 701  $\text{cm}^{-1}$ ).

m/z (APCI) 244  $[\text{M} + \text{H}]^+$

D) Synthesis of N-benzyl-3,5-dihydroxy-N-methylbenzamide **44**:

As described for **42**. M.p. 103-104 °C. Yield is not available.

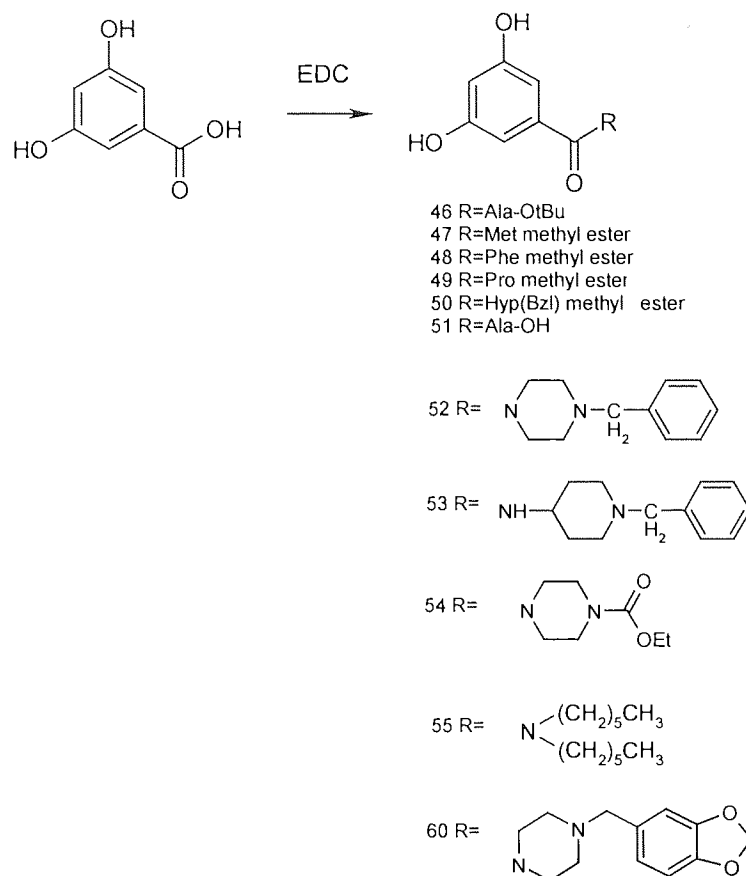
$^1\text{H}$  NMR (CDCl<sub>3</sub>):  $\delta$  2.77 (3H, s, CH<sub>3</sub>-N-), 4.43 (2H, d, Ph-CH<sub>2</sub>-), 6.26 (2H, s, 2 x Ar-H), 7.34 (6H, m, 5 x Ar-H and Ar-H), 8.77 (2H, br, 2 x Ar-OH).

$^{13}\text{C}$  NMR (DMSO):  $\delta$  20.81 ( $\text{NCH}_3$ ), 59.80 ( $\text{CH}_2\text{Ph}$ ), 103.29 (2 x  $\text{ArC}$ ), 104.55 ( $\text{ArC}$ ), 126.78 ( $\text{ArC}$ ), 126.84 ( $\text{ArC}$ ), 127.56 ( $\text{ArC}$ ), 128.14 ( $\text{ArC}$ ), 128.69 ( $\text{ArC}$ ), 136.99 ( $\text{ArC-CO}$ ), 138.18 ( $\text{ArC-CH}_2$ ), 150.83 (2 x  $\text{ArC}$ ), 158.44 ( $\text{NHCO}$ ).  
IR (KBr disc):  $\nu_{\text{max}}$  (3238, 1711, 1594, 1438, 1400, 1339, 1162, 1084, 1006, 950, 867, 694  $\text{cm}^{-1}$ ).  
 $m/z$  (APCI) 258  $[\text{M} + \text{H}]^+$

D) Synthesis of N-(3,5-dimethoxyphenyl)-3,5-dihydroxybenzamide **45**:

As described for **42**. M.p.  $>150$  °C. Yield is not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  3.74 (6H, s, 2 x  $\text{OCH}_3$ ), 6.45 (1H, s,  $\text{ArH}$  between two  $\text{OCH}_3$ ), 6.82 (2H, s, 2 x  $\text{ArH}$  between amide and methoxyl), 6.90 (1H, d,  $\text{ArH}$  between two alcohols), 7.34 (2H, d, 2 x  $\text{ArH}$  between OH and amide), 9.62 (1H, s,  $\text{CONH}$ ), 9.97 (2H, s, 2 x  $\text{OH}$ );  
 $^{13}\text{C}$  NMR (DMSO):  $\delta$  55.43 ( $\text{OCH}_3$ ), 55.66 ( $\text{OCH}_3$ ), 105.45 ( $\text{ArC}$ , between two  $\text{OCH}_3$ ), 105.78 ( $\text{ArC}$ , between amide and  $\text{OCH}_3$ ), 111.83 ( $\text{ArC}$ , between amide and  $\text{OCH}_3$ ), 112.27 ( $\text{ArC}$ , between two OH), 132.86 ( $\text{ArC}$ , between OH and amide), 137.31 ( $\text{ArC-CONH}$ ), 145.07 ( $\text{ArC-NHCO}$ ), 148.44 (2 x  $\text{ArC-OH}$ ), 158.38 (2 x  $\text{ArC-OCH}_3$ ), 165.44 ( $\text{CONH}$ );  
IR (KBr disc):  $\nu_{\text{max}}$  (3410, 2959, 2842, 2369, 1607, 1522, 1346, 1233, 1152, 1004, 842, 801, 756, 670, 531  $\text{cm}^{-1}$ ).  
 $m/z$  (APCI) 290  $[\text{M} + \text{H}]^+$



#### Synthesis of tert-butyl-2-[(3,5-dihydroxybenzoyl)amino]propanoate **46**:

To an ice-cold solution of 3,5-dihydroxybenzoic acid (212 mg, 1.37 mmol), Ala-OtBu (300 mg, 1.65 mmol) in THF 15 ml was added EDC (316 mg, 1.65 mmol). Stirring was continued at 0 °C for 6 hours. The mixture was extracted with ethyl acetate 40 ml and 2 M HCl 20ml, washed with 2M HCl (2 x 20 ml) and brine (2 x 20 ml), dried over Na<sub>2</sub>SO<sub>4</sub>. The organic phase was removed in vacuum to afford the white solid. TLC R<sub>f</sub>=0.4 in DCM: MeOH (10:1). M.p. 104-106 °C. Yield is not available.

<sup>1</sup>H NMR (DMSO): δ 1.34 (3H, d, CH<sub>3</sub>), 1.39 (9H, s, Boc), 4.23 (1H, m, -CHCH<sub>3</sub>), 6.37 (1H, s, Ar-H), 6.70 (2H, d, 2 x Ar-H), 8.46 (1H, d, NHCO), 9.47 (2H, s, 2 x OH).

$^{13}\text{C}$  NMR (DMSO):  $\delta$  13.80 ( $\text{CHCH}_3$ ), 29.97 (Boc), 51.97 ( $\text{CHCH}_3$ ), 105.42 (ArC), 105.76 (2 x ArC), 135.95 (ArC-CO), 158.79 (2 x ArC-OH), 167.11 (ArCO), 172.55 (COO).  
IR (KBr disc):  $\nu_{\text{max}}$  (3304, 2980, 1712, 1601, 1522, 1446, 1362, 1158, 1052, 1012, 848, 684,  $520\text{cm}^{-1}$ )  
m/z (APCI): 282 [M + H] $^+$

Synthesis of methyl-2-[(3,5-dihydroxybenzoyl)amino]-4-(methylthio)-butanoate **47**:

As described for **46**. M.p. 107-110 °C. Yield is not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  2.04 (3H, s,  $\text{SCH}_3$ ), 2.55 (4H, m,  $-(\text{CH}_2)_2-$ ), 3.64 (3H, s,  $\text{COOCH}_3$ ), 4.51 (1H, m,  $\text{CHCOO}$ ), 6.38 (1H, s, Ar-H), 6.71 (2h, d, 2 x Ar-H), 8.60 (1H, d,  $\text{NHCO}$ ), 9.52 (2H, s, 2 x Ar-OH).  
 $^{13}\text{C}$  NMR (DMSO):  $\delta$  16.70 ( $\text{SCH}_3$ ), 28.96 ( $-\text{CH}_2\text{SCH}_3$ ), 39.62 ( $\text{CH}_2\text{CH}_2\text{S}$ ), 48.93 ( $\text{COOCH}_3$ ), 80.33 ( $\text{NHCH}$ ), 105.27 (ArC), 105.68 (2 x ArC), 136.22 (ArC-CO), 158.22 (2 x ArC-OH), 166.76 (ArCO), 171.99 (COO).  
IR (KBr disc):  $\nu_{\text{max}}$  (3318, 1729, 1607, 1540, 1450, 1361, 1159, 1006, 853, 768,  $678\text{cm}^{-1}$ ).  
m/z (APCI): 300 [M + H] $^+$

Synthesis of methyl-2-[(3,5-dihydroxybenzoyl)amino]-3-phenylpropanoate **48**:

As described for **46**. M.p. 95-97 °C. Yield is not available.



$^1\text{H}$  NMR (DMSO):  $\delta$  3.11 (2H, m,  $\text{CH}_2\text{Ph}$ ), 3.62 (3H, s,  $\text{OCH}_3$ ), 4.59 (1H, m,  $\text{CHCOO}$ ), 6.34 (1H, s, Ar-H), 6.62 (2H, s, 2 x Ar-H), 7.27 (5H, m,  $\text{CH}_2\text{Ph}$ ), 8.64 (1H, d,  $\text{NHCO}$ ), 9.49 (2H, s, 2 x OH).

$^{13}\text{C}$  NMR (DMSO):  $\delta$  36.10 ( $\text{CH}_2\text{Ph}$ ), 51.95 ( $\text{OCH}_3$ ), 54.22 ( $\text{CHCOO}$ ), 105.37 (ArC), 105.64 (2 x ArC), 126.49 (ArC in Phe), 128.27 (2 x ArC in Phe), 129.07 (2 x ArC in Phe), 135.90 (ArC-CO), 137.85 (ArC- $\text{CH}_2$  in Phe), 158.24 (2 x ArC-OH), 166.74 (Ar-CO), 172.27 (COO).

IR (KBr disc):  $\nu_{\text{max}}$  (3368, 2948, 2858, 2357, 1748, 1636, 1600, 1573, 1443, 1367, 1170, 1004, 875, 704  $\text{cm}^{-1}$ ).

m/z (APCI): 316  $[\text{M} + \text{H}]^+$

Synthesis of methyl-1-(3,5-dihydroxybenzoyl)pyrrolidine-2-carboxylate **49**:

As described for **46**. M.p. 100-102  $^\circ\text{C}$ . Yield is not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  1.83 (4H, m,  $\text{NCH}_2\text{CH}_2\text{CH}_2-$ ), 3.44 (2H, m,  $\text{NCH}_2$ ), 3.65 (3H, s,  $\text{OCH}_3$ ), 4.41 (1H, m,  $\text{NCHCOO}$ ), 6.27 (1H, s, Ar-H), 6.30 (2H, d, 2 x Ar-H), 9.52 (2H, s, 2 x Ar-OH).

$^{13}\text{C}$  NMR (DMSO):  $\delta$  24.96 ( $\text{NCH}_2\text{CH}_2$ ), 28.89 ( $\text{NCH}_2\text{CH}_2\text{CH}_2$ ), 49.57 ( $\text{NCH}_2$ ), 52.14 ( $\text{OCH}_3$ ), 58.68 ( $\text{NCHCOO}$ ), 103.46 (ArC), 105.00 (2 x ArC), 137.92 (ArC-CO), 158.28 (2 x ArC-OH), 168.37 (Ar-CO), 172.48 (COO).

IR (KBr disc):  $\nu_{\text{max}}$  (3268, 1726, 1588, 1437, 1362, 1296, 1158, 1007, 852, 768, 684, 520  $\text{cm}^{-1}$ ).

m/z (APCI): 266  $[\text{M} + \text{H}]^+$

Synthesis of methyl-3-(benzyloxy)-1-(3,5-dihydroxybenzoyl)pyrrolidine-2-carboxylate **50**:

As described for **46**. M.p. 89-92  $^\circ\text{C}$ . Yield is not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  2.40 (2H, m,  $\text{NCH}_2\text{CH}_2$ ), 3.54 (2H, m,  $\text{NCH}_2$ ), 3.66 (3H, s,  $\text{OCH}_3$ ), 4.15 (1H, m,  $-\text{CH}-\text{O}$ ), 4.50 (3H, m,  $\text{CHCOO} + \text{CH}_2\text{Ph}$ ), 6.31 (1H, s,  $\text{Ar-H}$ ), 6.32 (2H, d, 2 x  $\text{Ar-H}$ ), 7.29 (5H, m,  $\text{CH}_2\text{Ph}$ ), 9.57 (2H, s, 2 x  $\text{OH}$ ).

$^{13}\text{C}$  NMR (DMSO):  $\delta$  34.20 ( $\text{NCH}_2\text{CH}_2$ ), 52.02 ( $\text{OCH}_3$ ), 54.86 ( $\text{NCH}_2$ ), 57.47 ( $\text{NCH}$ ), 69.82 ( $\text{CH-O-CH}_2$ ), 76.73 ( $\text{OCH}_2\text{Ph}$ ), 104.33 ( $\text{ArC}$ ), 105.23 (2 x  $\text{ArC}$ ), 127.51 (2 x  $\text{ArC}$  in Hyp), 128.32 (2 x  $\text{ArC}$  in Hyp), 137.39 ( $\text{ArC-CH}_2$ ), 138.15 ( $\text{ArC-CO}$ ), 158.35 (2 x  $\text{ArC-OH}$ ), 168.91 ( $\text{Ar-CO}$ ), 172.28 ( $\text{COO}$ ).

IR (KBr disc):  $\nu_{\text{max}}$  (3318, 2954, 2382, 1745, 1586, 1464, 1359, 1173, 1077, 1014, 850, 741  $\text{cm}^{-1}$ ).

$m/z$  (APCI): 372  $[\text{M} + \text{H}]^+$

#### Synthesis of 2-[(3,5-dihydroxybenzoyl)amino]propanoic acid **51**:

To an ice cold solution of tert-butyl-2-[(3,5-dihydroxybenzoyl)amino]propanoate **46** (550mg, 1.9mmol) in DCM 10ml was added TFA 7ml. The mixture was stirred at room temperature for 1 hour. The solution was removed in vacuum. The residue was extracted with EtOAc 20ml and 2M HCl 10ml, washed with HCl and brine (2 x 10 ml) and dried over  $\text{Na}_2\text{SO}_4$ . The EtOAc was removed to afford the white solid. M.p. 98-100  $^\circ\text{C}$ . Yield is not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  1.34 (3H, d,  $\text{CH}_3$ ), 4.38 (1H, m,  $\text{NHCH}$ ), 6.37 (1H, s,  $\text{ArH}$ ), 6.72 (2H, s, 2 x  $\text{ArH}$ ), 8.47 (1H, d,  $\text{NHCO}$ ), 9.61 (2H, br, 2 x  $\text{OH}$ );

$^{13}\text{C}$  NMR (DMSO):  $\delta$  16.82 ( $\text{CH}_3$ ), 48.10 ( $\text{CHCH}_3$ ), 105.7 ( $\text{ArC}$ ), 107.34 (2 x  $\text{ArC}$ ), 136.24 ( $\text{ArC-CO}$ ), 158.25 (2 x  $\text{ArC-OH}$ ), 166.51 ( $\text{CONH}$ ), 174.33 ( $\text{COOH}$ );  
IR (KBr disc):  $\nu_{\text{max}}$  (3376, 1724, 15989, 1545, 1455, 1356, 1226, 1168, 1010, 867, 768, 674  $\text{cm}^{-1}$ ).

$m/z$  (APCI): 224  $[\text{M} + \text{H}]^+$

Synthesis of 5-[(4-benzylpiperazin-1-yl)carbonyl]benzene-1,3-diol **52**:

As described for **46**. M.p. 138-140 °C. Yield is not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  2.33 (4H, m, 2 x  $\text{CH}_2\text{-N}$ ), 3.34 (4H, m, 2 x  $\text{N-CH}_2$ ), 3.47 (2H, s,  $\text{CH}_2\text{Ph}$ ), 6.16 (2H, s, 2 x  $\text{ArH}$ ), 6.26 ( $\text{ArH}$ ), 7.29 (5H, m, 5 x  $\text{ArH}$ ), 9.52 (2H, s, 2 x  $\text{OH}$ );

$^{13}\text{C}$  NMR (DMSO):  $\delta$  53.06 (2 x  $\text{CH}_2\text{N}$ ), 59.82 (2 x  $\text{CH}_2\text{N}$ ), 61.89 ( $\text{CH}_2\text{Ph}$ ), 103.39 ( $\text{ArC}$ ), 104.72 (2 x  $\text{ArC}$ ), 127.07 ( $\text{ArC}$ ), 128.23 (2 x  $\text{ArC}$ ), 128.97 (2 x  $\text{ArC}$ ), 137.71 ( $\text{ArC-CH}_2$ ), 158.76 (2 x  $\text{ArC-OH}$ ), 170.42 ( $\text{CO}$ ).

IR (KBr disc):  $\nu_{\text{max}}$  (3196, 2932, 2819, 1747, 1603, 1446, 1338, 1302, 1199, 1163, 1001, 849, 750,  $701\text{cm}^{-1}$ )

$m/z$  (APCI): 312  $[\text{M} + \text{H}]^+$

Synthesis of N-(1-benzylpiperidin-4-yl)-3,5-dihydroxybenzamide **53**:

As described for **46**. M.p. 144-146 °C. Yield is not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  1.67 (4H, m, 2 x  $\text{CHCH}_2$ ), 1.98 (2H, m,  $\text{CH}_2\text{N}$ ), 2.81 (2H, m,  $\text{CH}_2\text{N}$ ), 3.43 (2H, s,  $\text{CH}_2\text{Ph}$ ), 3.70 (1H, m,  $\text{NHCH}$ ), 6.35 (1H, s,  $\text{ArH}$ ), 6.64 (2H, s, 2 x  $\text{ArH}$ ), 7.29 (5H, m, 5 x  $\text{ArH}$ ), 8.05 (1H, d,  $\text{CONH}$ ), 9.44 (2H, s, 2 x  $\text{OH}$ );

$^{13}\text{C}$  NMR (DMSO):  $\delta$  30.18 (2 x  $\text{CH}_2$ ), 44.66 ( $\text{NHCH}$ ), 52.87 (2 x  $\text{CH}_2\text{N}$ ), 62.79 ( $\text{CH}_2\text{Ph}$ ), 105.04 ( $\text{ArC}$ ), 105.56 (2 x  $\text{ArC}$ ), 127.06 ( $\text{ArC}$ ), 128.30 (2 x  $\text{ArC}$ ), 128.77 (2 x  $\text{ArC}$ ), 136.89 ( $\text{ArC-CH}_2$ ), 158.20 (2 x  $\text{ArC-OH}$ ), 166.09 ( $\text{CO}$ ).

IR (KBr disc):  $\nu_{\text{max}}$  (3339, 2958, 2801, 1589, 1543, 1350, 1162, 997, 859, 772,  $689\text{cm}^{-1}$ ).

$m/z$  (APCI): 326  $[\text{M} + \text{H}]^+$

Synthesis of ethyl 4-(3,5-dihydroxybenzoyl)piperazine-1-carboxylate **54**:

As described for **46**. M.p. >160 °C. Yield is not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  1.17 (3H, t,  $\text{CH}_2\text{CH}_3$ ), 3.39-3.51 (8H, m, 4 x  $\text{CH}_2\text{N}$ ), 4.03 (2H, m,  $\text{CH}_2\text{CH}_3$ ), 6.17 (2H, s, 2 x ArH), 6.26 (1H, s, ArH), 9.53 (2H, s, 2 x OH);  
 $^{13}\text{C}$  NMR (DMSO):  $\delta$  14.58 ( $\text{CH}_2\text{CH}_3$ ), 43.34 (2 x  $\text{NCH}_2$ ), 46.90 (2 x  $\text{CH}_2\text{N}$ ), 60.97 ( $\text{OCH}_2\text{CH}_3$ ), 103.49 (ArC), 104.80 (2 x ArC), 137.48 (ArC), 154.63 (COO), 158.45 (2 x ArC-OH), 169.25 (CON);  
IR (KBr disc):  $\nu_{\text{max}}$  (3154, 1699, 1659, 1593, 1437, 1348, 1232, 1156, 983, 841, 765, 667,  $560\text{cm}^{-1}$ )  
m/z (APCI): 294 [M + H] $^+$

Synthesis of N,N-dihexyl-3,5-dihydroxybenzamide **55**:

As described for **46**. Yellow oil. Yield and m p are not available.

$^1\text{H}$  NMR (DMSO):  $\delta$  0.87 (6H, m, 2 x  $\text{CH}_3$ ), 1.07-1.26 (12H, m, 6 x  $\text{CH}_2$ ), 1.54 (4H, m, 2 x  $\text{NCH}_2\text{CH}_2$ ), 2.83 (2H, m,  $\text{NCH}_2$ ), 3.14 (2H, m,  $\text{NCH}_2$ ), 6.07 (2H, s, 2 x ArH), 6.22 (1H, s, ArH), 9.47 (2H, s, 2 x OH);  
 $^{13}\text{C}$  NMR (DMSO):  $\delta$  13.89 (2 x  $\text{CH}_3$ ), 21.97 (2 x  $\text{CH}_2$ ), 25.89 (2 x  $\text{CH}_2$ ), 26.33 (2 x  $\text{CH}_2$ ), 30.88 (2 x  $\text{CH}_2$ ), 47.34 (2 x  $\text{CH}_2$ ), 102.79 (ArC), 104.11 (2 x ArC), 139.06 (ArC-CO), 158.37 (2 x ArC-OH), 170.40 (CO).  
IR (KBr disc):  $\nu_{\text{max}}$  (3232, 2959, 2932, 2854, 1582, 1464, 1373, 1345, 1172, 1004, 845,  $686\text{cm}^{-1}$ ).  
m/z (APCI): 321 [M + H] $^+$

Synthesis of 5-{{4-(1,3-benzodioxol-5ylmethyl)piperazin-1-yl}carbonyl}benzene-1,3-diol **60**:

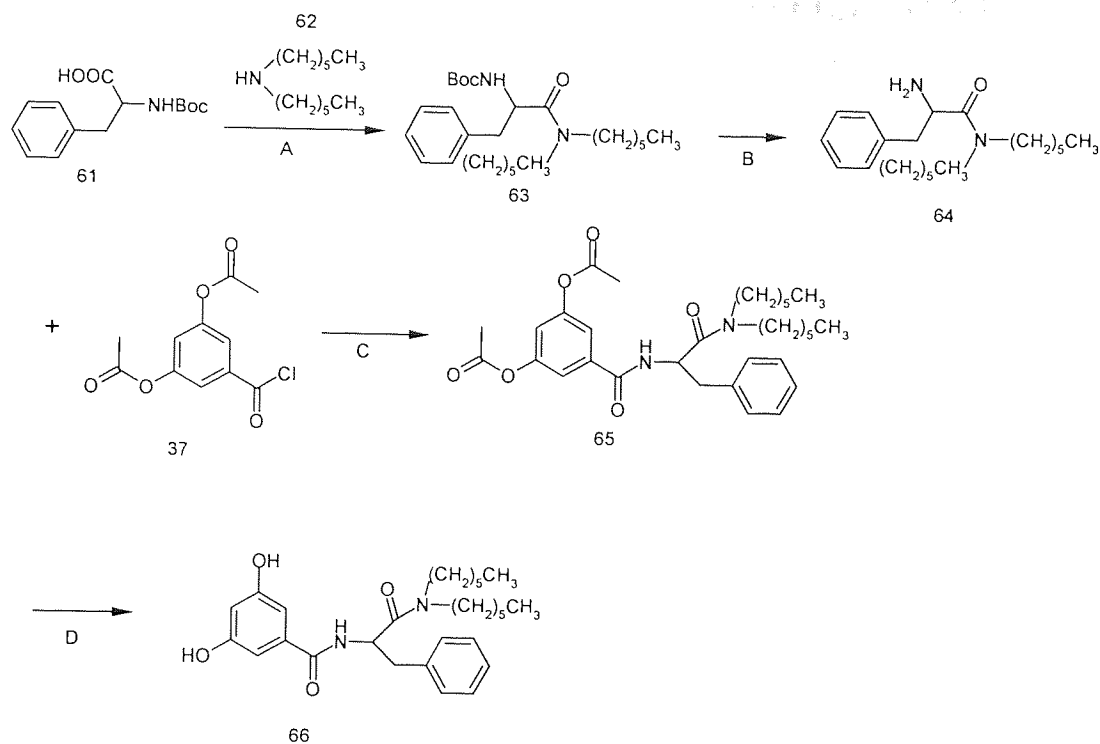
As described for **46**. M.p. 148-150 °C. Yield is not available.

<sup>1</sup>H NMR (DMSO): δ 2.51 (4H, s, 2 x CH<sub>2</sub>N), 3.40 (4H, s, 2 x NCH<sub>2</sub>), 3.52 (2H, m, CH<sub>2</sub>Ph), 5.99 (2H, s, OCH<sub>2</sub>O), 6.15 (2H, m, 2 x ArH), 6.25 (1H, m, ArH), 6.76-6.87 (3 x ArH), 9.50 (2H, s, 2 x OH).

<sup>13</sup>C NMR (DMSO): δ 59.18 (2 x NCH<sub>2</sub>), 61.5 (2 x CH<sub>2</sub>N), 100.29 (CH<sub>2</sub>Ph), 103.35 (OCH<sub>2</sub>O), 104.72 (2 x ArC), 107.86 (2 x ArC), 109.15 (ArC), 122.09 (ArC), 131.61 (ArC), 137.75 (ArC-CO), 146.27 (ArC-O), 147.27 (ArC-O), 158.44 (2 x ArC-OH), 168.95 (CO).

IR (KBr disc): ν<sub>max</sub> (3209, 2959, 2842, 2368, 1607, 1522, 1346, 1233, 1152, 1004, 842, 801, 756, 670, 531 cm<sup>-1</sup>).

m/z (APCI) 357 [M + H]<sup>+</sup>



- A) DCC, TEA in DCM at 0 °C for 6 hours.  
 B) TFA in DCM r.t. for 1 hour.  
 C) Pyridine in DCM r.t. for 2.5 hours.  
 D) 2M NaOH r.t for 45 mins.

#### Synthesis of Boc-Phenylalanine-dihexylamide **63**:

To Boc-Phe-OH **61** (0.8g, 3 mmol), DCC (679 mg, 3.3 mmol) and TEA (0.5 ml, 3.3 mmol) in DCM 15 ml at 0 °C for 0.5 hours was added dihexylamine **62** (0.7 ml, 3.3 mmol). Stirring was continued for 6 hours. The mixture was extracted with DCM and 2M HCl, washed with 2M HCl, saturated Na<sub>2</sub>CO<sub>3</sub> and brine and dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed to afford the white oil. R<sub>f</sub> = 0.25 in DCM: MeOH (10:1). Yield is not available.

$^1\text{H}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  0.86 (6H, s, 2 x  $\text{CH}_3$ ), 1.25 (16H, m, 8 x  $\text{CH}_2$ ), 1.37 (9H, s, Boc), 3.17 (2H, m,  $\text{CH}_2\text{Ph}$ ), 3.39 (4H, m, 2 x  $\text{NCH}_2$ ), 4.71 (1H, m,  $\text{CHNH}$ ), 5.34 ( $\text{CONH}$ ), 7.18 (5H, m, 5 x  $\text{ArH}$ ).

$^{13}\text{C}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  13.9 (2 x  $\text{CH}_3$ ), 22.47 (2 x  $\text{CH}_2$ ), 26.27 (2 x  $\text{CH}_2$ ), 28.18 (2 x  $\text{CH}_2$ ), 28.37 (3 x  $\text{CH}_3$  in Boc), 31.22 (2 x  $\text{CH}_2$ ), 34.80 ( $\text{CH}_2\text{Ph}$ ), 45.97 (2 x  $\text{CH}_2$ ), 48.90 ( $\text{CHCO}$ ), 78.72 ( $\text{C}(\text{CH}_3)_3$ ), 126.93 ( $\text{ArC}$ ), 128.39 ( $\text{ArC}$ ), 128.56 ( $\text{ArC}$ ), 129.10 ( $\text{ArC}$ ), 129.42 ( $\text{ArC}$ ), 136.68 ( $\text{ArC}-\text{CH}_2$ ), 154.79 ( $\text{CO}$ ), 171.17 ( $\text{CO}$ ).

IR (KBr disc):  $\nu_{\text{max}}$  (3324, 2932, 2842, 2117, 1711, 1621, 1459, 1360, 1247, 1175, 1053, 891, 688  $\text{cm}^{-1}$ ).

$m/z$  (APCI) 355  $[\text{M} + \text{H}]^+$

#### Synthesis of Phenylalanine-dihexylamide **64**:

To a solution of **63** (432mg, 1 mmol) in DCM 10 ml was added TFA 4 ml. Stirring was continued for 1 hour. The product was extracted with DCM and 10 %  $\text{Na}_2\text{CO}_3$ , washed with 10 %  $\text{Na}_2\text{CO}_3$  and brine, dried over  $\text{Na}_2\text{SO}_4$ . The solvent was removed to afford the yellow oil. Yield is not available.

$^1\text{H}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  0.86 (6H, s, 2 x  $\text{CH}_3$ ), 1.25 (16H, m, 8 x  $\text{CH}_2$ ), 3.17 (2H, m,  $\text{CH}_2\text{Ph}$ ), 3.39 (4H, m, 2 x  $\text{NCH}_2$ ), 4.71 (1H, m,  $\text{CHNH}$ ), 7.18 (5H, m, 5 x  $\text{ArH}$ ).

$^{13}\text{C}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  13.85 (2 x  $\text{CH}_3$ ), 22.40 (2 x  $\text{CH}_2$ ), 26.67 (2 x  $\text{CH}_2$ ), 29.18 (2 x  $\text{CH}_2$ ), 31.30 (2 x  $\text{CH}_2$ ), 33.85 ( $\text{CH}_2\text{Ph}$ ), 46.38 (2 x  $\text{CH}_2$ ), 48.36 ( $\text{CHCO}$ ), 126.53 ( $\text{ArC}$ ), 128.36 ( $\text{ArC}$ ), 128.50 ( $\text{ArC}$ ), 129.10 ( $\text{ArC}$ ), 129.22 ( $\text{ArC}$ ), 137.86 ( $\text{ArC}-\text{CH}_2$ ), 156.87 ( $\text{CO}$ ), 174.20 ( $\text{CO}$ ).

IR (KBr disc):  $\nu_{\text{max}}$  (3327, 2936, 2864, 2357, 1711, 1621, 1459, 1374, 1302, 1239, 898, 701  $\text{cm}^{-1}$ ).

#### Synthesis of **65**:

To a solution of **37** (691 mg, 2.1 mmol) and **64** (700mg, 2.1 mmol) in DCM 30 ml was added pyridine (0.8ml, 10 mmol). Stirring was continued for 2.5 hours. The

product was extracted with DCM and 2M HCl, washed with 2M HCl, saturated Na<sub>2</sub>CO<sub>3</sub> and brine and dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed to afford the white solid. M.p. 125-127 °C. Yield is not available.

<sup>1</sup>H NMR (CDCl<sub>3</sub>): δ 0.83 (6H, m, 2 x CH<sub>3</sub>), 1.22 (16H, m, 8 x CH<sub>2</sub>), 2.26 (6H, s, 2 x CH<sub>3</sub>), 3.16 (2H, m, CH<sub>2</sub>Ph), 3.39 (4H, m, 2 x NCH<sub>2</sub>), 4.51 (1H, m, CHCO), 7.20 (8H, m, 8 x ArH).

<sup>13</sup>C NMR (CDCl<sub>3</sub>): δ 13.89 (2 x CH<sub>3</sub>), 22.12 (2 x CH<sub>3</sub>), 22.44 (2 x CH<sub>2</sub>), 26.62 (2 x CH<sub>2</sub>), 28.70 (2 x CH<sub>2</sub>), 31.14 (2 x CH<sub>2</sub>), 33.75 (CH<sub>2</sub>Ph), 45.75 (2 x CH<sub>2</sub>), 48.50 (CHCO), 112.14 (ArC), 112.17 (2 x ArC), 126.53 (ArC), 128.32 (ArC), 129.67 (ArC), 134.57 (ArC), 135.5 (ArC), 136.41 (ArC-CO), 157.51 (ArC-CH<sub>2</sub>), 158.78 (2 x ArC-O), 165.58 (CO), 170.85 (2 x CO), 171.10 (CO).

IR (KBr disc): ν<sub>max</sub> (3313, 2914, 2847, 2323, 1747, 1717, 1371, 1207, 1038, 888, 746 cm<sup>-1</sup>).

Synthesis of N-[1-benzyl-2-(dihexylamino-2-oxoethyl)]-3,5-dihydroxybenzamide **66**:

To a solution of **65** (650 mg, 1 mmol) in THF 10 ml was added 10 ml 2 M NaOH. Stirring was continued for 45 minutes. PH was adjusted to 1 by adding 2M HCl. The product was extracted with EtOAc and 2M HCl, washed with 2M HCl and brine and dried over Na<sub>2</sub>SO<sub>4</sub>. Removal of the solvent afforded the yellow oil. R<sub>f</sub> = 0.2 in DCM: MeOH (10:1). Yield is not available.

<sup>1</sup>H NMR (CDCl<sub>3</sub>): δ 0.83 (6H, m, 2 x CH<sub>3</sub>), 1.22 (16H, m, 8 x CH<sub>2</sub>), 3.16 (2H, m, CH<sub>2</sub>Ph), 3.39 (4H, m, 2 x NCH<sub>2</sub>), 4.51 (1H, m, CHCO), 7.20 (8H, m, 8 x ArH).

<sup>13</sup>C NMR (CDCl<sub>3</sub>): δ 13.91 (2 x CH<sub>3</sub>), 22.42 (2 x CH<sub>2</sub>), 26.27 (2 x CH<sub>2</sub>), 28.69 (2 x CH<sub>2</sub>), 31.22 (2 x CH<sub>2</sub>), 33.67 (CH<sub>2</sub>Ph), 45.97 (2 x CH<sub>2</sub>), 48.90 (CHCO), 105.34 (ArC), 106.05 (2 x ArC), 126.93 (ArC), 128.39 (ArC), 129.54 (ArC), 134.64 (ArC), 135.80 (ArC), 137.68 (ArC-CO), 158.09 (ArC-CH<sub>2</sub>), 158.39 (2 x ArC-OH), 166.93 (CO), 171.17 (CO).

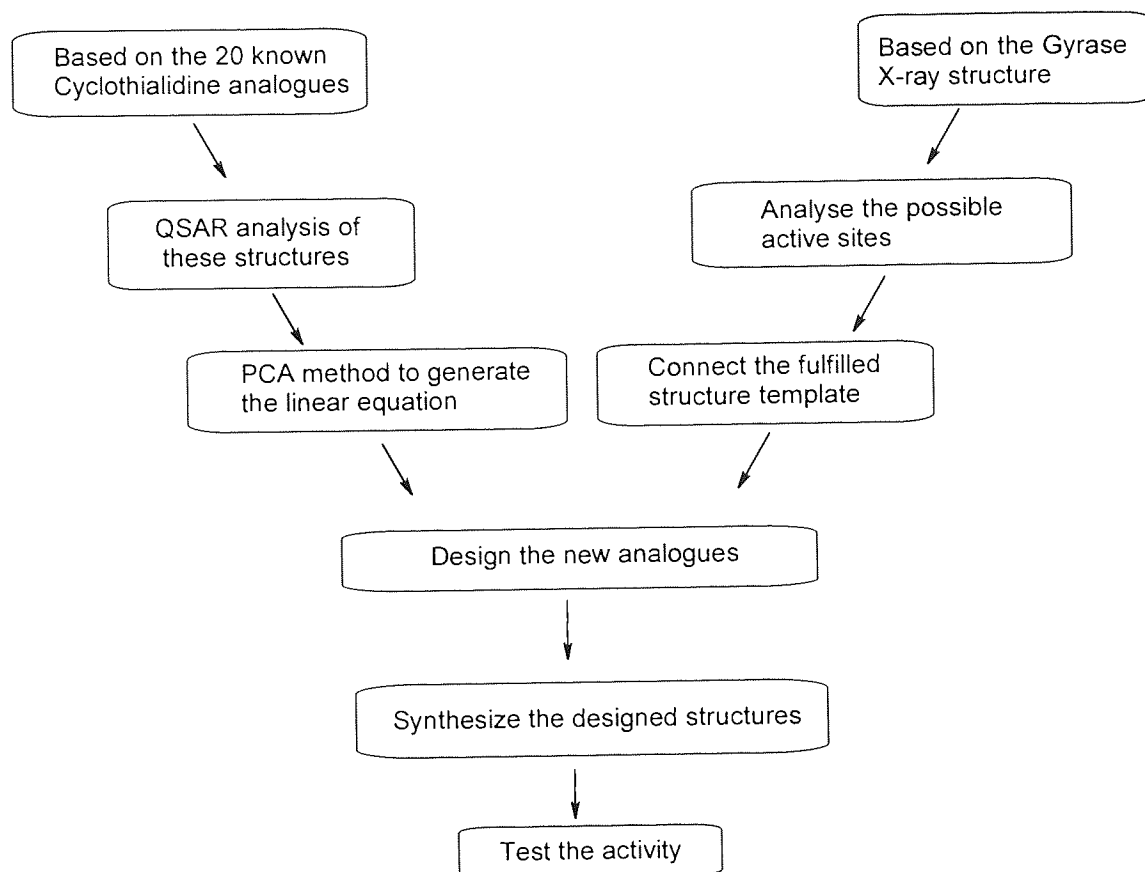


IR (KBr disc):  $\nu_{\max}$  (3226, 2930, 2850, 1598, 1497, 1457, 1373, 1147, 1001, 864, 767, 710, 519).

$m/z$  (APCI) 469  $[M + H]^+$

## **Chapter 4. Results and Discussion**

## 1. The whole plan.



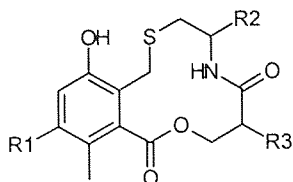
As the science is always tested by different methods, we tried to perform two kinds of widely used molecular modeling methods: QSAR analysis and structure based ligand design to discover new compounds, which could show better activity against bacteria.

QSAR analysis is performed based on the reported 20 Cyclothialidine analogues. The properties of the two substituent groups ( $R_2$  and  $R_3$ ) are calculated and analyzed with PCA (Principal Component Analysis) to generate a linear equation, which can be used to predict the activity of the designed structures. In contrast the structure-based drug design focuses on the enzyme's structure. The computer builds up the drug structure's skeleton according to the possible binding

site, then joins these templates to be a new compound. The structures discovered by these two methods are synthesized and tested for the MICs.

## 2. QSAR analysis.

It is surprising that so few analogues have been reported since Cyclothialidine was first discovered in 1994. The only 20 compounds with the completed gyrase and antibacterial activity testing were cited from E. Goetschi (126) in **Table 1**.



No.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	Gyrase Inhibitor Activity	MNEC (ug/ml)	In vitro Antibacterial Activity MIC (ug/ml)			
						<i>E.coli</i> DC2	<i>N.men</i> 69480	<i>S.aureus</i> 887	<i>M.luteus</i> ATCC8340
Cyclo	OH	NH-Ala-OH	NH-3Hyp-Ser		0.05	>	128	>	128
43-4106	OH	COOMe	NH-3Hyp-Boc		0.05	>	64	>	8
42-9416	OH	COOMe	NHBoc		0.5	>	16	>	16
43-3386	OH	COOMe	NHAc		0.4	>	32	>	128
43-4109	OH	COOMe	H		0.2	>	32	>	16
43-9052	OMe	COOMe	NHBoc		0.2	>	16	64	16
44-4731	OMe	COOMe	NHAc		0.2	64	8	64	16
46-2252	OMe	COOMe	NHCHO		0.1	32	2	2	8
46-2592	OMe	COOMe	NHCOOMe		0.2	64	8	32	4
46-5734	OMe	COOMe	NH <sub>2</sub>		1	>	128	>	64
47-0268	OMe	COOMe	OH		0.1	64	8	4	16
44-4728	OMe	COOMe	H		0.5	64	16	32	32
44-4729	OMe	CH <sub>2</sub> OH	H		>5	>	128	>	128
44-4730	OMe	CH <sub>2</sub> OAc	H		2.5	128	16	64	64
47-0143	OMe	CONHCH <sub>2</sub> CHCH <sub>2</sub>	NHBoc		0.1	>	4	16	8
46-4229	OMe	CONHCH <sub>2</sub> CH <sub>2</sub> OH	NHBoc		0.2	>	4	>	32
47-1056	OMe	CONHCH <sub>2</sub> CHCH <sub>2</sub>	NHAc		0.05	128	4	>	32
44-6974	OMe	CONHCH <sub>2</sub> CH <sub>2</sub> COOH	NHAc		0.1	>	>	>	>
46-5216	OMe	CONHCH <sub>2</sub> CH <sub>2</sub> OMe	NHCHO		0.1	64	4	128	128
46-9288	OMe	COOEt	NHCSMe		0.2	64	2	2	4

Supercoiling assay using purified Gyrase from *Escherichia coli*.

MNEC: Maximum Non-effect Concentration; highest inhibitor concentration at which no DNA gyrase inhibition can be observed visually.

">" = > 128 µg/ml; was assumed to be 256 µg/ml in the calculation.

MIC: Minimal Inhibitory Concentration; lowest inhibitor concentration at which the bacterial activity is completely inhibited visually.

**Table 1. The structures and antibacterial activity of Cyclothialidine analogues**

In most researches, the *E.coli* and *S. aureus* are usually selected as the samples for Gram-negative and Gram-positive bacteria. However, the activities to these two most widely used strains in our table are not ready to be analyzed yet. The “>” signal means that no activity has been found within the scope of the tests, we could only suppose it was 256 µg/ml in our statistical calculation. These inaccurate data limit the accuracy of the equation. Fortunately, *N. meningitidis 69480* and *M. luteus ATCC8340* as the Gram negative and Gram-positive coccus, are generally more sensitive. But *N. Meningitidis 69480* is seldom used for the testing because of the difficult growth and dangerous handling. However, in this thesis they have been used for the QSAR analysis as the samples for Gram-negative and Gram-positive bacteria, because their activities to Cyclothialidine analogues are amenable to numeric testing without too much assumption.

Another good idea in the analysis is using µmol/ml instead of µg/ml, which could prevent the undesired influences made by the compounds’ molecular weight. The minus logarithm of the activity was set as Y in the equation, which has been proved to more similar to the biological system. From these hypotheses, the more useful **Table 2** is displayed:

No.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	Gyrase Inhibitor Activity log 1/MNEC ( $\mu\text{mol/ml}$ )	In vitro Antibacterial Activity log 1/MIC ( $\mu\text{mol/ml}$ )	
					<i>N.men</i> 69480	<i>M.luteus</i> ATCC8340
Cyclo	OH	NH-Ala-OH	NH-3Hyp-Ser	4.108	0.7	0.7
43-4106	OH	COOMe	NH-3Hyp-Boc	4.076	0.969	1.872
42-9416	OH	COOMe	NHBoc	2.986	1.481	1.481
43-3386	OH	COOMe	NHAc	3.027	1.124	0.522
43-4109	OH	COOMe	H	3.266	1.062	1.363
43-9052	OMe	COOMe	NHBoc	3.396	1.493	1.493
44-4731	OMe	COOMe	NHAc	3.342	1.74	1.439
46-2252	OMe	COOMe	NHCHO	3.629	2.328	1.726
46-2592	OMe	COOMe	NHCOOMe	3.358	1.756	2.057
46-5734	OMe	COOMe	NH <sub>2</sub>	2.6	0.493	0.794
47-0268	OMe	COOMe	OH	3.601	1.698	1.397
44-4728	OMe	COOMe	H	2.884	1.379	1.078
44-4729	OMe	CH <sub>2</sub> OH	H	1.851	0.443	0.443
44-4730	OMe	CH <sub>2</sub> OAc	H	2.198	1.392	0.79
47-0143	OMe	CONHCH <sub>2</sub> CHCH <sub>2</sub>	NHBoc	3.718	2.116	1.813
46-4229	OMe	CONHCH <sub>2</sub> CH <sub>2</sub> OH	NHBoc	3.421	2.119	1.217
47-1056	OMe	CONHCH <sub>2</sub> CHCH <sub>2</sub>	NHAc	3.968	2.065	1.162
44-6974	OMe	CONHCH <sub>2</sub> CH <sub>2</sub> COOH	NHAc	3.696	0.288	0.288
46-5216	OMe	CONHCH <sub>2</sub> CH <sub>2</sub> OMe	NHCHO	3.671	2.069	1.768
46-9288	OMe	COOEt	NHCSMe	3.371	2.371	2.07

**Table 2. The logarithm of molar activity**

The R<sub>2</sub> and R<sub>3</sub>'s parameters are calculated in the TSAR program, followed by the first filtration: correlation matrix analysis. This method aims to detect the highly correlated data, which could have brought mistakes in the next multiple regression analysis. The matrix (**Table 3**) shows the correlation coefficients for the data; the red stands for the highest correlation, whose coefficients are above 0.9. The yellow is for the medium and blue for the low correlation, whose coefficients are around 0.8 and below 0.5, respectively.

	Molecular Mass (Subst. 2)	Molecular Mass (Subst. 3)	Molecular Surface Area (Subst. 2)	Molecular Surface Area (Subst. 3)	Molecular Volume (Subst. 2)	Molecular Volume (Subst. 3)	Verloop L (Subst. 2)	....
Molecular Mass (Subst. 2)	1	0.35564	0.96778	0.33887	0.98991	0.34341	0.86174	
Molecular Mass (Subst. 3)	0.35564	1	0.27453	0.99153	0.3184	0.99737	0.26588	
Molecular Surface Area (Subst. 2)	0.96778	0.27453	1	0.27207	0.98061	0.26973	0.87764	
....								

**Table 3. The correlation matrix for all the parameters (partly displayed)**

The reason for this analysis is that if we do not know the correlations between any two parameters, we cannot tell whether a high correlation between the activity and a parameter is meaningful or merely the result of high correlation with another parameter. So we must make sure to select only one parameter out of a set of highly correlated parameters.

After a careful overview of the whole matrix, we can divide all these data into three fields: (I) the steric effect, which is including structure mass, structure surface area, structure volume and Verloop parameter. (II) The charge effect: total dipole, dipole moment component, substituent bond dipole, and (III) the lipophilic effect: total lipole, lipole components, substituent bond lipole, Log P, molar refractivity. It is reasonable to select one parameter in each field, which is proposed to be structure volume, dipole moment component and Log P, respectively.

However, there are still too many data left. The PCA (Principal Component Analysis) is used for the further data reduction. The parameter and principal component matrix is generated in TSAR as **Table 4**.



	Principal Comp. 1	Principal Comp. 2	Principal Comp. 3	Principal Comp. 4	Principal Comp. 5
Molecular Volume (Subst. 2)	-0.34894	0.36497	-0.08288	-0.29567	-0.047536
Molecular Volume (Subst. 3)	-0.36961	0.070506	0.50075	0.096785	-0.02381
Dipole Moment X Component (Subst. 2)	-0.023118	-0.4833	0.393	-0.27834	0.093304
Dipole Moment X Component (Subst. 3)	0.41789	0.053717	0.24478	0.045544	0.5947
Dipole Moment Y Component (Subst. 2)	-0.34747	-0.33759	-0.2923	0.32701	0.16365
Dipole Moment Y Component (Subst. 3)	-0.24576	0.27119	0.36479	0.54739	-0.26994
Dipole Moment Z Component (Subst. 2)	0.10289	-0.44858	-0.27245	0.44774	-0.18017
Dipole Moment Z Component (Subst. 3)	-0.36554	-0.30675	0.26905	0.0573	0.39558
log P (Subst. 2)	0.28178	-0.30544	0.35462	-0.15768	-0.57779
log P (Subst. 3)	0.40215	0.22188	0.1927	0.43414	0.11121
Fraction of variance explained	0.33341	0.23943	0.17035	0.096458	0.068974
Total variance explained	0.33341	0.57284	0.74319	0.83964	0.90862

**Table 4. The parameters and principal component matrix**

These five principal components can explain 90 % of the whole possibility. Generally, the principal component PC1 is the most important factor (explains 33 %), and the larger coefficient means the more contribution to the component. After numerous attempts, the Dipole Moment Y axis Component (Substituent 2) whose coefficient is  $-0.34747$ , Dipole Moment X axis Component (Substituent 3), coefficient  $0.41789$ , and Log P (Substituent 3), coefficient  $0.40215$ , are proved to be the best equation generation for *N. Meningitidis*, an example of gram negative bacteria (**Equation 1**).

**Equation 1: Linear equation for gram-negative bacteria.**

$$Y=0.75X_1+0.63X_2-0.33X_3+1.96$$

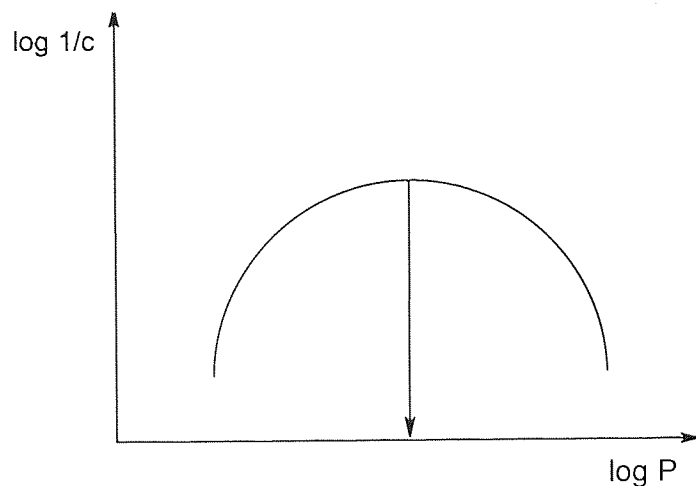
$$R^2=0.76 \quad s=0.32 \quad n=20$$

$X_1$ : Dipole moment X (substituent 3)

$X_2$ : Dipole moment Y (substituent 2)

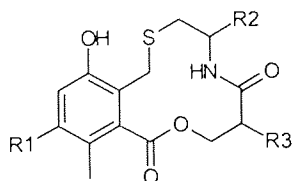
$X_3$ : Log P (substituent 3)

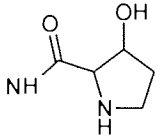
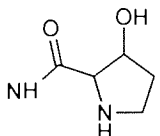
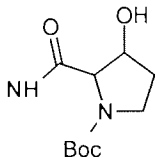
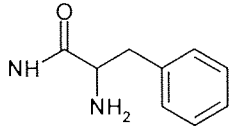
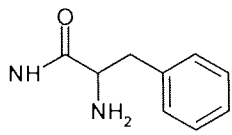
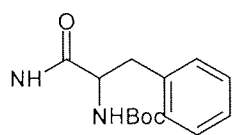
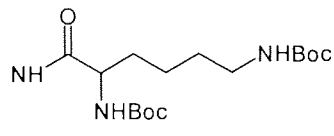
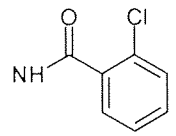
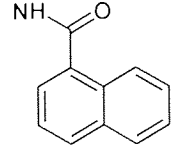
This equation shows the substituent 2 and 3 dipole moments give a positive effect to the activity, while the Log P gives the negative one. It is surprising because the membrane is generally proposed to be a lipophilic environment, which should require a larger Log P. Usually the more lipophilic, the easier to penetrate. According to Hansch's equation (68), the graph between the activity and Log P would be a curve, there is an optimal value of the hydrophobicity: too low and the compounds would not partition into the cell membranes; too high and the compound would partition into the membranes but tend to remain there rather than proceeding to the actual target. So we supposed that the log P data for these 20 analogues are only ranged in the right half of the curves.



The R square of this equation is 0.76, while the cross validation R square is 0.24. This means that the linear regression is acceptable ( $R^2 = 0.76$ ), but the prediction ability of this equation is lower ( $R^2_{\text{cross}} = 0.24$ ). Anyway, as the early equation had been generated, we could do further work to modify it. Depending on the equation, we synthesized the new analogues, which are mainly modified at the substituent 3 position to try to explore the more active compounds (**Table 5**).

The same procedure is repeated to the gram-positive bacteria (the *M. luteus*). Unfortunately, there was no linear relationship found between any of the parameters and the activity.



No.	R <sub>2</sub>	R <sub>3</sub>	Dip. Comp. X (R <sub>3</sub> )	Dip. Comp. Y (R <sub>2</sub> )	Log P R <sub>3</sub>	Predicted Activity Log <sub>1</sub> /MIC (umol/ml)	Actual Activity Log <sub>1</sub> /MIC (umol/ml)
19	COOMe		-0.6841	1.3465	-1.504	2.79	-0.01
24	COOEt		-3.3171	-1.2453	-1.504	-0.815	0
18	COOMe		1.3409	1.4721	-0.380	4.019	--
30	COOMe		-2.0522	-1.261	0.6135	-0.576	-0.02
31	COOEt		-2.129	-1.233	0.6135	-0.616	0.03
27	COOEt		-0.5713	-1.1623	1.8992	0.173	--
33	COOEt		11.189	-0.1821	1.2657	9.819	--
58	COOEt		3.5709	-1.5139	1.7242	3.115	0.64
59	COOEt		4.3598	-1.2705	2.2084	3.707	0.64

-- No activity was observed within 512 ug/ml

Table 5. The new analogues' activity to gram-negative bacteria

In **Table 5**, the predicted activity is measured by the minus logarithm of micromoles of compound per milliliter. The greater predicted activity number means the smaller MIC, and the better activity. So in the designed new compounds, some should have quite good activity, which was predicted up to  $1.57 \times 10^{-10}$   $\mu\text{mol/ml}$  (Compound **33**) if tested by the NCCLS standards (See Methods). However, the experimental antibacterial activity testing shows not many of the new compounds have significantly improved activity to *E. coli* DC2 and *S. aureus* OXFORD, which are the most sensitive strains to antibiotic. There is a slightly improved activity for the compounds **58** and **59**, possibly because of the greater Log P. While compound **33** does not have the actual activity, which predicted activity is even bigger than **58** and **59**. The reason could be that the structure is too big to come through the membrane.

The reasons for this disappointing analysis were concluded that: First of all, the QSAR analysis, especially the multiple linear equation always depends on the samples' numbers as well as a well planned range of data. If we have either a great enough number of samples or well ranged data, we have more chances to get satisfactory results. The researches that have been done for this kind of potential antibiotic are nowhere near enough, and the reported 20 analogues' substituents have big similarities between each other. In other words, the data are gathered in a very small scope. The linear equation generated by these data will surely have poor predictive abilities. We did not intend to apply the non-linear analysis, such as neural networks and genetic algorithms, which have been developed and proved to be successful in 1990's (69) because although the new technologies can give the better prediction, it is difficult to know what are the significant parameters to the activity. While the linear equation clearly shows the important parameters, the dipole component and the log P for example, so it is possible to change them through modifying the substituents.

Another reason could be the fact that we generate the equation by analysis of the strains *N. meningitidis* 69480 and *M. luteus* ATCC8340, but test the activity by *E. coli* DC2 and *S. aureus* OXFORD. This is because in the previous papers, there is no activity shown to the *E. coli* and *S. aureus*, but obvious MIC to *N.*

*meningitidis* and *M. luteus*, so we have to use these two as the starting analysis. It is possible that although *E. coli*, *N. meningitidis* are gram-negative bacteria and *M. luteus*, *S. aureus* are gram-positive bacteria, they still have a slight difference in their membranes; for example, some big antibiotic can come through *N. meningitidis*, but never did to *E. coli*. However, the *N. meningitidis* is too dangerous to handle in the lab, so we cannot test the MICs to this organism.

The last reason is that it is postulated that the *in vitro* antibacterial activity of DNA gyrase inhibitor can be separated into two components: (i) penetration through the bacterial membrane and (ii) inhibition of bacterial DNA gyrase. According to this hypothesis, if we selected the cyclothialidine analogues, which show the similar gyrase inhibitory activity, but different *in vitro* antibacterial activity, then we can regard the MIC as the membrane penetration properties. However, the anti-gyrase activity testing is not as easy as MIC's. We cannot do this in our university and cannot find a collaborator in other research groups either, so this plan had to be given up.

### **3. Structure-based ligand design.**

In Lewis's crystal structure of Cyclothialidine with DNA gyrase (140), the active sites of the gyrase were defined and the resorcinol ring was identified as the pharmacophore (**Figure 14 and Figure 15**)

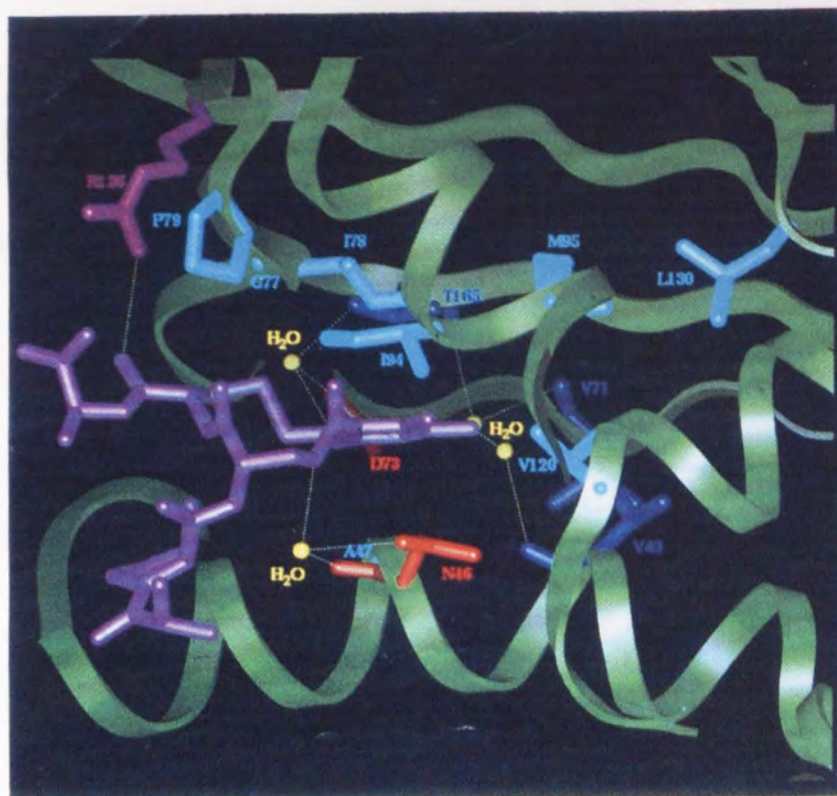


Figure 14. The crystal structure of Cyclothialidine with gyrase B

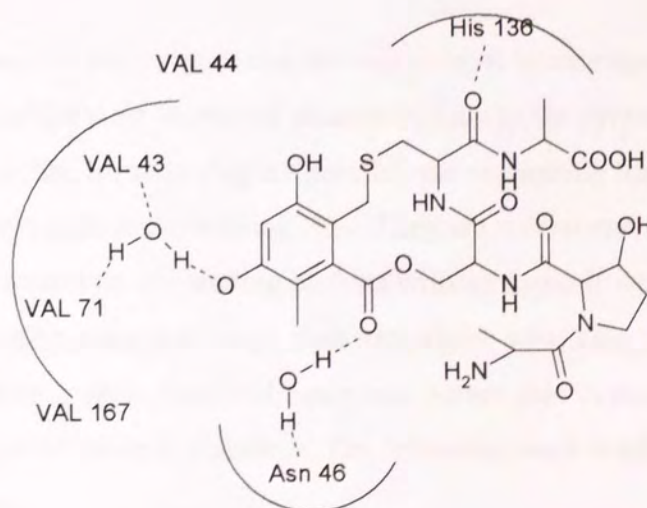
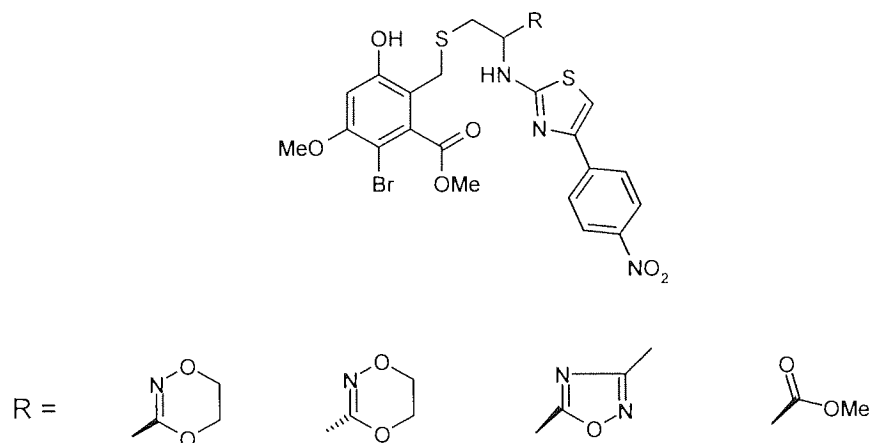


Figure 15. The 2D structure of Cyclothialidine with some binding sites

In the recent reports, Rudolph (127) and co-workers in Bayer Company synthesized four seco-Cyclothialidine analogues, which opened the 12-membered lactone ring. And three of them have good MIC activities to gram-positive bacteria as well as the gyrase activity. (Figure 16)



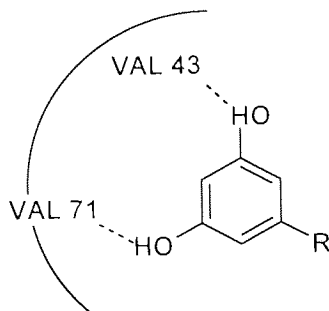
**Figure 16. Rudolph's seco-Cyclothialidine structures**

Based on these studies, we can suppose that the two oxygen substituents in the resorcinol ring comprise the most important pharmacophore to the gyrase's target sites. The side chains on the benzene ring are possibly the supporting fragments to help deliver the compounds to the binding sites. They are crucial and the compounds could not come into the binding pockets without them. If we keep the pharmacophore groups (the resorcinol ring), then investigate what kind of side chains can help the compounds to bind to the enzyme, we are able to modify the pharmacophore to discover the new inhibitors. The following work is all based on the computer programs.

Let's first examine Rudolph's structure. The thiazole and benzene rings are believed to improve the hydrophobicity, which could be helpful through the membrane. The two heterocyclic rings are providing the possible hydrogen bond

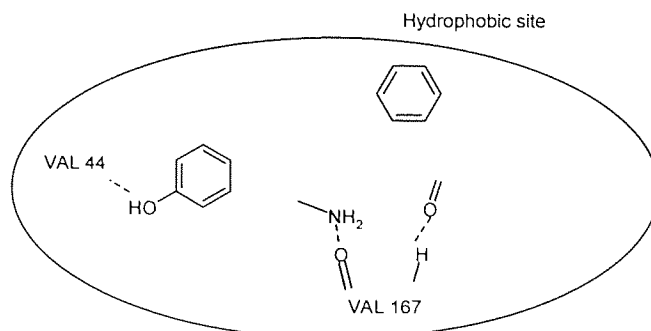


donor and acceptor as well as the hydrophobic region. According to this hypothesis, we propose the key pharmacophore for Gyrase B is (**Figure 17**):



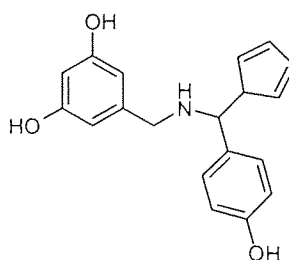
**Figure 17. The key pharmacophore for Gyrase B**

The two amino acid residues VAL 44 and VAL 167 in gyrase were set as the active sites instead of the VAL 43 and VAL 71. This was based on the SPROUT program. If, as in this case, the ligand is binding with the enzyme, the ligand would be set as the probe to explore the active sites. If the ligand is unknown, the binding sites (VAL 43 and VAL 71) would be used to explore the active sites. The VAL 44 and VAL 167 were generated by the ligand, which was in unknown situation. All the amino acids within 10 Å radius around the active site were included for analyzing. The SPROUT program then generated a possible binding pocket, with two hydrogen bond acceptors, one hydrogen bond donor and one hydrophobic site in it (**Figure 18**). Definition of the pocket was followed by the proper templates to fulfill, for example, the phenol and amine was defined as the hydrogen bond donors, the carbonyl as the hydrogen bond acceptor and the aromatic ring as the hydrophobic group.

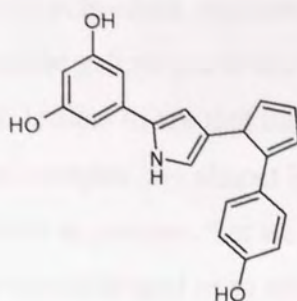


**Figure 18. The fulfilled templates to the correlated sites**

The computer then used the necessary group fragments to connect all the fulfilled structure templates together subject to the steric constraints, such as the staggered bond conformation and proper bond length. More than one structure was generated. There are two main clusters: one has side-chains in the N-terminal, in which the hydrogen acts as the hydrogen donor. One chain raises the aromatic ring up to the hydrophobic site, the other chain comes down to reach the hydrogen bond acceptor. The structure is **Figure 19a**.



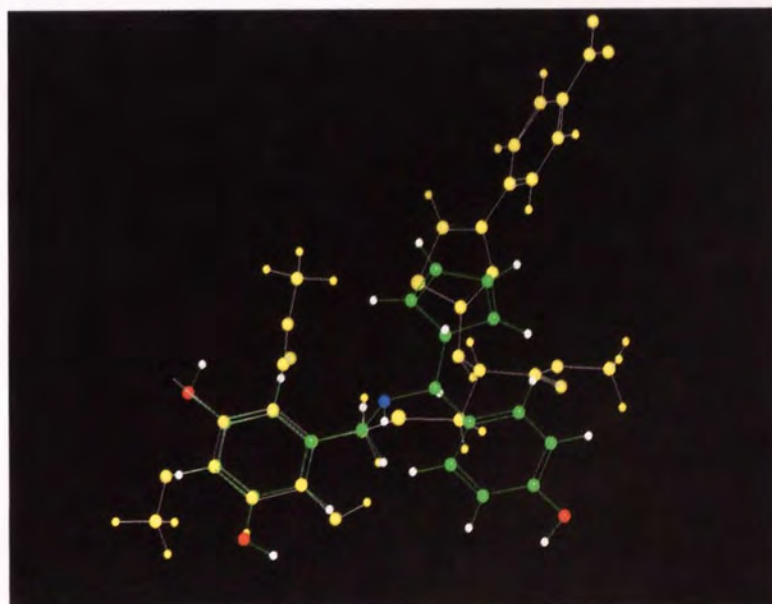
**Figure 19a**



**Figure 19b**

**Figure 19a and 19b. The designed structures by SPROUT**

The other kind of structure is a long chain **Figure 19b**, which needs a special conformation to raise the hydrophobic site up, then twist back to reach the hydrogen bond acceptor. The **Figure 19a** is found to be very similar to the Rudolph's structure; the superimposed structures of these two are displayed in **Figure 20**.

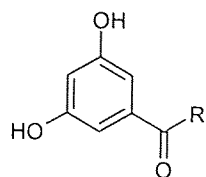


**Figure 20. The Superimposition of Figure 19a and Rudolph's structure**

The green compound is the side-chain structure (**Figure 19a**) after minimizing the energy by MOPAC and the yellow one is the Rudolph's structure with R=COOMe (**Figure 20**). It is clear to see that these two significant sites, which are hydrophobic and hydrogen acceptor, are almost lying in the same position. These results gave us the confidence to propose that we could have found a correct direction in the research and encouraged us to synthesize the designed structure.

### 3.1 Figure 19b leading synthesis

The straight chain structures are designed to have the amide, which is the potential hydrogen bond donor, with the hetero atom and the aromatic rings. Some structures with the secondary amine have to be protonated in order to serve as a hydrogen bond donor, which should be possible since they are good bases. Our aim is to generate a 3D pharmacophore through exploring the distance between the hydrophobic site and the hydrogen bond by changing the length of the linear chain. The synthesized structure and their MICs are in the **Table 6**.



No.	Structure R	<i>S.aureus</i>	<i>E.coli</i>
51		>512	>512
45		>512	>512
52		512	512
53		>512	>512
54		>512	>512
46		>512	>512
42		>512	>512
43		>512	>512
49		>512	>512
60		>512	>512

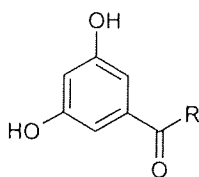
The MICs in this section are all tested with NCCLS standards (see Methods) as the unit  $\mu\text{g/ml}$

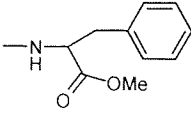
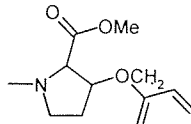
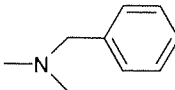
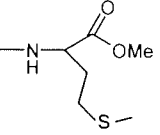
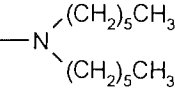
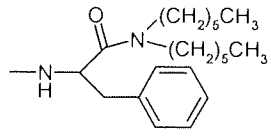
**Table 6. The synthesized linear chain analogues**

While screening the MICs, it is very disappointing that not many of them have the greatly improved activity. The reason could be there are too many single rotated bonds in the structures, so that there are too many possibilities for the exact conformations of the structures. It is very difficult to match the enzyme's conformation.

### 3.2 Figure 19a leading synthesis

The side chain structures are also designed to have the amide, with which the side chains connected to the resorcinol ring. Some structures with the secondary amine have to be protonated in order to serve as a hydrogen bond donor, which should be possible since they are good bases. The synthesized structures and tested MICs are in **Table 7**



No.	Structure R	S.aureus	E.coli
48		512	512
50		128	256
44		512	512
47		>512	>512
55		8	16
66		64	128

**Table 7. The synthesized side chain analogues**

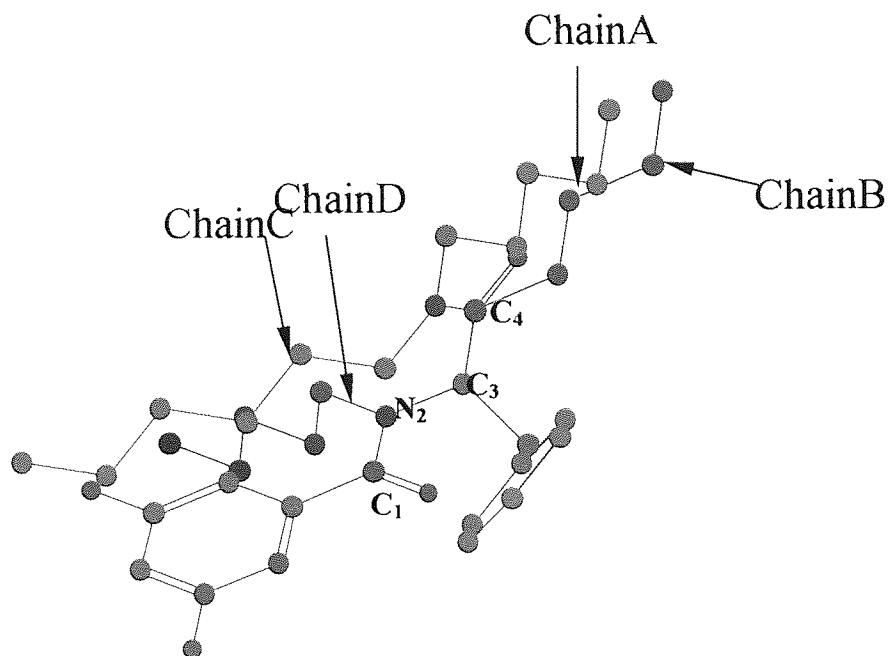
From the results, the structure **55** and **66** show the improved activity to both of the *E.coli* and *S.aureus*. The further MICs testing is performed to the strains *Micrococcus luteus*, *Enterococcus faecalis* (7JI), *Enterococcus faecium* (7JII), *Staphylococcus aureus* (7FII), *Staphylococcus aureus* (7FIV), *Klebsiella pneumoniae* (8DV), *Serratia marcescens* (8AIV) and *Pseudomonas aeruginosa*. The first five strains are gram-positive organisms, while the others are gram-negative organisms. The results are in **Table 8**.

Strains	Compounds	55 (ug/ml)	66 (ug/ml)
<i>Micrococcus luteus</i>		8	32
<i>Enterococcus faecalis</i> (7JI)		16	32
<i>Enterococcus faecium</i> (7JII)		512	512
<i>Staphylococcus aureus</i> (7FII)		512	512
<i>Staphylococcus aureus</i> (7FIV)		16	256
<i>Klebsiella pneumoniae</i> (8DV)		>512	>512
<i>Serratia marcescens</i> (8AIV)		>512	>512
<i>Pseudomonas aeruginosa</i>		64	32

**Table 8. The MICs to another eight organisms**

Although these two compounds lack the activity against the gram-negative organisms, except *Pseudomonas aeruginosa*, they do have quite good activity to gram-positive organisms. Therefore it is worthwhile to analyze the characters of **55** and **66**. In the further look at these two structures, we find they both contain the dihexylamine groups. The alkanes are believed to match the hydrophobic binding site on the enzyme. After minimizing the energy with MOPAC program, we fit them together in the Chem-X program to find out the further common characters between them. The picture is in **Figure 21**.

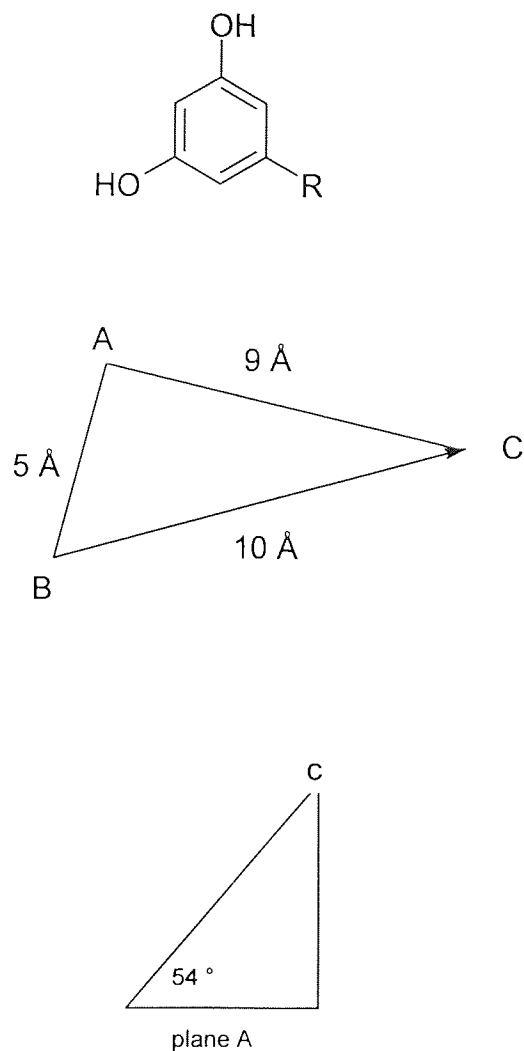




**Figure 21. The Superimposition of 55 and 66**

The green structure is **66** and the purple one is **55**. The resorcinol rings are fitted together exactly, as well as the following  $C_1$ ,  $N_2$ ,  $C_3$ ,  $C_4$  in **55** and **66**, which shows that these two compounds' conformations are very similar. While **55** extends one hexyl chain from  $N_2$  (**Chain D**), the other chain follows **66** exactly, until they separate after  $C_4$  (**Chain B**). The program shows both chains after  $C_4$  (**Chain A and Chain B**) are parallel. The distance between them is only 2-4 Å (about one or two  $sp^3$  carbon-carbon bond lengths). Otherwise, the other two chains (**Chain C and Chain D**) extend in different directions, the distance between them is up to 8 Å. For this reason, we think the parallel chains (**Chain A or Chain B**), which act as the hydrophobic sites, are also significant factors to the antibacterial activity.

After all this analysis, we generate a possible pharmacophore for the DNA gyrase B: it contains the two hydrogen bond donors and one hydrophobic site. In **Figure 22**:



**Figure 22. The possible pharmacophore of Gyrase B**

Points A and B stand for the two phenol groups, which are the hydrogen bond donors. Their distance should be about 5 Å. This is fairly rigid because of the benzene ring connecting them and is always the same in all of our compounds. Point C is the center of the hexyl group (hydrophobic site). The distance to A and B is 9 Å and 10 Å. The plane A is the benzene plane with the torsion angle 54 ° to the hydrophobic sites. So far, we can define the 3D conformation of the pharmacophore, which means that structures having the proper pharmacophore conformation could have the activities to DNA gyrase B.

A well-known example of a drug discovery procedure is the design of HIV protease inhibitors in Merck research group in 1994 (142), based on the 3D

pharmacophore. The 3D structure hits would be searched in the Cambridge Structural Database, which is also available in Dr. Carl Schwalbe's lab. In what could be the most difficult stage in the design flow chart, the full imagination would lead to a possible inhibitor's structure. There is still a lot of work to do after we can find a possible inhibitor candidate capable of synthesis. I wish my early work would save time for the following researchers and hopefully lead in a right direction for finding a new gyrase inhibitor.

## Chapter 5. References

1. **Lewis, R.A et al. (1997)** Similarity measures for rational set selection and analysis of combinatorial libraries: The diverse property-derived (DPD) approach. *J.Chem.Inf. Comput. Sci*, 37, 599-614.
2. **Rishton, G.M. (1997)** Reactive compounds and in vitro false positive in HTS. *Drug Discovery Today* 2, 382-384.
3. **Walters, W. P. et al. (1998)** Virtual screening – An overview. *Drug Discovery Today* 3, 160-178.
4. **Hann, M. et al. (1999)** Strategic pooling of compounds for high throughput screening. *J. Chem. Inf. Comput. Sci.* 39, 897-902.
5. **Navia, M.A. and Chaturvedi, P. R. (1996)** Design principles for orally bioavailable drugs. *Drug Discovery Today* 1, 179-189.
6. **Chan, O. H. and Stewart, B. H. (1996)** Physicochemical and drug-delivery considerations for oral drug bioavailability. *Drug Discovery Today* 1, 461-473.
7. **Kramer, S. D. (1999)** Absorption prediction for physicochemical parameters. *Pharm. Sci. Technol. Today* 2, 373-380; 2, 426.
8. **Moriguchi, I, et al. (1992)** Simple method for calculating octanol/water partition coefficient. *Chem. Pharm. Bull.* 40, 127-130.

9. **Dhose, A. K. et al. (1999)** A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem. 1*, 55-68.
10. **Veber D. F., Johnson S. R., Cheng H., Smith B. R., Ward K. W. and Kopple K. D. (2002)** Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*, 2002, 45, 2615-2623.
11. **Pardridge, W. M. (1998)** CNS drug design based on principles of blood-brain barrier transport. *J. Neurochem. 70*, 1781-1792.
12. **Kuntz I. D. (1992)** Structure-based strategies for drug design and discovery. *Science 257*, 1078-1082
13. **DiMasi J. -A., Bryant N. R. and Lasagna L. (1991)** New drug development in the United States from 1963 to 1990 *Clin. Pharmacol. Ther 50*, 471-486
14. **Gerysen H. M., Melon R. H. and Bartelme S. J. (1984)** Use of peptide synthesis to probe vital antigens for epitopes to a resolution of a single amino acid, *Proc. Natl, Acad. Sci. USA 81*, 3998-4002
15. **Gordon E. M., Barren R. W., Dower W. J., Foder S. P. A. and Gallop M. A. (1994)** Applications of combinatorial technologies to drug discovery, 2. Combinatorial organic synthesis, library screening strategies, and future directions, *J. Med. Chem, 37*, 1385-1400
16. **Schultz P.G. and Lerner R. A. (1995)** From molecular diversity to catalysis: lessons from the immune system. *Science 269*, 1835-1842

17. **Perez-Paya E., Houghton R. A. and Blondelle S. E. (1996)** Functionalised protein-like structures from conformationally defined synthetic combinatorial libraries. *J. Biol. Chem.* 271, 4120-4126.
18. **Thompson L. A. and Ellman J. A. (1996)** Synthesis and applications of small molecule libraries, *Chem. Rev.* 96, 555-600.
19. **Boyd D. B. (1990)** Successes of computer-assisted molecular design. *Rev. Comput. Chem.* 1, 355-371
20. **Navia M. A. and Murcko M. A. (1992)** Use of structural information in drug design. *Curr. Opin. Struct. Biol.* 2, 202-210
21. **Erickson J. W. and Fesik S. W. (1992)** Macromolecular X-ray crystallography and NMR as tools for structure-based drug design. *Annu. Rep. Med. Chem.* 27, 271-289
22. **Bugg C. E., Carson W. M. and Montgomery J.A. (1993)** Drugs by design. *Sci. Am.* 92-98.
23. **Peters R. and Mckinstry R. C. (1994)** Three-dimensional modelling and drug development: Has rational drug design arrived? *Biotechnology* 12, 147-150
24. **Whittle P. J. and Blundell T. L. (1994)** Protein structure-based drug design, *Annu. Rev. Biophys. Biomol. Struct.* 23, 349-375
25. **Hol W. G. J. and Verlinde C. L. M. J. (1994)** Structure-based drug design: Process, results and challenges. *Structure* 2, 577-587

26. **Greer J., Erickson J. W., Baldwin J. L. and Varney M. D. (1994)**  
Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* 37, 1035-1054
27. **Guida W. C. (1994)** Software for structure-based drug design. *Curr. Opin. Struct. Biol.* 4, 777-781
28. **Jackson R. C. (1995)** Update on computer-aided drug design. *Curr. Opin. Biotechnol.* 6, 646-651
29. **Desjarlais R., Sheridan R. P., Seibel, G. L. Dixon J. S., Kuntz I. D. and Venkataraghavan R. (1988)** Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.* 31, 722-720
30. **Kuntz, I. D. Meng E. C. and Shoichet B. K. (1994)** Structure-based molecular design. *Acc. Chem. Res.* 27, 117-123
31. **Martin Y. C. (1992)** 3D database searching in drug design. *J. Med. Chem.* 35, 2145-2154
32. **Artymiuk P. J. Bath P. A., Grindley H. M., et al. (1992)** Similarity searching in database of 3-dimensional molecules and macromolecules. *J. Chem. Inf. Comput. Sci.* 32, 617-630.
33. **Goodford P. J. (1985)** A computational procedure for determining energetically favourable binding sites on biologically important macromolecules. *J. Med. Chem.* 28, 849-857
34. **Miranker A. and Karplus M. (1991)** Functionality maps of binding sites- a multiple copy simultaneous search method. *Protein Struct. Funct. Genet.* 11, 29-34



35. **Bohacek R. S. and McMartin C. (1994)** Multiple highly diverse structures complementary to enzyme binding sites-results of extensive application of de novo design method incorporating combinatorial growth. *J. Am. Chem.Soc.* 116, 5560-5571
36. **Bodian D. L., Yamasaki R. B., Buswell R. L., Stearns J. F., White J. M. and Kuntz I. D. (1993)** Inhibition of the fusion-inducing conformational change of influenza-hemagglutinin by 1,4-benzoquinones and hydroquinones, *Biochemistry* 32, 2967-2978
37. **Grootenhuis P. D. J., Roe D. C., Kollman P. A. and Kuntz I. D. (1994)** Finding potential DNA binding compounds by using molecular shape. *J. Comput.. Aided Mol. Design* 8 731-750
38. **Kerwin S. M., Kuntz I. D. and Kenyon G. L. (1991)** The design of a DNA binding compound using an automated procedure for screening potential ligand. *Med. Chem. Res* 1, 361-368
39. **Qi Chen, Kuntz I. D. and Shafer R. H. (1996)** Spectroscopic recognition of guanine dimeric hairpin quadruplexes by a carbocyanine dye, *Proc. Natl. Acad. Sci. USA* 93 2635-2639
40. **DasJarlais R. L., Seibal G. L., Kuntz I. D., Ortiz P. R. de Montellano, Furth S. P., Alvarez J. C., DeCamp D. L., Babe L. M. and Craik C. S. (1990)** Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease, *Proc. Natl. Acad. Sci. USA* 87 6644-6648
41. **Li Z., Chen X., Davidson E., Zwang O., Mendis C., Ring C. S., Roush W. R., Fegley G., Li R., Rosenthal P. J., Lee G. K., Kenyon G. L., Kubntz I. D. and Cohen F. E. (1994)** Anti-malarial drug development using models of enzyme structure, *Chem. Biol.* 1,31-37

42. Ring C.S., Sun E., McKerrow J. H., Lee G. K., Rosenthal P. J., Kuntz I. D. and Cohen F. E. (1993) Structure-based inhibitor design using models for the development of antiparasitic agents. *Proc. Natl. Acad. Sci, USA* 90, 3583-3587
43. Rutenber E., Fauman E. B., Keenan R. J., Fong S., Furth P. S., Ortiz de Montellano P. R., Meng E., Kuntz I. D., DeCamp D. L., Salto R., Rose J. R., Craik C. and Stroud R. M.. (1993) Structure of a non-peptide inhibitor complexed with HIV-1 protease. *J. Biol. Chem.* 268, 15343-15346
44. Shoichet B. K., Stroud R. M., Santi D. V., Kuntz I. D., and Perry K. M. (1993) Structure-based discovery of inhibitors of thymidylate synthase, *Science* 259, 1445-1450
45. Karplus P. A. and Faerman C. (1994) Ordered water in macromolecular structure. *Curr. Opin. Struct. Biol.* 4, 770-776
46. Reich S. H., Fuhry M. A. M., Nguyen D., Pino M. J., Welsh K. M., Webber S., Janson C. A., Jordan S. R., Matthews D. A., Smith W. W., Bartlett C. A., Booth C. L. J., Herrmann S. M., Howland E. F., Morse C. A., Ward R. W. and White J. (1992) Design and synthesis of novel 6,7-imidazotetra-hydroquinoline inhibitors of thymidylate synthase using iterative protein crystal structure analysis, *J. Med. Chem*, 35, 847-858.
47. Halgren T. A. (1996) Merck molecular force field. I Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem*, 17, 490-519
48. Halgren T. A. (1996) Merck molecular force field. II MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem*, 17, 520-552.
49. Halgren T. A. (1996) Merck molecular force field. III Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem*, 17, 553-86.

50. **Halgren T. A. and Nachbar R. B. (1996)** Merck molecular force field. IV conformational energies and geometries for MMFF94. *J. Comput. Chem*, 17, 587-615.
51. **Halgren T. A. (1996)** Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem*, 17, 614-41
52. **Gillet V. J., A. P. Johnson, P. Mata, S. Sike and P. Williams (1993)** SPROUT; A program for structure generation. *J. Comput. Aided Mol. Design*, 7, 127-153.
53. **Gillet V. J., Newell W., Mata P., Myatt G., Sike S., Zsoldos Z. and Johnson A. P. (1994)** SPROUT-recent developments in the de novo design of molecules. *J. Chem. Inf. Comput. Sci*, 34, 207-217.
54. **Bunin B. A. and Ellman J. A. (1993)** A general and expedient method for the solid-phase synthesis of 1,4-benzodiazopenes. *J. Am. Chem. Soc* 114, 10997-10998.
55. **Martin E. J., Blaney J. M., Siani M. A., Spellmeyer D. C., Wong A. K. and Moos W. H. (1995)** Experimental design of combinatorial libraries for drug discovery. *J. Med. Chem.* 38, 1431-1436.
56. **Ridky T. and Leis J. (1995)** Development of drug resistance to HIV-1 protease inhibitors. *J. Biol. Chem.* 270, 29621-29623
57. **Hermes J. D., Blacklow S. C. and Knowles J. R. (1990)** Searching sequence space by definably random mutagenesis; Improving the catalytic potential of an enzyme. *Proc. Natl. Acad. Sci. USA* 87, 696-700.
58. **Kuroda M. J., el-Farrash M. A., Choudhury S. and Harada S. (1995)** Impaired infectivity of HIV-1 after a single point mutation in the POL gene to escape the effect of a protease inhibitor in vitro. *Virology* 210, 212-216.

59. **Partaledis J. A., Yamaguchi K., Tisdale M., et al. (1995)** In vitro selection and characterization of human immunodeficiency virus type 1 (HIV-1) isolates with reduced sensitivity to hydroxyethylamino sulfonamide inhibitors of HIV-1 aspartyl protease. *J. Virol.* 69, 5228-5235
60. **Reddy M. R., Varney M. D., Kalish V., Viswanadhan V. N. and Appelt K. (1994)** Calculation of relative differences in the binding free energies of HIV-1 protease inhibitors: A thermodynamic cycle perturbation approach. *J. Med. Chem.* 37, 1145-52.
61. **Zuckermann R. N., Martin E. J., Spellmeyer D. C., et al. (1994)** Discovery of nanomolar ligands for 7-transmembrane G-protein-coupled receptors from a diverse N- (substituted) glycine peptide libraries. *J. Med. Chem.* 37, 2678-85.
62. **Sheridan R. P. and Kearsley S. K. (1995)** Using a genetic algorithm to suggest combinatorial libraries, *J. Chem. Inf. Comput. Sci.* 35, 310
63. **Weber L., Wallbaum S., Broger C. and Gubernator K. (1995)** Optimization of the biological activity of combinatorial compound libraries by a genetic algorithm. *Angew. Chem. Int. Ed. Engl.* 34, 2280-2282.
64. **Singh J., Ator M. A., Jaeger E. P., Allen M. P., Whipple D. A., Soloweij J. E., Chowdhary S. and Treasurywala A. M. (1996)** Application of genetic algorithms to combinatorial synthesis: a computational approach to lead identification and lead optimization. *J. Am. Chem. Soc.* 118, 1669
65. **Barton, G. H. R. (1950)** The conformation of the steroid nucleus. *Experientia* 6, 316-320.
66. **Blaney, J. M. and Dixon J. S. (1993)** A good ligand is hard to find: Automated docking methods. *Perspective in drug discovery and design* 1, 301-319.

67. **Goodsell D. S. and Olson A. J. (1990)** Automatic docking of substrates to proteins by simulated annealing. *Protein: Structure, Function and Genetics* 8, 195-202
68. **Hansch, C. (1969)** A quantitative approach to biochemical Structure- Activity Relationships. *Accounts of Chemical Research* 2, 232-239.
69. **Andrea, T. A. and Halayeh H. (1991)** Applications of neural networks in quantitative structure-activity relationships of dihydrofolate-reductase inhibitors. *Journal of Medicinal Chemistry* 34, 2824-2836.
70. **Manallack, D. T., Ellis D. D. and Livingstone D. J. (1994)** Analysis of linear and nonlinear QSAR data using neural networks. *Journal of Computer-Aided Molecular Design* 37, 3758-3767.
71. **Rogers, D. and Hopfinger A. J. (1994)** Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *Journal of Chemical Information and Computer Science* 34, 845-866.
72. **Chatfield, C. and A. J. Collins (1980)** *Introduction to Multivariate Analysis*. London Chapman and Hall.
73. **H. Martens and T. Naes, Multivariate Calibration, chapter 3, John Wiley and Sons, 1989.**
74. **Wold S., Technometrics, 1978, 20(4), 39.**
75. **Cramer, R. D III, D. E. Patterson and J. D. Bunce (1988)** Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Building of Steroids to Carrier Proteins. *The Journal of the American Chemical Society* 110, 5959-5967.

76. **Badertscher M., Welti M., Portmann P., and Pretsch E., (1986)** Calculation of interaction energies in host-guest systems. *Curr. Chem. (Biomimetic Bioorg. Chem. 3)* 136, 17-80.
77. **Stanton R. V., Dixon S. L., Merz K. M., Jr (1996)** Free energy perturbation calculations within quantum mechanical methodologies, *ACC Symp. Ser. (Chemical Applications of Density-Functional Theory)* 629, 142-153.
78. **Eisenberg D. and McLachlan A. D., (1986)** Solvation energy in protein folding and binding. *Nature* 319, 199-203.
79. **Dolata D. P., Leach A. R., and Prout K., (1987)** WIZARD: AI in conformational analysis, *Journal of Computer-Aided Molecular Design* 1, 73-85.
80. **Leach A. R. and Prout K., (1990)** Automated conformational analysis-directed conformational search using the A\* algorithm. *J. Comput. Chem. II*, 1193-1205.
81. **Rusinko A., Sheridan R. P., Nilakatan R., Haraki K. S., Bauman N. and Venkataghavan R. (1989)** Using CONCORD to construct a large database of 3-dimensional coordinates from connection tables. *J. Chem. Inf. Comput. Sci.* 29, 251-255.
82. **Rubicon**, Daylight Chemical Information System, Mission Viejo CA.
83. **Allinger N. L., Yuh Y. H. and Lii J. J. (1989)** Molecular mechanics in the MM3 force field for hydrocarbons. *J. Am. Chem. Soc.* 111, 8551-8566.
84. **Allen F. H., Bellard S., Brice M. D., Cartwright B. A., Doubleday A., Higgs H., Hummelink T., Hummelink-Peters B. G., Kennard O.,**

- Motherwell W. D. S., Rodgers J. R. and Watson D. G.. (1979)** The Cambridge Crystallographic Data Centre: Computer-based search retrieval, analysis, and display of information. *Acta. Crystallogr. Sect. B* 35, 2331-2339.
85. **Milne G. W. A., Nicklaus M. C., Driscoll J. S. and Wang S.. (1994)** National Cancer Institute Drug Information System 3D Database. *J. Chem. Inf. Comput. Sci* 34, 1219-1224.
86. **Jones G., Willett P. and Glen R. C. (1995)** Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* 245, 43-53.
87. **Gschwend D. (1995)** Molecular docking, PhD. Dissertation, University of California, San Francisco.
88. **Meng E. C., Gschwend D. A., Blaney J. M. and Kuntz I. D. (1993)** Orientational sampling and rigid-body minimization in molecular docking. *Proteins Struct. Funct. Genet* 17, 266-278.
89. **Bairoch, A. and B. Boeckmann. (1992)** The SWISS-PORT protein sequence data bank. *Nucleic Acids Res*, 20, 2019-2022.
90. **Bernstein, F. C., Koetzle T. F., Williams G. J. B., Meyer E. F., Brice M. D., Rodgers J. R., Kennard O., Shimanovich T. and Tasumi M.. (1997)** The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112, 535-542.
91. **Chou, P. Y. and Fasman G. D.. (1974)** Prediction of protein conformation. *Biochemistry* 13, 222-245.

92. **Garnier, J., D. J. Osguthorpe and B. Robson. (1978)** Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Bio.* 120, 97-120.
93. **Benner, S. A. (1992)** Prediction de novo of the folded structure of proteins. *Curr. Opin. Struct. Biol.* 2, 402-412.
94. **Martin, Y. C (1992)** 3D database searching in drug design, *J. Med. Chem*, 35, 2145.
95. **Martin, Y. C (1989)** ALADDIN: an integrated tool for computer-assisted molecular design and pharmacophore recognition from geometric, steric, and substructure searching of three-dimensional molecular structures. *J. Comput-Aided Mol. Design* 3. 225.
96. **Bartlett, P. A., Shea, G. T. and Telfer, S. J. (1989)** In Roberts, S. M. (Ed): Molecular Recognition: Chemical and Biological Problems, The Royal Society of Chemistry, London, 182-196.
97. **Desjarlais, R. L., Sheridan R. P., Seibel G. L., Dixon, J. S. Kuntz, I. D., and Venkataghavan, R. (1988)** Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure *J. Med. Chem*, 31, 722.
98. **Martin, Y. C. (1990)** Computer design of potentially bioactive molecules by geometric searching with ALADDIN. *Tetrahedron Comput. Method* 3, 15-25.
99. **Lewis, R. A. and Dean, P. M. (1989)** Automated site-directed drug design: the concept of spacer skeletons for primary structure generation. *Proc. R. Soc. London Ser. B*, 236, 125.



100. **Lewis, R. A. (1990)** Automated site-directed drug design: approaches to the formation of 3D molecular graphs. *J. Comput.-Aided Mol. Design* 4, 205.
101. **Lewis, R. A. (1992)** Automated site-directed drug design using molecular lattices. *J. Mol. Graphics* 10, 66.
102. **Nishibata, Y. and Itai, A. (1991)** Automatic creation of drug candidate structures based on receptor structure. Starting point for artificial lead generation. *Tetrahedron* 47, 8985.
103. **Moon, J. B. and Howe, W. J. (1991)** Computer design of bioactive molecules: a method for receptor-based de novo ligand design *Protein Struct. Funct. Genet.* 11, 314.
104. **Bohm, H. J. (1992)** The computer program LUDI: a new method for the de novo design of enzyme inhibitors *J. Comput.-Aided Mol. Design* 6, 61.
105. **Lewis, R. A., Reo, D. C., Huang, C. Ferrin, T. E. Langridge, R. and Kuntz, I. D. (1992)** Automated site-directed drug design using molecular lattices *J. Mol. Graphics* 10, 66.
106. **Verlinde, C. L. M. J., Rudenko, G. and Hol, W. G. J. (1992)** In search of new lead compounds for trypanosomiasis drug design: a protein structure-based linked-fragment approach *J. Comput.-Aided Mol. Design* 6, 131.
107. **Wipke, W. T., Pitman, M. C., Koehler, R. T., Kislin, B.S. and Anderson, G. D. (1991)** Abstracts of Papers, 202<sup>nd</sup> National Meeting of the American Chemical Society, New York, NY, Fall 1991, American Chemical Society, Washington, DC.

108. **Doweyko, A. M. (1988)** The hypothetical active site lattice. An approach to modelling active sites from data on inhibitor molecules *J. Med. Chem.* 31, 1396.
109. **Cohen, N. C. Blaney, J. M., Humblet, C., Gund, P. and Barry, D. C. (1990)** Molecular modeling software and methods for medicinal chemistry *J. Med. Chem.* 33, 883.
110. **Nilsson, N. J. (1982)** *Principle of Artificial Intelligence, Springer-Verlag, Berlin.*
111. **Danziger, D. J. and Dean, P. M. (1989)** Automated site-directed drug design: a general algorithm for knowledge acquisition about hydrogen-bonding regions at protein surfaces *Proc. R. Soc. London Ser. B.* 236, 101.
112. **Tomioka, N., Itai, A. and Iitaka, Y. (1987)** A method for fast energy estimation and visualization of protein-ligand interaction *J. Comput.-Aided Mol. Design* 1, 197.
113. **C. K. Mathews and K. E. van Holde (1990)** *Biochemistry. Benjamin/Cummings Publishing Company. Inc. Redwood City, CA.*
114. **L. F. Liu and J. C. Wang (1978)** DNA-DNA Gyrase complex: the wrapping of the DNA duplex outside the enzyme. *Cell*, 15, 979-984.
115. **L. F. Liu et al. (1994)** *Advance in Pharmacology Volume 29A, DNA Topoisomerases: Biochemistry and Molecular Biology. Academic Press Ltd, London.*

116. **M. K. Campell.**(1995) *Biochemistry, 2<sup>nd</sup> Ed. Saunders College Publishing, Orlando.*
117. **D.B.Wigley,G. J. Davies, E .J. Dodson, A. Maxwell and G. Dodson** (1991). Crystal structure of an N-terminal fragment of the DNA gyrase B protein. *Nature* 351,624-629.
118. <http://cmgm.stanford.edu/biochem201/Slides/DNA%20Topoisomerase/114%20DNA%20Gyrase%20Struct.JPG>
119. **Menzel, R. and Gellert, M. (1994).** The biochemistry and biology of DNA gyrase. *Advances in Pharmacology Volume 29A, Lui, L. F, Academic Press Limited, 24-28 Oval Road, London, 39-61.*
120. **Lodish, H. Berk, Zipursky A., Matsudaira S.L., Baltimore P., Darnell, J. (2000).** *Molecular Cell Biology, Tenney S. W.H. Freeman and Company, 41 Madison Avenue, New York, 468-72.*
121. **Kuo, M. S, Yurek, D.A, Chirby, D. G, Cialdella, J. I and Marshall, V. P. (1991).** Microbial o-carbamoylation of novobiocin. *J. Antibiot. 44, 1096-1100*
122. **Gellert, M (1981).** DNA topoisomerases. *Annu. Rev. Biochem.* 50;879-910
123. **Gitschi E., Jenny C. J., Reindl P., Ridelin F. (1996).** Total synthesis of cyclothialidine. *Helvetica Chimica Acta. 79; 2219-2234*

124. **Drlica, K. and R. J. Franco (1988).** Inhibitors of DNA topoisomerases. *Biochemistry* 27; 2253-2259
125. **Nakada N., Gmunder H., Hirata T. and Arisawa M. (1995)** Characterisation of the binding site for cyclothialidine on the B subunit of DNA gyrase. *J. Bio. Chem.* 270; 14286-14291
126. **Gotschi E., C. Jenny, P. Reindl and F. Ricklin (1996).** Total synthesis of Cyclothialidine. *Helvetica Chimica Acta, Vol 79, 2219-2233.*
127. **Rudolph J., Theis H., Hanke R., Endermann R., Johannsen L. and Geschke F. (2001).** Seco-Cyclothialidine: new concise synthesis, inhibitor activity toward bacterial and human DNA topoisomerases, and antibacterial properties. *J. Med. Chem.* 44, 619-626.
128. **Sheehan J. C, Hess G. P (1955)** A new method of forming peptide bonds. *J. Amer. Chem. Soc.* 77, 1067-1068.
129. **Schussler H, Zahn H (1962)** Peptide XXIII. The mode of reaction of carbobenzyloxyamine acid with dicyclohexylcarbodiimide. *Chem. Ber* 95, 1076-1080.
130. **Nefkens G. H. L, Tesser G. I (1961)** A novel activated ester in peptide synthesis. *J. Amer. Chem. Soc* 83, 1263
131. **Anderson G. W, Zimmermann J. E, Callahan F. M (1963)** N-hydroxysuccinimide esters in peptide synthesis. *J. Amer. Chem. Soc* 85, 3039

132. **Beaumont, S. M., Hanford, B.O., Jones, J. H., Young, G. T. (1965)** Amino-acids and peptide. Part XXIV. The use of esters of 1-hydroxypiperidine and other N,N-dialkylhydroxylamines in peptide synthesis and as selective acylating agents. *J. Chem. Soc.* 6814-6827.
133. **Konig, W, Geiger, R (1970)** Racemization in peptide syntheses *Chem. Ber.* 103, 788, 2024
134. **Konig, W, Geiger, R (1973)** Synthesis of peptides with the properties of human proinsulin C peptides (hC peptide). I. Scheme for the synthesis and preparation of the sequence 28-31 of human proinsulin C peptide *Chem Ber.* 106, 188.
135. **Savrda, J. (1977)** An unusual side reaction of 1-succinimidyl esters during peptide synthesis *J. Org. Chem.* 42, 3199
136. **Mitsunobu, O. and Eguchi, M. (1971)** Preparation of carboxylic esters and phosphoric esters by the activation of alcohols, *Bull. Chem.Soc. Jpn*, 44, 3427-3430.
137. **Mitsunobu, O. (1981)** The use of diethyl azodicarboxylate and triphenylphosphine in synthesis and transformation of natural products, *Synthesis*, 1-28
138. **Waddell, S. T. and Blizzard T. A. (1993)** Chimeric azalides with simplified western portion. *Tetrahedron Letters*, 34, 5385-5388.
139. **Valerie, G, Peter Johnson A., Mata P., Sike S., and Williams P. (1993)** SPROUT: A program for structure generation. *Journal of Computer-Aided Molecular Design*, 7, 127-153
140. **Lewis, R.J., Singh, O.M.P., Smith, C.V., Skarzynski, T., Maxwell, A., Wonacott, A.J., and Wigley, D.B. (1996)**. The nature of inhibition of DNA

gyrase by the coumarins and the cyclothialidines revealed by x-ray crystallography. *The EMBO journal*, 15, 1412-20

141. **Boehm, H.J., Boehringer, M., Bur, D., Gmuender, H., Huber, W., Klaus, W., Kostrewa, D., Kuehne, H., Luebbers, T., Meunier-Keller, N., and Mueller, F. (2000).** Novel Inhibitors of DNA gyrase: 3D structure based biased needle screening, Hit validation by biophysical methods, and 3D guided optimisation. A promising alternative to random screening. *Journal of Medicinal Chemistry* 43, 2664-74.
142. **Lam, P.Y. S., Jadhav, P. K., Eyermann, C. E., Hodge, C. N., Ru, Y., Bacheler, L.T., Meek, J. L., Otto, M. J., Rayner, M. M., Wong, Y. N., Chang, C-H, Weber, P. C., Jackson, D. A., Sharpe, T. R. and Erichson-Viitanen, S. (1994)** Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitor. *Science* 263:380-384.
143. **Leach Andrew. (2001)** Molecular Modelling: Principles and Applications. 684.