

Target-Based development of novel lysine histone demethylase one (LSD1) inhibitors

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat)

vorgelegt der

Naturwissenschaftlichen Fakultät I

Biowissenschaften

der Martin-Luther-Universität Halle-Wittenberg

von

Herr MSc Luca Carlino

geb. 02 July 1984 in Caserta, Italien

Gutachter/in:

1. Prof. Dr. Wolfgang Sippl, Halle (Saale) -Deutschland-
2. Prof. Gabriele Costantino, Parma -Italien-
3. Prof. Dr. Gerhard Ecker, Vienna, -Austria-

Verteidigungsdatum: 12/03/2014

Acknowledgments

The present work was carried out at the Institute of Pharmaceutical Chemistry at Martin-Luther-University Halle-Wittenberg from 2009 to 2013.

The first person that I would like to thank is Prof. Wolfgang Sippl, for giving me the opportunity to work in the extraordinary field of computational modelling, for the support and for the very interesting scientific discussions that I had with him. Greatly acknowledged are also Prof. Manfred Jung, Dr. Martin Leo Smith and Dr. Alexander-Thomas Hauser from the University of Freiburg for all the biological tests and the synthesis carried out. Without their job most of the models here presented would not be reported.

Again a special thank go to the people of Prof. Sippl's group for the numerous scientific discussions that we had. I should reserve a special thank to Ralf Heinke for introducing me in the Halle community and for his help with all the bureaucratic documents, to Dr. Dina Robaa for her friendship and for helping me during the revision of the thesis, to Martin Pippel not only for his thesis revision but also for his true friendship demonstrated in several circumstances and his constant help concerning computer science and not only. Regarding computer science I should also thank Micheal Scharfe and German Erenkamp. Finally I want particularly to thank Suhaib Shekfeh, not only for the scientific discussions that I had with him, but also for helping me out of the University environment, for introducing me to the PhD community in Halle/Saale (especially to Kachaber Lominadze) and to being my very first friend in Germany. From the University of Halle I should also thank Prof. Reinhard Neubert and Dr. Matthias Schimdt for giving me the chance to play football regularly and for the neverending discussions on Italian football "way to play".

I want to thank also, some of the people I had the pleasure to meet in Halle and became very good friends like Sherif and Salem, for helping me several times in several ways during these years and for showing me the "party face" of Halle. I also would like to thank Francesca Petricca (my first flatmate) for the very interesting umanistic discussions, for encouraging me and for the strong friendship born in Halle/Saale.

A special thank go to Prof. Gabriele Costantino from the University of Parma for showing me the opportunities to start a PhD experience, for hosting me three months at the University of Parma and for very interesting scientific discussions.

Last but not least I would like to thank my parents and my brother for their personal and financial support, as always, and Aura Grassi for her constant love and her encouragements that helped me several time when the right way seemed lost.

Table of Contents

INTRODUCTION	1	1
1.1 A brief introduction to epigenetics		1
1.2 From DNA to chromosome: the nucleosome complex		2
1.3 Histone modifications		3
1.4 Histone lysine demethylases		4
1.4.1 Histone lysine demethylase 1 (LSD1)		4
1.4.2 Jumonji C (JmjC) domain-containing histone demethylases		10
1.5 LSD1 inhibitors		11
1.5.1 LSD1 inhibitors derived from MAOs inhibitors		12
1.5.2 LSD1 inhibitors derived from PAO inhibitors		14
1.5.3 LSD1 reversible inhibitors		15
1.6 Aim of the work		16
MATERIALS AND METHODS	2	17
2.1 Quantum mechanics and empirical force field methods		17
2.1.1 Energy minimisation and molecular dynamic simulation		18
2.2 Molecular docking		20
2.2.1 Scoring functions		21
2.2.2 Docking programs		22
2.3 Virtual screening		23
2.4 Binding free energy calculation methods		24
2.4.1 Molecular mechanics Poisson-Boltzmann surface area (MM-PBSA)		25
2.4.2 Quantum-mechanics, molecular mechanics Poisson-Boltzmann surface area (QMMM-PBSA)		27
2.6 In vitro assays for LSD1 demethylases activity		28
2.6.1 Peroxidase assay		28
2.6.2 Formaldehyde-dehydrogenase-dependent (FDH) assay		29
2.6.3 Antibody based assay		30
2.6.4 Mass-spectrometry (MS) assay		32
RESULTS AND DISCUSSION	3	33
3.1 Analysis of LSD1 crystal structures		33
3.1.1 LSD1 crystal structures		33
3.1.2 Structural analysis		34
3.1.3 LSD1 similarity search		37

Table of Contents

3.1.4 Molecular dynamic simulations.....	38
3.1.4.1 MD simulation of LSD1 complexed with the N-terminal histone tail and unmethylated Lys4'	40
3.1.4.2 MD simulation of LSD1 complexed with the N-terminal histone tail and mono-methylated Lys4'	43
3.1.4.3 MD simulation of LSD1 complexed with the N-terminal histone tail and di-methylated Lys4'	44
3.1.4.4 MD simulation of LSD1 complexed with the N-terminal histone tail and tri-methylated Lys4'	45
3.1.5 Discussion.....	46
3.2 Propargylamine derivatives.....	47
3.2.1 Docking studies.....	49
3.2.2 Virtual screening.....	50
3.2.3 New propargylamine derivatives.....	52
3.2.4 Discussion.....	55
3.3 Basic polyamino and guanidino derivatives	57
3.3.1 Docking studies.....	57
3.3.2 Molecular Dynamics simulations	63
3.3.3 Multi-Step target-based virtual screening	65
3.3.4 Discussion.....	68
3.4 4H-chromen-4-one derivatives	69
3.4.1 Structure-Based approach	72
3.4.1.1 Docking studies	72
3.4.1.2 Binding free energy calculation studies	73
3.4.1.3 QM/MM-GBSA model and new putative in silico generated LSD1 inhibitors activity prediction.	79
3.4.2 Structure activity relationships (SAR) of chromane derivatives	82
3.4.3 Discussion.....	83
3.5 Coumarine derivatives	85
3.5.1 Preliminary similarity search	85
3.5.2 Docking studies.....	86
3.5.3 Similarity search.....	89
3.5.4 Discussion.....	94
CONCLUSION AND OUTLOOK 4	95
SUMMARY 5	99
BIBLIOGRAPHY.....	101
APPENDIX A	113
Force Field Terms	113
Bond Stretching term	113

Table of Contents

Angle Bending term.....	113
Torsion angle term	114
Out-of-plane bending term.....	114
Van der Waals term.....	114
Electrostatic term.....	115
Computational Details	115
Docking studies	115
Molecular dynamics	116
MD of LSD1 with 15 residues of the N-terminal histone tail.....	116
MD of LSD1 with polyamine inhibitors.....	118
MM-PB(GB)SA and QM/MM-GBSA calculations.....	118
Assays details.....	120
In vitro peroxidase assays for propargylamine derivatives	120
In vitro DELFIA® assays.....	121
APPENDIX B.....	123
Table 1: Propargylamine derivatives dataset	123
Table 2: 4<i>H</i>-chromen-4-one derivatives dataset using Namoline as query structure	126
Table 3: 4<i>H</i>-chromen-4-one false positive derivatives dataset using Shaolin as query structure.....	127
Table 4: Coumarine derivatives from preliminary similarity search using Chembridge database	129
List of abbreviations.....	133
Amino acids letter codes.....	137

Table of Contents

INTRODUCTION

1.1 A brief introduction to epigenetics

The term “epigenetics” has Greek origin and literally means above (epi-) the genetics. During the last years several definitions of epigenetics have been proposed. In 2003 R. Jaenisch and A. Bird defined epigenetics as the “heritable changes in gene expression or phenotype that are stable between cell divisions, and sometimes between generations, but do not involve changes in the underlying DNA sequence of the organism”.^{1,2}

Every eukaryotic cell has a specific epigenetic state that changes during the cell differentiation and the development of an organism.² Epigenetic changes are also involved in cellular reprogramming and in response to the environment.² For these reasons, all epigenetic mechanisms may have an important role in diseases related to lifestyle, diet, surrounding environment and early life experiences.^{2,3} Hence, epigenetic mechanisms found therapeutic importance in multiple diseases like cancer, metabolic diseases, inflammation, neuropsychiatric disorders and in regenerative medicine.^{2,4,5,6} Manipulations of the epigenetic mechanisms associated with cell diseases might modify the cell phenotype and restore the original cell function without affecting the DNA sequence.

Up till today epigenetic modifications have been classified into three main mechanisms: DNA methylation or hydroxymethylation; methyl groups are added or removed from the deoxyribose chain by a group of protein called DNA methyltransferases (DNMTs),⁷ micro RNAs (mRNAs) modifications,⁸ and N-terminal histone tails post-translational modifications (also named histone modifications).

Nowadays there are several examples of DNA methylation inhibitors and histone deacetylase (HDAC) inhibitors that are approved for clinical use in cancer therapy,^{2,9,10} which confirms the importance of the epigenetic regulating mechanisms as new targets for cancer diagnosis and therapy.

1.2 From DNA to chromosome: the nucleosome complex

In eukaryotic cells genetic information is organized in a basic structure called DNA. The DNA molecule is a polymer formed by 3×10^9 nucleotides and a backbone made of sugar and phosphate groups joined by ester bonds. This structure forms approximately 25000 genes (the basic units for proteins synthesis) and its total length is about two meters. The DNA location is the cell nucleus, and for a two-meter long structure, the folding process is a crucial step to maintain the polymer in such a limited space. DNA organization into the nucleosome structure is the first step of the folding process. The nucleosome has a characteristic structure which consists of four histone proteins: H2A, H2B, H3 and H4 (each one in two copies forming an octamer) and ~147 base pairs of DNA wrapped around them (figure 1).¹¹

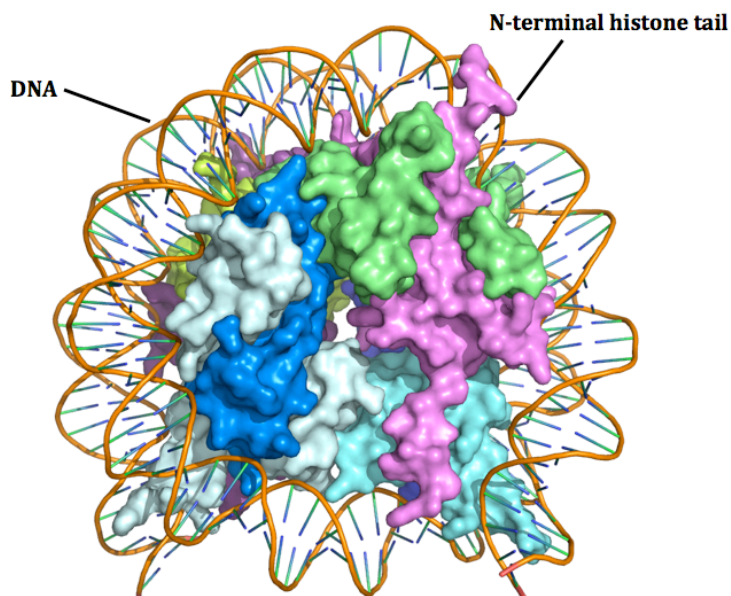


Figure 1: Nucleosome crystal structure (PDB code 3AFA). The histone proteins form the core of the structure. They are represented as protein surface and are coloured differently. The DNA structure is represented as a ribbon.

The nucleosome basic units generate a polymeric structure called chromatin. The organization of chromatin domains form a structure called chromatid, while two chromatids linked by a centromere are organized in a structure known as chromosome. In humans there are 23 chromosomes, carrying all the necessary information.¹²

Chromatin can be classified into two main structures: highly condensed chromatin fibres (heterochromatin) related to transcriptional repression and arrest of cell division; and less compacted chromatin fibres (euchromatin), that usually refer to an active transcriptional state. Chromatin might be considered as a dynamic polymer whose function is to organize the whole

Introduction

genome, determining which genes have to be expressed or repressed, according to the circumstances.¹²

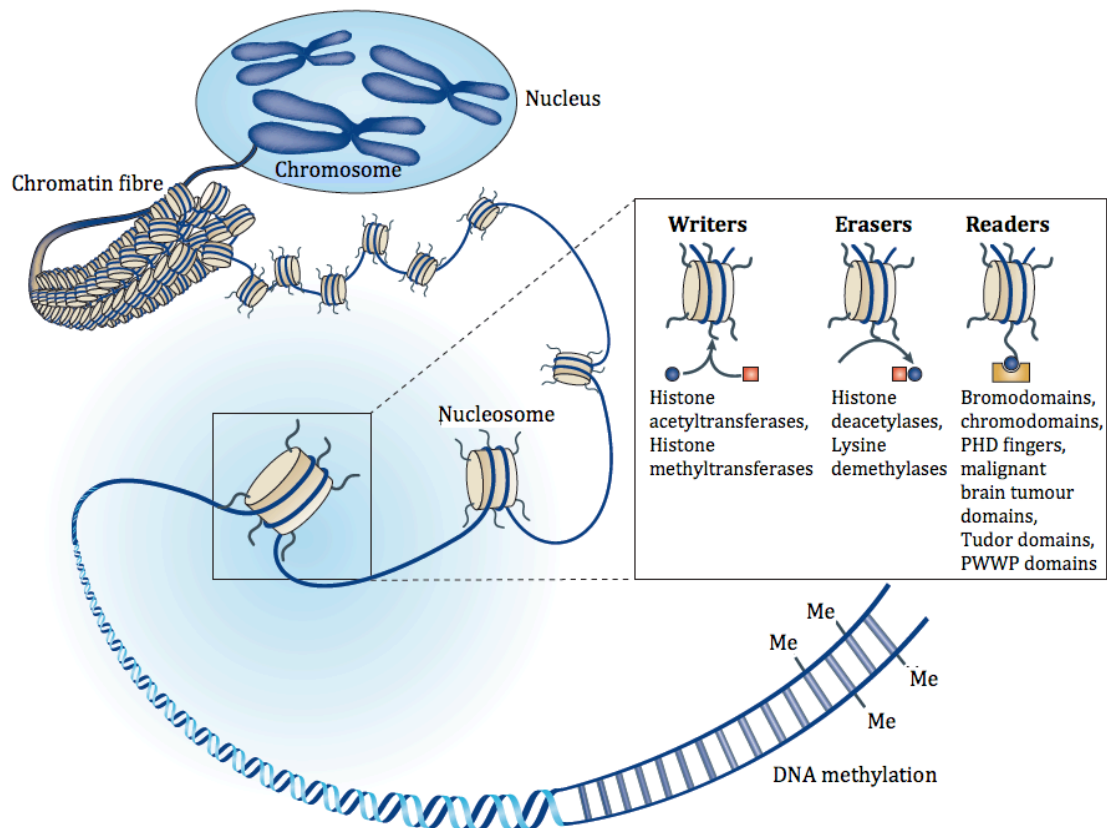


Figure 2: Graphical representation of nucleosome, chromatin fibre and chromosome (adapted from Arrowsmith C.H. et al.²).

1.3 Histone modifications

Histone proteins show covalent chemical flags. These covalent adducts, called post-translational modifications, are present on the amino acids in the N-terminal histone tails or in the histone core, therefore they have a direct role in gene expression/repression. The most important modifications include phosphorylation, acetylation, ubiquitinylation, methylation, sumoylation and ribosylation.¹² Usually active marks include acetylation, arginine methylation, and some lysine methylations (lysine number four on histone number three H3K4, or lysine number 36 on histone number three H3K36). On the other hand, repressive marks include lysine methylation of H3K9, H3K27, and H4K20.¹² All combinations of these modifications that might occur on every histone and/or nucleosome generate a specific code that rules the transcriptional properties of the involved genes.² Histone modifications are mediated by specific proteins.¹² The general classification of the involved proteins is according to their functions: “writers” are proteins that form the code attaching the

covalent flags (for instance histone acetyltransferases and histone methyltransferases); “readers” are proteins that read the histone code without introducing any modification to the covalently bound flags (e.g. bromodomains, plant homeodomains (PHDs) and members of the royal family of methyl-lysine-binding domains); while “erasers” are all enzymes that remove the covalently bound histone marks from the histones (for instance histone deacetylases and histone demethylases) (figure 2).²

1.4 Histone lysine demethylases

The first histone demethylase (HDM) was identified in 2004 by Shi and co-workers.^{13,14} It was found to be active on lysine number 4 in histone H3 and was named lysine-specific-demethylase-1 (LSD1).^{13,14,15} Besides, it was also found that LSD1 shows a high sequence-homology with FAD-dependent amine oxidases (Chapter 3.1).¹³ Two years later another type of HDMs, operating by different demethylation mechanism, was identified and termed Jumonji C (JmjC) domain-containing histone demethylases.^{13,16}

1.4.1 Histone lysine demethylase 1 (LSD1)

Lysine specific demethylase-1 (also called LSD1, KDM1A, BHC110 and AOF2) is a flavin-dependent amine oxidase enzyme where the FAD cofactor is reversibly bound to the active site of the enzyme.¹⁷

The catalytic mechanism of LSD1 involves an oxidation of the substrate, using one molecule of Flavin Adenine Dinucleotide (FAD) as a cofactor, passing through an imine intermediate (figure 3).^{17,14,18}

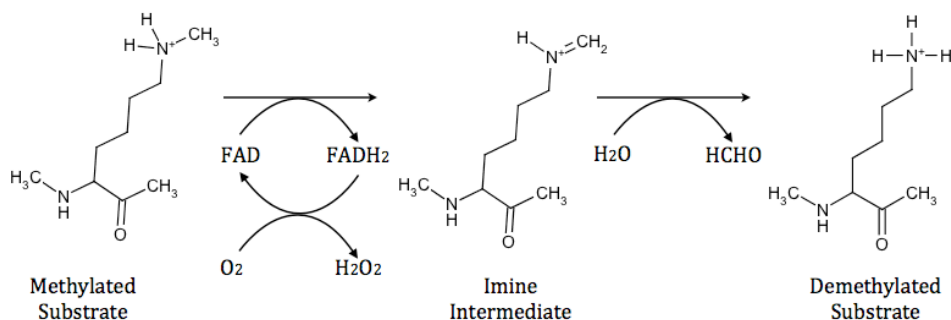


Figure 3: Catalytic mechanism of LSD1.

Introduction

For the flavin-catalysed substrate-dehydrogenation reaction three putative redox reaction mechanisms were hypothesized (figure 4):¹⁹ (i) direct reaction: there is a direct transfer of a hydride anion from the substrate to the flavin ring; (ii) radical mechanism: the reaction is characterized by the formation of a radical pair at C4a and its subsequent collapse; (iii) carbanion mechanism: in this case, an active-side base removes a proton from the substrate and the newly formed carbanion is able to donate two electrons to the flavin ring.¹⁹

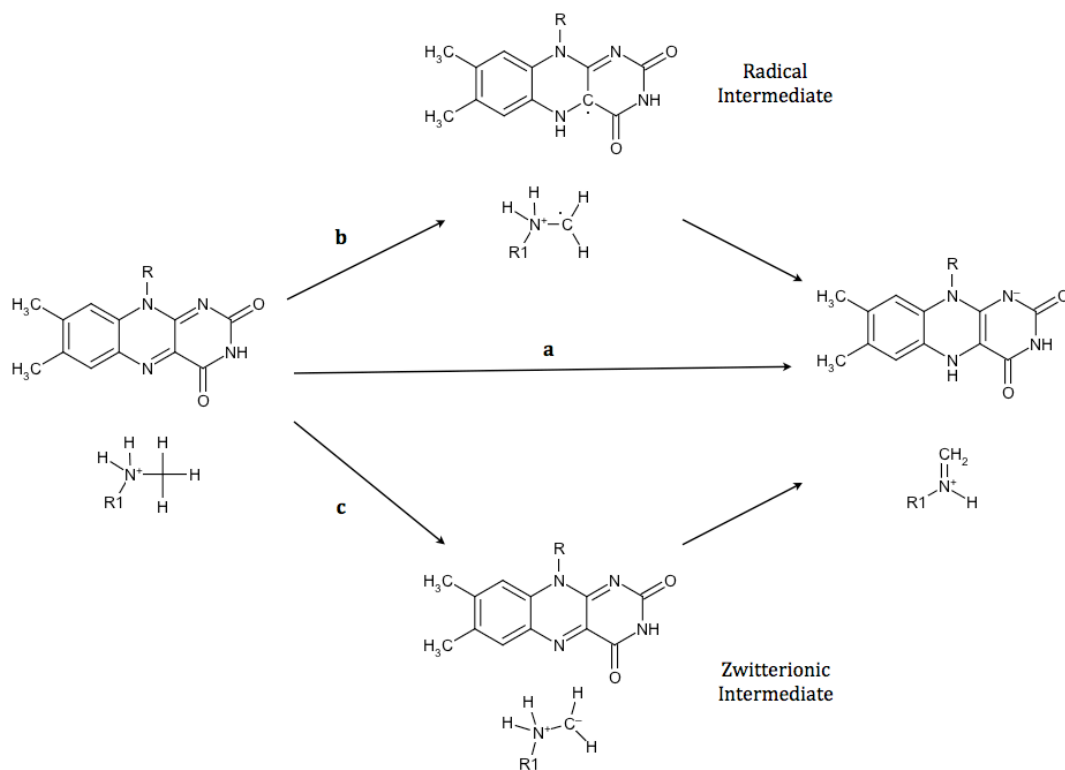


Figure 4: Proposed catalytic mechanism¹⁹ for the flavin-catalysed substrate-dehydrogenation reaction.

During the LSD1 catalytic reaction, a formaldehyde molecule is also produced from one water molecule (figure 3) following the cleavage of the imine intermediate.^{17,14,18} Some suggested that the produced formaldehyde might be recycled as a methyl donor (figure 5) for a class of enzymes called histone methyltransferases (HTM).²⁰

Introduction

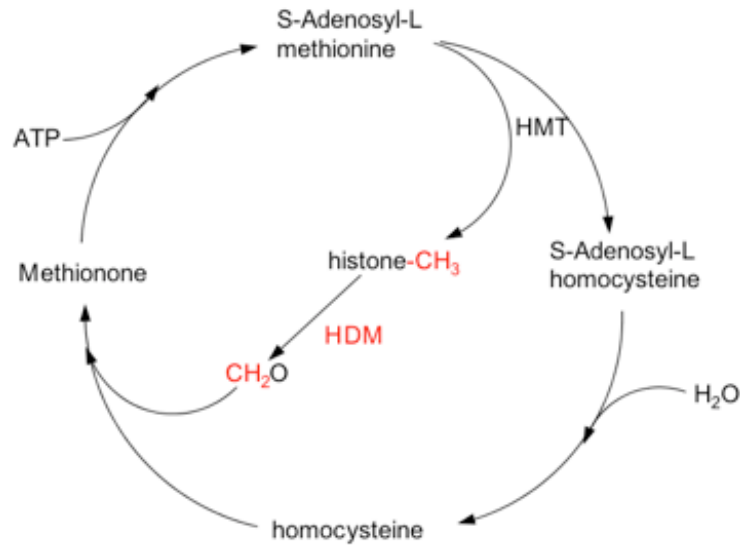


Figure 5: Production of the formaldehyde molecule by LSD1 and recruited by HMT as S-Adenosyl-L-methionine (adapted from Bannister A.J. et al.²¹).

After the LSD1 catalytic reaction is finished, an oxygen molecule is responsible to restore the FAD functionality; the FAD isoalloxazine ring transfers two electrons to the oxygen molecule. This allows the re-oxidation of the cofactor, which now is ready for another catalytic cycle. However, this step generates one molecule of hydrogen peroxide (figure 3) whose role in the cell biology is not completely understood. In fact the generation of hydrogen peroxide in the chromatin environment might favour oxidative damage of DNA and might be harmful.¹⁷ Nevertheless, it has been described that hydrogen peroxide acts as a signalling molecule in a variety of physiological processes including apoptosis, cell cycle progression, cell differentiation, transcriptional regulation, and also in pathological events including neurodegeneration.^{22,23}

LSD1 is a huge protein formed by 852 amino acids and it consists of three separate domains (figure 6). The catalytic domain of LSD1 is the amine oxidase like (AOL) domain that shares sequence homology with other classes of flavin-dependent amine-oxidases.^{18,24} Here the FAD-cofactor is located and the catalytic reaction takes place. Linked to the AOL domain is the so-called SWIRM domain. This is the N-terminal region of LSD1 and in contrast to other SWIRM domains present in nature, in LSD1 it seems not to be directly involved in the substrate binding and it is only responsible for the stabilization of the enzyme.¹⁸

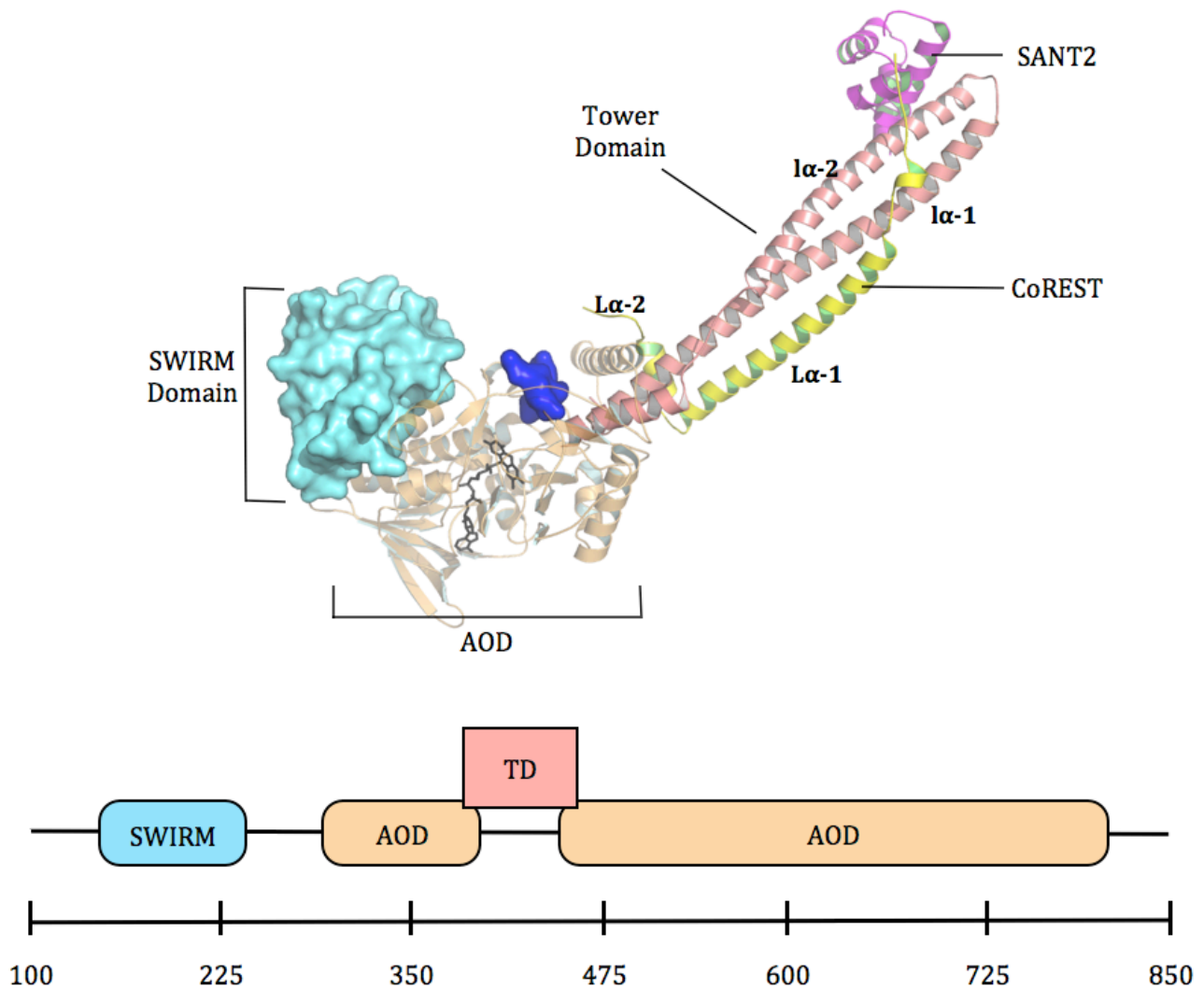


Figure 6: LSD1 structure interacting with CoREST (PDB code 2UXN). The SWIRM domain is represented as cyan protein surface. AOL domain is represented as orange cartoon, while the tower domain is represented as salmon cartoon. FAD is shown as black sticks while the substrate is represented as blue surface. CoREST is represented as yellow cartoon, while SANT2 domain is shown as purple cartoon. In the bottom part there is a schematic representation of the LSD1 domains according to residues enumeration.

The last domain present in the structure is a double α -helical tower domain; a region found to link the allosteric LSD1 modulator like the repressor element-1 silencing transcription factor co-repressor (CoREST).^{25,24,18,26}

Co-REST is a protein member of the class of histone modification complexes, and interacts through engaging the tower domain of LSD1. The interaction is modulated by three main interfaces (figure 6): (i): Between L α 1 helix of Co-REST and a loop of LSD1. In this interface there are mostly hydrophobic interactions. (ii): Among L α 2 helix of Co-REST and the two α -helices of LSD1 (L α 1 and L α 2). In this interface there are not only hydrophobic interactions but also ionic interactions. (iii): Between SANT2 and one L α 2-helix of LSD1. This interface seems to be unimportant for the activity of LSD1. The main function of this domain is to interact with DNA, since the linkage to the

Introduction

nucleosome is required for the *in vivo* activity of LSD1, even if there is the hypothesis that CoREST might enhance LSD1 in a different manner.^{24,27}

Up to now, several proteins have been found to be LSD1 substrates. The mono- or dimethylated lysine 4 on the tail of histone number 3 (H3K4me1/2) active mark is one of the well-known targets of LSD1.²⁸ From the shape and the properties of the LSD1 active cavity (Chapter 3.1) and from the catalytic mechanism, which requires a protonated nitrogen atom, it could be well demonstrated that LSD1 can demethylate only mono- and dimethylated lysine residues.^{29,30} Considering the H3K4me1/2 role in the histone code, demethylation of this target is directly linked to repression of the gene expression and for this reason LSD1 is found in co-repressor complexes.^{28,14} As a component of co-repressor complexes, LSD1 might associate with other histone-modifying enzyme such as histone deacetylases and methyltransferases.^{28,29} Interestingly, LSD1 is not acting on nucleosome substrates as long as it is not associated with CoREST, but acts only after its recruitment by the silencer REST, suggesting that the association LSD1-CoREST is mandatory for the *in vivo* activity of LSD1.^{26,31,32} BHC80, a protein of one of the complexes that mediate REST repression, inhibits the nucleosomes demethylation mediated by the LSD1-CoREST complex (figure 7).³² Again, LSD1, is stimulated by the presence of a member family of histone deacetylases (HDAC-1), showing a strong relationship between demethylase and deacetylase activities.²⁹

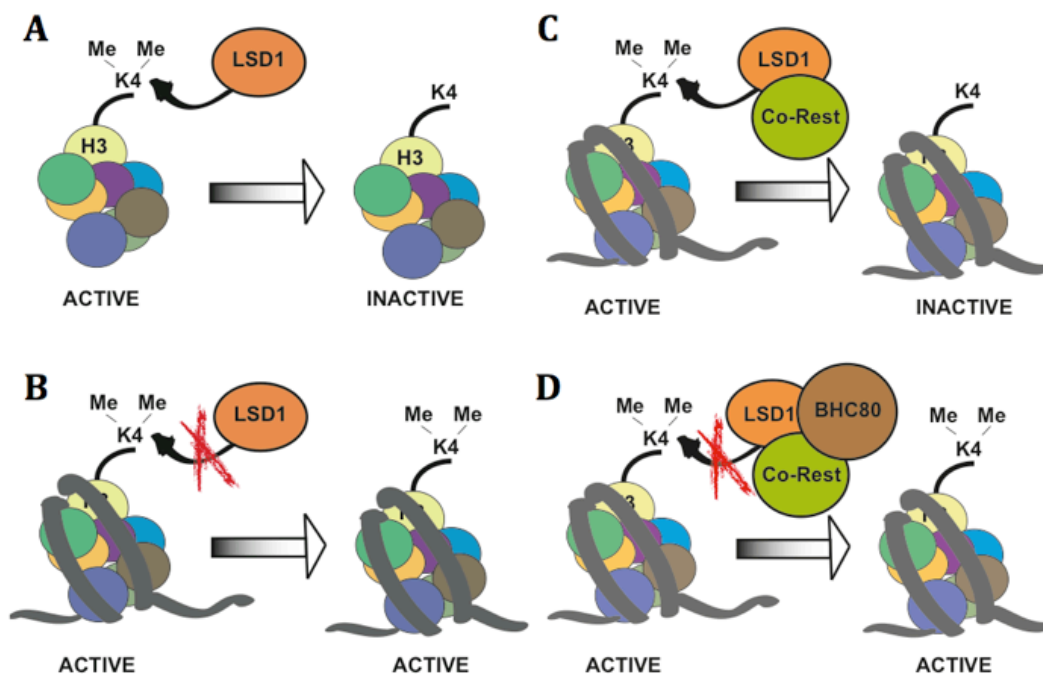


Figure 7: A) LSD1 acts on histone complex in order to demethylate H3K4me1/2 and inactivate the complex. B) LSD1 alone does not act on H3K4me1/2 only when complexed with the nucleosome structure (*in vivo*). C) LSD1-CoREST complex can demethylate H3K4me1/2. D) BHC80 works as signal terminator, as its interaction with LSD1-CoREST complex causes the end of the activity *in vivo* (adapted from Wysocka, J. et al.²⁸).

Introduction

Chen et al.³³ described that the increase of the androgen receptor (AR) activation and its related target genes is an important mechanism for the conversion of metastatic-prostate cancer to a drug-resistant form.^{28,33} In 2005 it was demonstrated by Metzger et al.¹⁵ that LSD1 associated with AR acts as co-activator for the transcriptional activation of AR by ligand-induced demethylation of mono- or dimethylated lysine number 9 on histone number 3 (H3K9me1/2).^{15,28} The interaction of LSD1 with AR definitely causes substrate switching of LSD1 from H3K4me1/2 to H3K9me1/2 (figure 8). Subsequently, the demethylation of H3K9me1/2 promotes hormone-induced gene activation especially in those tissues where the androgen receptor has a key physiological role such as the prostate or testis. Other data supporting the role of LSD1 in prostate-cancer was assessed by Kahl et al.³⁴ and the overexpression of LSD1 in prostate cancer with high Gleason score, a prostate cancer evaluation parameter, could be demonstrated.³⁴

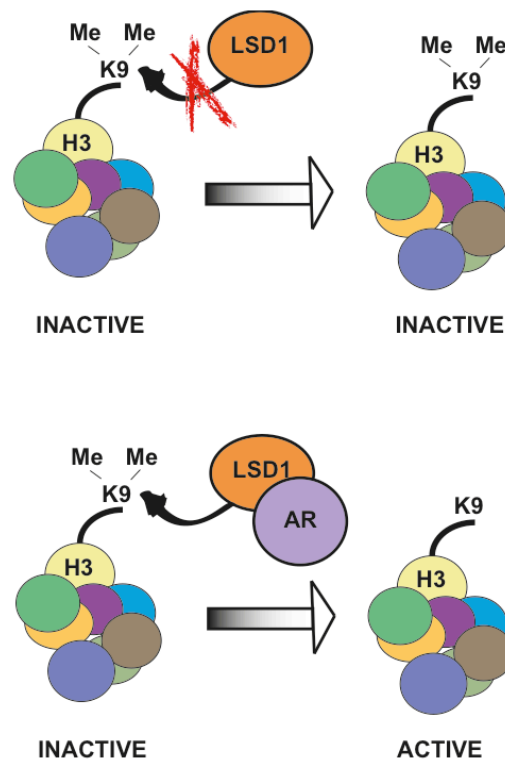


Figure 8: LSD1 cannot demethylate H3K9me1/2 in histone complex, as long as is not interacting with AR. Demethylation of H3K9 causes activation of the gene transcription.

However, the mechanism, that controls the androgen-dependent gene activation *in vivo*, is more complicated and several enzymes participate in the regulation creating a concerted mechanism not completely solved until now. For instance the full regulation of the LSD1 switching substrate is still unknown. It is well described in literature, that the gatekeeper of the androgen-dependent gene expression is the protein kinase C-related kinase 1 (PRK1), which after ligand-dependent recruitment of AR target genes, phosphorylates threonine number 11 on histone number 3

(H3T11).³⁵ This modification accelerates the demethylation by the JmjC domain-containing protein JMJD2C, the first histone tridemethylase regulating AR function, which together with LSD1 cooperatively stimulate AR-dependent gene transcription.³⁶ On the other hand, androgen-dependent recruitment of PRK1, activates protein kinase C beta I (PKC β _I) that phosphorylates threonine 6 on histone tail number 3 (H3T6), which is considered as the key event that prevents LSD1 from demethylating H3K4me1/2 during AR-dependent gene activation.³⁷ In support of this mechanism it was demonstrated that the level of PRK1, and its related phosphorylated H3T11, as well as the level of PKC β _I and the related phosphorylated H3T6 correlates with high Gleason score for prostate carcinoma,^{35,37} and that JMJD2C co-localizes with AR and LSD1 in normal prostate and in prostate carcinomas.³⁶

It has also been recently reported that LSD1 might act on other non-histone substrates.^{38,39} LSD1 represses the tumour suppressor and transcriptional activator p53, through demethylation of the p53-lysine number 370 (K370).³⁸ Moreover, LSD1 stabilizes the DNA-methyltransferase 1 (DNMT-1), though DNMT-1 residue-demethylation, inducing progressing loss of DNA methylation.³⁹

Interestingly, in 2009 a second mammalian FAD dependent amine oxidase enzyme was found and named LSD2 (KDM1B, AOF1).^{40,41} The protein shows the same number of residues present in LSD1, but rearranged in different domains. Indeed LSD2 is devoid of the tower domain, that provides the interaction with the corepressor protein CoREST in LSD1, while is showing a zinc-finger domain not present in LSD1 with still unknown functionality.⁴⁰ This new enzyme shows substrate-specific properties highly conserved to LSD1 but most probably LSD2 is involved in chromatin-remodelling complexes that are different from those involving LSD1,⁴⁰ increasing the complexity of the epigenetic code regulation.

1.4.2 Jumonji C (JmjC) domain-containing histone demethylases

The second class of HDMs was discovered in 2006, when the first member JmjC domain-containing histone demethylase 1 (JHDM1) was characterized.¹⁶ JHDM1 specifically demethylates mono-, di- and even trimethylated lysine number 36 on histone tail number 3 (H3K36) in a different manner than LSD1. The oxidation of the substrate is allowed by the reduction of an iron atom as co-enzyme using one molecule of α -ketoglutarate as co-substrate.^{13,16} During the catalytic reaction formaldehyde and succinate are produced (figure 9).^{13,16}

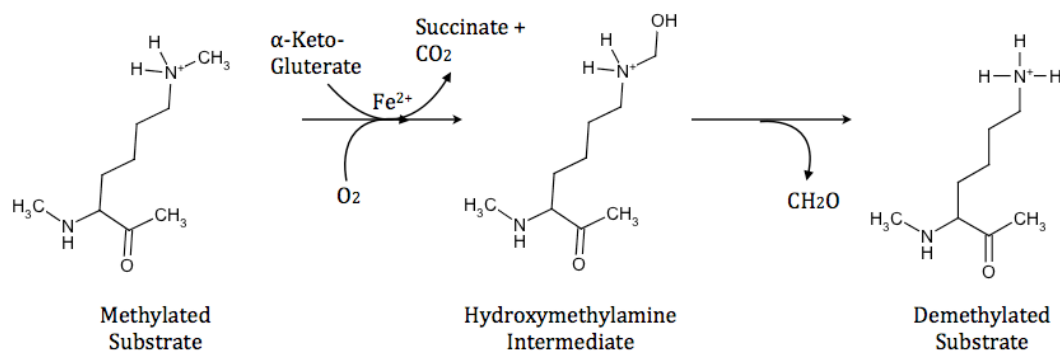


Figure 9: Catalytic mechanism of JmjC domain-containing histone demethylase.

Hitherto, 28 members of JmjC domain-containing histone demethylases have been characterized and linked to cancer.⁴² Klose et al.⁴³ divided the members into seven different JmjC domain-containing histone demethylases subfamilies. For a more complete description and classification see Hoffman et al.⁴².

1.5 LSD1 inhibitors

Until now several histone deacetylase inhibitors are in clinical trials or already on the market as anticancer drugs (Verinostat®, Romidepsin®),² while some HDMs inhibitors are in the pre-clinical phase,⁴² emphasizing the importance of these epigenetic targets as a new therapy approach not only for carcinomas, but also for other pathologies such as psychiatric diseases, age-dependent neurodegenerative disorders, and neurologic pathologies.

The generation of the first LSD1 inhibitors was based on the high sequence homology between monoamine oxidases (MAOs) and the LSD1 amine oxidase domain (AOD).⁴⁴ The approved anti-depressant MAO inhibitors, such as *trans*-1-phenylcyclopropylamine (PCPA), were found to be also active on LSD1.^{42,2} Subsequently, several well-known small molecule inhibitors of proteins carrying a flavin-dependent amine-oxidase domain (AOD) were also tested. Until today, known LSD1 active inhibitors are still small in number. They have been mainly derived from existing monoamine oxidase, polyamine oxidase or histone methyltransferase inhibitors (MAOIs, PAO, HMT respectively). In addition a few reversible LSD1 inhibitors have been described (Chapter 3.4).^{45,46}

1.5.1 LSD1 inhibitors derived from MAOs inhibitors

MAO inhibition has a clinical significance in psychiatry and neurology,⁴⁷ more precisely, selective and reversible MAO-A inhibitors have been widely employed for the treatment of depression and anxiety disorder, while MAO-B inhibitors are used as adjuncts in treatments of Parkinson's disease and they might contribute to the treatment of other neurodegenerative diseases such as Alzheimer's disease.^{48,49,47,50} To date, several MAO inhibitors have been found and they have been classified in selective, non-selective, reversible and irreversible inhibitors.⁴⁷

PCPA (**1**) was the first MAO-inhibitor found to be also active in a micromolar range as a covalent LSD1 inhibitor ($IC_{50} \sim 2\text{-}21\mu\text{M}$).⁴² Schenk et al.⁵¹ demonstrated that the combination of all-trans retinoic acid (ATRA) with PCPA decreases the engraftment of primary human acute myeloid leukemia (AML) cells *in vivo* in nonobese diabetic (NOD)-severe combined immunodeficient (SCID) mice, demonstrating that the combination of classical therapy plus PCPA which targets LSD1 might fight leukemia-initiating cells.^{42,51} Moreover, inhibition of LSD1 by PCPA causes a suppression of AR dependent transcription in bladder carcinoma cells.⁵² Other covalent MAO inhibitors such as pargyline (**2**) and phenelzine (**3**) and clorgyline (**4**) showed inhibitory activity on LSD1, but with a greatly decreased potency compared with PCPA.^{53,54} It has been also demonstrated that pargyline decreases the androgen-sensitive human prostate adenocarcinoma cells (LNCap) growth, and that it increases the levels of H3K9me1/2 by inhibition of LSD1 function.^{15,42} While Benelkebir et al.⁵⁵ showed that the more potent *p*-bromo or *p*-phenyl-analogs of PCPA inhibit the growth rate of LNCap cells in a micromolar range.^{42,55}

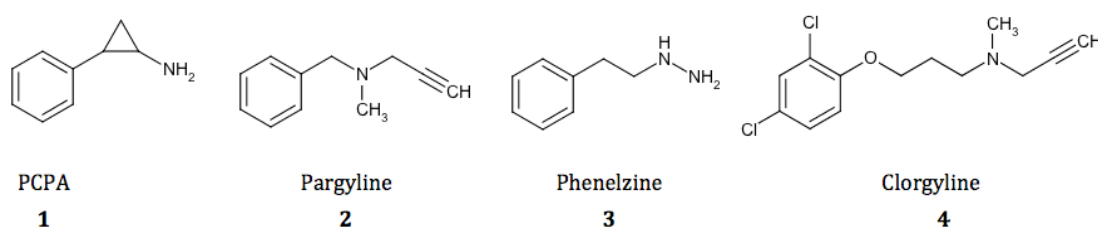


Figure 10: Well-known MAOs inhibitors active on LSD1.

In 2010 a structurally similar compound was patented (**5** is shown in figure 11).⁵⁶ It showed activity in *in vitro* assays, where it was more potent than PCPA (IC_{50} PCPA = $15.7\mu\text{M}$; IC_{50} **5** = 96 nM). *Ex vivo* experiments using MLL human AML cell lines,⁴² showed a significant reduction of the colony-forming cells and resulted in a more potent inhibition compared to PCPA (IC_{50} PCPA = $8\mu\text{M}$; IC_{50} **5** = 50 nM). This compound was able to reduce the clonogenic potential of cells and improve the induction of differentiation in murine and primary human MLL leukemia cells.^{42,57} Moreover in

Introduction

2010 and during the early 2011 other two N-alkylated PCPA derivatives were patented (**6** and **7** are shown in figure 11).^{58,59} These compounds inhibit LSD1 with high potency and selectivity over MAO-A and MAO-B (K_i **6** LSD1 = 9.0 nM; K_i **6** MAO-A = 15 μ M and K_i **6** MAO-B > 40 μ M, while K_i **7** LSD1 = 5 nM; K_i **7** MAO-A = 16 μ M and K_i **7** MAO-B = 74 μ M) although full details have not been released.⁶⁰ However, further new covalent inhibitors were presented by Binda et al.⁴¹ (**8**) that showed inhibition selectivity for LSD1 over MAO-A but not MAO-B, with K_i of 1.1 μ M on LSD1 (K_i PCPA = 271 μ M)⁴¹. The selectivity could be attributed to the presence of a bulky peptidomimetic moiety in the *ortho*- or *para*- position of the phenyl ring of PCPA, as found in similar derivatives reported by Ueda et al (**9**) in 2009.⁶¹ Meanwhile, some new derivatives (**10**) described by Mimasu et al.⁶² in 2010 showed a full selectivity for LSD1 over MAOs and a more potent inhibition of LSD1 with an IC_{50} value lower than 1 μ M (figure 11).⁶² Novel PCPA derivatives acting as LSD1 inhibitors were presented by Pollock et al in 2012.⁶³ These new derivatives are *p*-phenoethers (**11**) of PCPA (with an IC_{50} in the micromolar range) which showed an inhibitory activity on the cell growth of estrogen receptor alpha (ER α)-positive and ER α -negative tumors.⁶³ Since LSD1 was demonstrated to be mandatory for estrogen-dependent ER α -mediated transcription,⁶⁴ Pollock et al. could argue that inhibition of LSD1 with small molecules influence the proliferation of both ER α -positive and ER α -negative breast cancer cells.^{42,63}

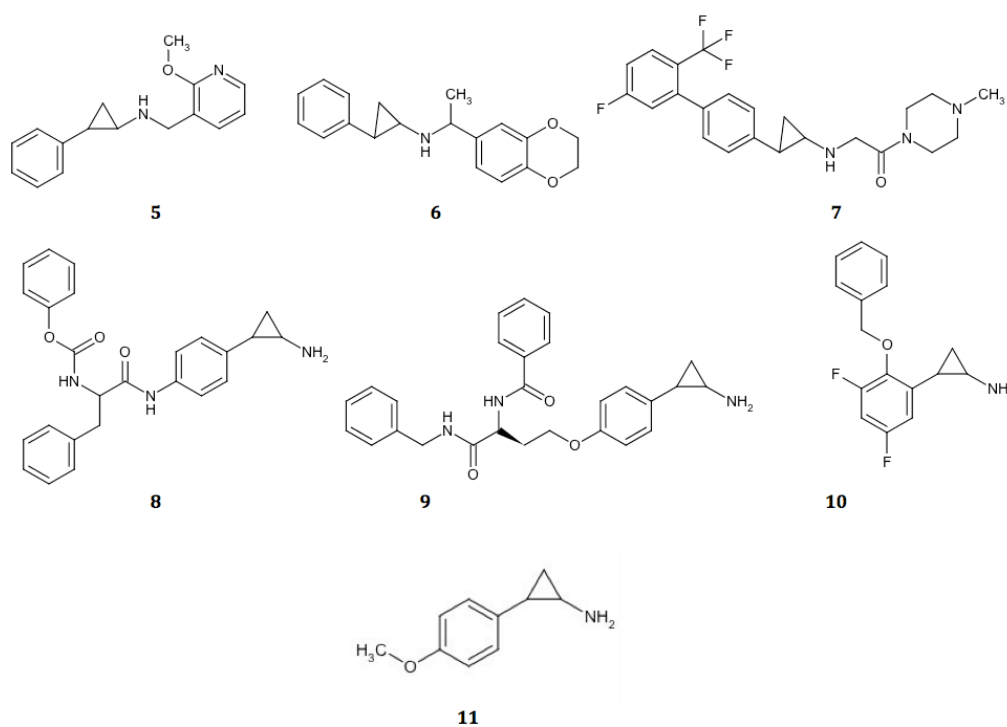


Figure 11: New LSD1 inhibitors derived from PCPA.

1.5.2 LSD1 inhibitors derived from PAO inhibitors

Polyamine oxidases (PAOs) represent a heterogeneous group of enzymes able to catabolise ubiquitous amines through oxidative deamination of spermidine, spermine and/or their acetylated derivatives.⁶⁵ The physiological value of PAOs substrates (essential growth factors) and the production of cytotoxic derivatives (aminoaldehydes and hydrogen peroxide) during PAOs active mechanism, give to these enzymes a strong regulative role in cellular proliferation and cellular death.⁶⁵ Only recently the interest in this class of enzymes was turned from tumour growth inhibition through PAO selective inhibitors, to PAOs catabolism pathways inhibition. In several tumour cells indeed PAOs catabolism pathways lead to apoptosis through the PAO mediated hydrogen peroxide production.^{65,66,67}

To date, the only three-dimensional structures of PAOs deposited in the PDB data bank have been obtained from maize or yeast.⁶⁸ However, the importance of *Zea mays* PAO crystal structure as a suitable model for designing new inhibitors, appropriate for all animal or plant PAO enzymes, has been demonstrated.^{65,69}

PAO inhibitors are mainly classified into four different classes:⁶⁵ (I) linear primary diamines,^{70,71} (II) agmatine and its analogs,⁷² (III) diguanidino inhibitors and analogs,⁷² (IV) polyamine analogs lacking terminal amino groups (figure 12).⁷¹ Some of the PAO inhibitors, also active as LSD1 inhibitors, were selected for further computational studies during this work using both PAO and LSD1 enzymes in order to figure out the binding differences. Since the similarity of LSD1 amine oxidase domain (AOD) to PAO including spermidine oxidase AOD is well demonstrated (Chapter 3.1), biguanidine (**12** as the most active molecule) and bisguanidine (**13** as the most active molecule) inhibitors were tested on LSD1 enzyme, and they were demonstrated to be non-competitive inhibitors *in vitro*.⁷³

Introduction

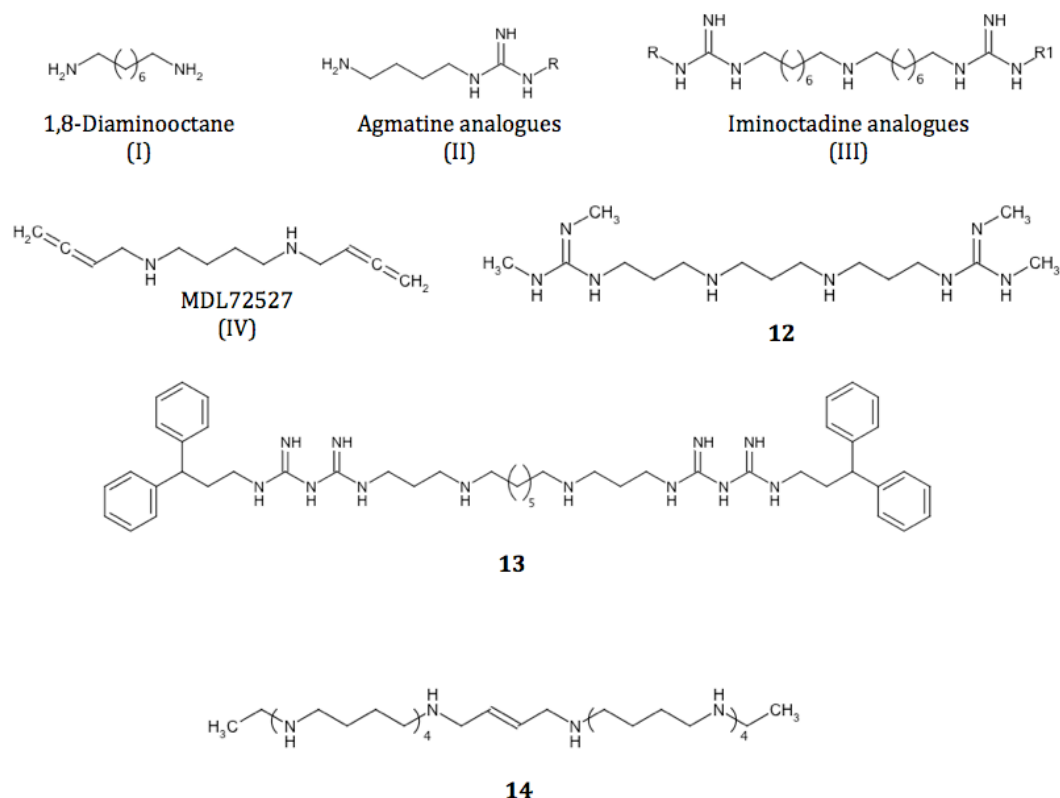


Figure 12: Classes of PAO active inhibitors and PAO inhibitors active on LSD1.

These two new compounds increased the H3K4 hypermethylation state at the promoter regions of the effected genes in HTC116 human colon carcinoma cells through LSD1 inhibition.^{42,73} Two years later the same group published other oligoamine inhibitors (**14** is the most active inhibitor) with similar effect on human colon carcinoma cells, but with a different active mechanism, indeed these new compounds showed reversible and substrate competitive inhibition in *in vitro* assays.^{42,74} It was also shown that the combination of these compounds (**13** and **14**) with DNMT inhibitors significantly reduced the tumour growth in athymic nude mice bearing HTC116 xenografts,⁴² indicating a new possible synergic approach for anticancer therapy.

1.5.3 LSD1 reversible inhibitors

In the last years novel small molecules showing a non-covalent allosteric inhibition of LSD1 were reported (the most active CBB1007 (**15**)).⁴⁶ These compounds contain a guanidino group in their structures and were demonstrated to inhibit cancer cells with pluripotent stem cell properties but not non-stem cell lineages, with *in vivo* IC₅₀ values lower than 5.27 μ M (selectivity over MAO has not been evaluated).⁴⁶

Further reversible LSD1 inhibitor has been presented by Willmann et al.⁴⁵ in 2012 (Namoline (**16**)). The compound shows a γ -pyrone scaffold and acts as non-covalent and LSD1 selective inhibitor with an *in vitro* activity of 51 μ M.⁴⁵ Treating LNCap prostate cancer cells with Namoline showed increased levels of H3K9me1/2, this indicates that the antiproliferative action might be linked to LSD1 inactivation and that Namoline has good cell permeability. Moreover, Namoline stops the tumour cells growth in LNCap xenografted mice.⁴⁵ This discovery opens the way to introducing new scaffold modifications in order to generate new potent LSD1 active inhibitors (Chapter 3.4).

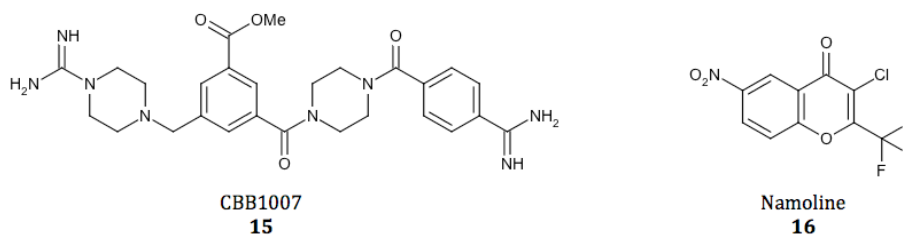


Figure 13: New non-covalent LSD1 inhibitors.

1.6 Aim of the work

It has been established that targeting the epigenetic code through small molecule inhibitors is a promising therapeutic strategy for a wide range of diseases. LSD1 as a well described epigenetic target reveals crucial importance in several tumorigenic mechanisms, especially in tissues where AR has a key physiological role for example prostate cancer. Up to date, there are only few well-characterized inhibitors of LSD1 available, and almost all of them show covalent inhibition and low selectivity over off-target proteins, such as MAOs or PAO enzymes.

From the exigency to find new inhibitors, our work was focused on the generation of new LSD1 inhibitors using bio-analytical and *in silico* studies. The experimental data were obtained in collaboration with Prof. Manfred Jung's group (University of Freiburg –DE-).⁷⁵

Starting from the analysis of LSD1 crystal structures (Chapter 3) the first step was to figure out differences and similarity in the binding modes among proteins showing AOD similarity like MAOs, and PAO enzymes. Known MAOs, PAOs inhibitors were docked to this target.

Molecular dynamic simulations were carried out in order to analyse the binding of LSD1 inhibitors as well as to discriminate between different binding modes derived by the docking studies. The validated models were then used for structure-based virtual screening experiments.

In order to generate suitable *in silico* models (descriptive and predictive) able to differentiate among LSD1 active and inactive compounds, also binding free energy calculations were carried out.

MATERIALS AND METHODS

2

Considering the development of rapid computer technology and the exponentially increasing number of protein structures deposited into structural databases such the Protein Data Bank (PDB),⁷⁶ it is easy to understand that computer-related disciplines like computer assisted molecular design (CAMD) have extremely improved in the last two decades. In this chapter CAMD tools, concerning hits selection and characterization, will be described. Binding free energy calculation methods used in this work will be also briefly described.

2.1 Quantum mechanics and empirical force field methods

Molecular modelling techniques can be divided into two main groups:

- Quantum mechanics.
- Empirical force field methods (molecular mechanics).

The main difference between quantum mechanics (QM) and molecular mechanics (MM) is the accuracy of the system. While QM models are taking into account both position of nuclei and electrons of each atom that belongs to the system; MM models simplify the system neglecting the electrons according to Born-Oppenheimer approximation.⁷⁷ This approximation drastically decreases the accuracy level (MM models are not able to investigate chemical reactions in which bond are broken or formed), but the number of atoms which MM models are able to handle and the calculation speed increases exponentially.

Empirical force field methods evaluate the potential energy of a system considering the position of the nuclei only. Two main interactions contribute to the final potential energy: bonded and non-bonded interactions. Bonded interactions can be further divided into bond stretching, which represents the energy to stretch a bond between two atoms, angle bending which expresses the energy necessary to modify the bond angle between three atoms compared to the equilibrium value and bond torsion, which describes the energy to rotate a torsion angle formed by four atoms. Non-bonded interactions refer to atoms that belong to the same molecule; separated at least by three atoms; or by two atoms that belong to different molecules. van der Waals interactions which considers the energy between two atoms not directly bonded and is usually represented by Lennard-Jones potential, and electrostatic interactions which represents the electrostatic energy between two

atoms not directly bonded and evaluated through the Coulombic interaction term are terms that belong to non-bonded interactions. The sum of all the contributions gives the potential energy of the system according to equation 2.1 (for a more accurate description see Appendix A).

$$E_{Pot} = \sum_{bonds} E_{Stretching} + \sum_{angles} E_{Bending} + \sum_{torsions} E_{Torsion} + \sum_{non-bonded} E_{vdW} + \sum_{non-Bonded} E_{Ele} \quad (2.1)$$

Up to now, several empirical force fields have been developed according to different classes of biomolecules.

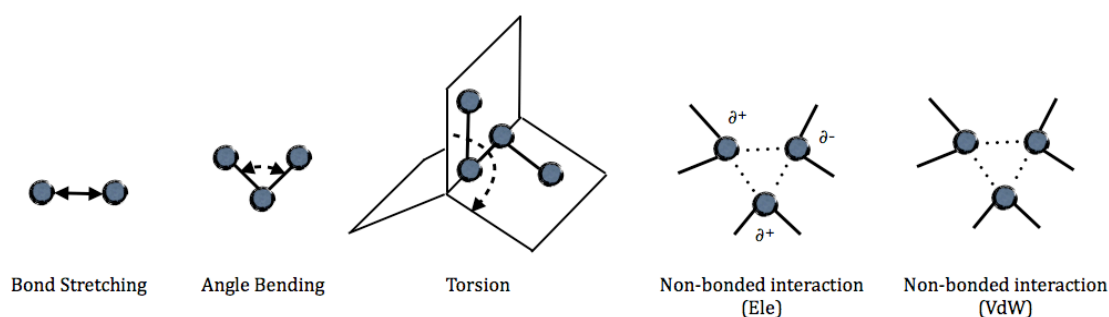


Figure 14: Bonded and non-bonded interactions implemented in a force field potential that contribute to the final potential energy

Warshel and Levitt introduced a good compromise between QM and MM models using QM-MM hybrid models.⁷⁸ In these models the system is divided into two regions: QM region (which is usually the inner region of the system) and MM region (which usually surrounds the QM region). The division allows the study of relevant processes that need to be accurately described electronically by quantum mechanics, for instance the active site of an enzyme, neglecting other regions of the system, described with empirical force field methods. The idea is that QM and MM atoms can recognize and interact among each other.⁷⁹ Here the potential energy is the sum of the QM energy term and MM energy term. During the calculation of QM energy term, MM atoms are considered as point charges, and they can influence the QM region, meanwhile van der Waals interactions are electron independent and therefore are calculated by classical molecular mechanics.⁷⁹

2.1.1 Energy minimisation and molecular dynamic simulation

The potential energy of a system is a multidimensional function of its coordinates. For instance the potential energy of a system formed by N atoms is a function of $3N$ Cartesian coordinates.⁷⁷ For a

system with more than one or two coordinates it is impossible to visualize the entire multi-dimension energy surface. In molecular modelling, the minimum points on the energy surface are considered the most interesting ones, because, according to Boltzmann distribution rule (2.2), the minimum points of the energy surface are the most populated and stable states of the system.

$$\frac{N_i}{N} = \frac{g_i e^{-E_i/k_b T}}{Z(T)} \quad (2.2)$$

Two different kinds of minima are present in a multi-dimensional energy surface: a so-called *global energy minimum* that is the very lowest energy point and a *local energy minimum*. The energy minimization process moves the system into the valley of the potential energy, which is located closest to the starting point. The energy minimisation procedure does not allow identifying the global minimum as long as the starting point is not close to the global minimum and it does not allow exploring large portions of the potential energy surface as well. Several algorithms are available to carry out energy minimization. These can be grouped into: energy based methods (e.g. simplex method) that are taking into account only the system energy value without deriving the energy surface; gradient based algorithm (e.g. steepest descent, and conjugate gradient) using the first derivative of the multi-dimensional energy surface; and algorithms that use the second derivative of the multi-dimensional energy surface (e.g. Newton-Raphson).^{77,79}

Molecular dynamic (MD) simulation generates several molecular configurations of a system, characterized by physical movement of atoms and molecules. This technique allows studying the atoms movements and the system properties like potential energy, kinetic energy, temperature, density, volume, etc. in a specific time range. Numerically, the trajectories obtained during the MDs are generated by integrating the Newton's law of motion:

$$F = m_i a_i \quad (2.3)$$

The equation describes the motion of a particle i with a mass m and acceleration a , with F as the force of the particle along a direction x .⁷⁷ MD simulation is a powerful tool that allows exploring the multi-dimensional surface of a generated system (conformational space exploring). This useful method finds also a wide use in the evaluation of protein models, for instance obtaining the energetically most favourable and stable structure. From the first MD simulation (almost 30 years

ago)⁸⁰ several progresses in Newton's equation integration algorithms and computer performances have been reached, allowing this method to handle bigger systems, switching from in vacuum simulations to more realistic simulations using explicit water molecules for instance, and longer simulations from picoseconds to microseconds in length⁸¹. Despite the strengths of MD simulations, this technique still has several limitations that must be considered. Electron neglecting does not permit studying bonds generations, isomerisation and bond breaking. Neither charge-transfer complexes can be simulated. Again, the size of the systems is still limited to "small dimensions" and to short time running, causing a limited exploration of the conformational space. Systems with high degrees of freedom or high energy barriers in the multi-dimensional surface might represent a limitation in the conformational space exploration, where MD simulations are not able to climb over energetic picks in relatively short time; especially for big systems.⁷⁹ For these reasons, the quality of the data obtained from MD simulations is strongly correlated with the running settings, the accuracy of the force field selected, and the user capabilities.⁸² The most used MD software are AMBER,⁸³ GROMACS⁸⁴ and CHARMM.⁸⁵

2.2 Molecular docking

Molecular docking is a CAMD tool that attempts to predict the molecular conformation between a ligand and a certain protein.⁷⁹ Nowadays, there are several algorithms able to solve the docking task. They are grouped mainly into deterministic and stochastic approaches.⁷⁹ However, up to now, a substantial problem still exists called the "Docking Problem" as described by Blaney and Dixon in 1993.^{77,79}

The docking problem is related to the flexibility and to the high number of degrees of freedom considering both the ligand and the protein.⁷⁹ In order to reduce the degrees of freedom, most of the docking algorithms consider the protein as a static object keeping the ligands flexible. Only in the last few years the main used docking programs implemented side chain flexibility that allow the mobility to some selected protein side chains.

The main docking algorithms are classified into: *Incremental Construction Methods* where the ligand is divided into small fragments which are rebuilt inside the protein active pocket.^{79,86} The *Genetic Algorithm* (GA) "mimics the evolutionary process manipulating a collection of structural data, called chromosomes".^{79,87} *Taboo Search* (TS) algorithm, starts allocating a completely random solution inside the active cavity and through imposing restrictions "enables a search process able to handle difficult region".^{79,88} *Monte Carlo Simulation* (MC) samples the conformational space by

random movements. Several docking program implement MC search in order to handle the degrees of freedom of the ligand.^{79,89}

2.2.1 Scoring functions

In order to rank the docking solutions generated from docking algorithms according to their binding constant K_i , every docking program provides a scoring function.⁷⁹ K_i is related to the Gibbs free energy of binding (ΔG) by the equation 2.4.

$$\Delta G = -RT \ln K_i \quad (2.4)$$

Where T is the temperature and R is the gas constant. For every docking software the ΔG approximation of a protein-ligand complex is an important goal as the prediction of the right ligand geometries inside the active pocket of a protein.⁷⁹

For all scoring functions, ΔG prediction is not an easy task. For this reason the accuracy of a scoring function is an important parameter that should be taken into consideration during a docking procedure.⁷⁹ Scoring functions are arranged into three main groups: (i) empirical scoring functions: they optimize, with a multilinear regression (MRA) the coefficients of every structural functions by using a training set of protein-ligand complexes, where the binding free energy has already been experimentally evaluated.^{79,90,91} (ii) Force-field-based scoring functions approximate the protein-ligand ΔG to the potential energy using the non-bonded term of a classical molecular mechanic force field. It uses the Lennard-Jones potential to describe the van der Waals interactions, and the Coulomb potential to describe the electrostatic interactions.⁷⁹ Measurements of the potential energy might however represent two disadvantages: on one hand, the entropic component of the binding free energy is omitted, on the other hand, the electrostatic term might heavily influence the final score.^{79,92} (iii) Knowledge-based scoring functions use a training set to derive the atom pair potentials and the final mathematical function.^{93,94,95,96} A major limitation, excluding the big number of protein-ligand complexes to use as a training set, is the capability for the function to rightly estimate the free energy of binding for those complexes that are not present in the training set.⁷⁹

Empirical scoring functions are fast and show a good prediction for the well-known protein-ligand complexes,^{91,97} meanwhile force-field-based scoring functions are the most computationally

demanding.^{98,99} Up to now, no scoring function is able to fully predict the ΔG for a protein-ligand complex,^{98,99} but all of them are relatively fast compared to other time consuming techniques such as energy perturbation¹⁰⁰ and are able to nearly estimate the free energy of binding. For this reason a combination of two or more scoring functions, so-called consensus ranking, might be a good compromise between speed and accuracy.

2.2.2 Docking programs

The following docking software were used in this work:

Gold (Genetic Optimization for Ligand Docking)⁸⁷ is based on the GA, which considers the ligand orientation and flexibility inside the protein binding pocket as a chromosome. Through chromosome genetic rules, like crossover, mutation and translocation, the GA is able to randomly create new conformations and positions inside the pocket. Only the favourable modifications, characterized by a score able to describe the suitability of the complex generated, are able to transfer the information to the next chromosome generations. The software is also provided with a force-field-based scoring function that takes into consideration protein-ligand van der Waals energy, protein-ligand hydrogen bonding, ligand internal van der Waals energy and ligand torsional strain energy.

Glide uses several filters in order to search for the possible ligand poses inside the protein binding pocket.⁹⁰ After the minimization step with a standard molecular mechanic energy function (OPLS-AA force field¹⁰¹ implemented in the software), only the best poses, i.e. the energetically most favourable, which are already located inside the cavity, are processed with MC sampling techniques.²⁰ This reduces the computational demand and the calculation time. *Glide* adopts an empirical scoring function that takes into consideration the protein-ligand lipophilic term (ligand-atom/receptor-atom pair defined as lipophilic), the ligand-protein hydrogen-bonding term, the metal-ligand interaction term, the protein polar but non-hydrogen-bonding atoms term, the protein-ligand van der Waals and Coulomb (electrostatic) interactions energy terms, and the solvation term (which uses an empirical scoring term that measures the exposure to explicit water for every atom of the ligand).⁹⁰

ParaDockS is an in-house software developed by René Meier et al.¹⁰² The software is available online as an open source package, and the download is free of charges. The algorithm of the

software follows the particle swarm optimization (PSO)¹⁰³ as an optimizer implementation. PSO is a metaheuristic method of search and optimization, inspired from nature (swarms movement: birds, fishes etc.). *ParaDockS* is also provided with two objective functions: (i) an empirically derived energy function (p-Score), which takes into account van der Waals interactions, van der Waals clashes, electrostatic interactions; and (ii) a knowledge-based objective function (PMF04).¹⁰²

2.3 Virtual screening

High-throughput screening (HTS) of chemical libraries is a commonly used approach in drug discovery to find new hit compounds.⁷⁹ However, despite the accuracy of the method, the high costs (taking also in consideration increasing size of the available databases) are rendering this methodology not affordable to everyone.⁷⁹ A computational approach, able to pre-select only a subset of compounds from the selected main database, might be useful to reduce the HTS costs. This approach is well-used and well-known as Virtual Screening (VS).¹⁰⁴ Nowadays, several VS technologies have been developed and validated. They might be classified into two main groups: Ligand-based and Structure-based VS. The first approach involves constraints based on the similarity of a set of known actives¹⁰⁵ or by using ligand pharmacophores,¹⁰⁶ while the second method is based on the 3D structure of the target, using docking procedures.¹⁰⁷

Considering both approaches, the preparation of the electronic library is one of the limiting steps for the VS success.⁷⁹ Database selection, among several commercially available screening collections that differs in size, type of molecules etc., is the first step that should be applied in a VS procedure.⁷⁹ Other steps should be applied in order to: reduce the number of molecules to screen, to reduce the computer resources and eventually to reduce the computational time. A good way of acting is to use a series of filters able to select and exclude highly reacting or toxic compounds. Up to now, the most used filter is the Lipinski's 'Rule-of-Five',¹⁰⁸ which evaluates the bioavailability of a compound, through the evaluation of the molecular weight (not over 500 Da), the lipophilicity, calculating the octanol-water partition coefficient (clogP; not more than 5), the total number of nitrogen and oxygen atoms (not more than 10), and the number of hydrogen bond donors (not more than 5) in the molecules. This rule allows keeping compounds in the database with predictable molecular absorption and cell permeability in the body compartments. More elaborate and specific filters; as filters that take into account ADME (adsorption, distribution, metabolism and excretion) properties or knowledge-based binary classification system¹⁰⁹ are also applicable during database filtering. Finally another filter that excludes the so called "promiscuous binders"¹¹⁰ (molecules with

high trend to bind several protein targets) might be also considered for a good virtual screening procedure.⁷⁹

2.4 Binding free energy calculation methods

Finding ligands that can reversibly bind to a specific target (protein agonist/antagonist, enzyme inhibitor etc.) is the main goal of every structure-based drug design project.¹¹¹ Measuring the energetics of molecular complexes and the energy prediction of protein-ligand non-covalent associations has been shown to be an important procedure in computational chemistry,¹¹¹ especially after demonstrating the agreement of these *in silico* measurements with experimental data.¹¹¹ Several methods have been developed, ranging from rapid methods to highly time-consuming methods. The rapid methods for binding free energy estimation are either empirical scoring approaches, which are based on a simple energy function,^{91,97,111,112} or knowledge-based scoring approaches, which are based on the frequency of occurrence of different atom-atom contact pairs in known structure complexes.^{95,96,111} The lack of structure sampling and explicit water treatment render this methods fast but not accurate.¹¹¹ More accurate methods are based on molecular force-fields and involve a structure ensemble of the system studied, which is generated by MD or by MC analysis techniques.¹¹¹ Obviously this causes an increase in the accuracy but also in the calculation time-demand, since a large number of pair-wise interactions should be calculated. These procedures are mostly used in order to estimate relative free energy differences between two equilibrium states of interest, where the relative binding free energy is expressed as ΔG .¹¹¹ For instance, estimation of the differences in the binding free energy between two similar ligands that bind the same receptor is one of the main applications of these techniques (figure 15).¹¹¹

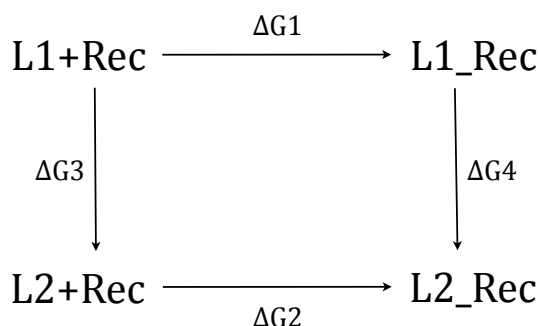


Figure 15: Thermodynamic cycle for two ligands: L1 and L2 to receptor R. L+Rec refers to ligand 1 or ligand 2 in solution. While L_Rec refers to ligand 1 or ligand 2 bound to the protein.

This application refers to the thermodynamic cycle shown in figure 15^{111,113} which is used to evaluate the binding free energy between ligand L1 and ligand L2, that are able to bind the same receptor Rec. $\Delta G1$ and $\Delta G2$ are the corresponding binding free energies of L1 and L2; $\Delta G3$ corresponds to the free energy differences of the two ligands in solutions meanwhile $\Delta G4$ is the free energy differences of the two ligands in complex with the protein. The relative binding affinity of L1 and L2 is the difference between $\Delta G2$ and $\Delta G1$, and it is written as $\Delta\Delta G$. The free energy is a state function; it means that its value round the thermodynamic cycle must be zero⁷⁷ so that the $\Delta\Delta G$ can be also written as difference between $\Delta G4$ and $\Delta G3$.

$$\Delta\Delta G = \Delta G2 - \Delta G1 = \Delta G4 - \Delta G3 \quad (2.5)$$

Nowadays the most accurate methods to evaluate binding free energies are free energy perturbation (FEP) and thermodynamic integration (TI). These methods calculate the binding free energy by decomposing the binary system (L+Rec \rightarrow L_Rec) into several unphysical intermediate-overlapping states through MD or MC simulations.^{77,111} Both technologies are not only time and computationally demanding, but their applicability expires if the examined molecules differ significantly from each other.¹¹¹ Other methods have been developed, like molecular mechanics Poisson-Boltzmann surface area (MM-PBSA, described in the next paragraph)¹¹⁴ and linear interaction energy (LIE),^{115,116} which considers only the physical relevant state (thermodynamic cycle corners) and approximates the binding free energy by calculating the electrostatic and van der Waals contributions for the ligand surrounded by the solvent and the ligand surrounded by the protein. The estimation of non-polar contribution is carried out using an empirically derived parameter that scales the van der Waals contribution from the MD simulation.^{111,115,116} This results in a less computationally demanding technique, compared to FEP and TI, keeping a good agreement with the accuracy of the calculations.

2.4.1 Molecular mechanics Poisson-Boltzmann surface area (MM-PBSA)

MM-PBSA method and its complementary molecular mechanics Generalized-Born surface area (MM-GBSA) method are based on the analysis of MD simulations using continuum solvent approach, deriving the free energy average of molecular complexes according to equation 2.6 (for more details see Appendix A):

$$\langle G \rangle = \langle E_{MM} \rangle + \langle G_{PBSA} \rangle - T \langle S_{MM} \rangle \quad (2.6)$$

Where $\langle E_{MM} \rangle$ is the average molecular mechanics energy that includes the internal energies, the electrostatic contributions and the van der Waals energies obtained from molecular mechanic force field and evaluated with non-bonded cutoff. The $\langle G_{PBSA} \rangle$ term is the average estimation of the nonpolar free energy contribution. This term is evaluated using the surface area and a numerical solution of Poisson-Boltzmann equation.¹¹⁷ The two terms together are usually obtained by averaging the most representative geometries from sampled MD trajectories of the system. Both terms combined approximate the enthalpy of the system when the simulations are performed at constant pressure and constant temperature (NTP, Gibbs binding free energy). The last term $-T \langle S_{MM} \rangle$ refers to the solute entropy. This is obtained by the quasi-harmonic analysis of the trajectories or by normal mode analysis.¹¹⁸ If initially MM-PBSA approach was used for the evaluation of DNA and RNA fragments' stability,¹¹⁸ nowadays, there are several examples where binding free energy calculation of protein-ligand complexes were carried out using MM-PBSA method through equation 2.7 (Appendix A).^{114,119,120}

$$\Delta G_{Bind} = \langle G_{complex} \rangle - (\langle G_{protein} \rangle + \langle G_{Ligand} \rangle) \quad (2.7)$$

There are two available ways to calculate the binding free energy of protein-ligand complexes using MM-PBSA method. In the first case the terms for the complex, the protein, and the ligand based on separate trajectories are calculated using equation number 2.6 and subsequently, ΔG_{Bind} might be calculated with equation 2.7. In the second case, ΔG_{Bind} is calculated determining every single term in equation 2.7 based on snapshots from a trajectory of the complex only.¹¹¹ The first pathway seems not to be useful in protein-ligand binding free energy calculations, since $\langle E_{MM} \rangle$ term for the protein and the complex of the system requires a long computational time to converge. Another possibility to calculate the protein-ligand binding free energy using MM-PBSA method is by using a single trajectory approach. The latter case is considered only suitable when no important conformational changes occur in the protein structure during MD.

The most difficult task to manage in the MM-PBSA technique is the accurate determination of entropy contribution, especially where significant conformational fluctuations occur. This problem opens two different scenarios. On one hand the possibility to omit the entropy contribution in case the binding free energy values are calculated for a series of ligand with similar size that bind a

common protein receptor.^{111,119,121} On the other hand this term becomes mandatory for a series of ligands with different structures and hence different degrees of freedom.^{111,119,121}

2.4.2 Quantum-mechanics, molecular mechanics Poisson-Boltzmann surface area (QMMM-PBSA)

In protein-ligand binding affinity, polarizations contributions and electronic contributions play an important role in processes like formation/cleavage of covalent bonds and fluctuation of charge during molecular dynamics.¹²¹ These contributions are often neglected using MD simulations with MM potentials. MM-PBSA and MM-GBSA are also neglecting these two important contributes during free energy calculations, especially in protein-ligand binding free energy calculations. In order to avoid this shortcome and increase the computational accuracy, several QM/MM models were carried out in recent years.^{121,122,123} Some studies showed better protein-ligand binding free energy results, using mixed quantum mechanic/molecular mechanic Poisson-Boltzmann surface area models (QM/MM-PBSA), according to equation 2.8 (Appendix A).^{121,124}

$$\langle G \rangle = \langle E_{QM/MM} \rangle + \langle G_{PBSA} \rangle - T \langle S_{MM} \rangle \quad (2.8)$$

The only difference to equation number 2.6 is the presence of a QM region in the term related to the potential energy of the system $\langle E_{QM/MM} \rangle$. This new term is calculated according to hybrid QM-MM models

$$E_{QM/MM} = E_{QM} + E_{MM} + E_{QM/MM} + E_{pol} + E_{boundaries} \quad (2.9)$$

Where E_{QM} and E_{MM} are the potential energies related to the QM and the MM region, respectively, while $E_{QM/MM}$ is the potential energy at the interface between the two regions. The last two terms refer to the influence that one region has on the other one, E_{pol} , and the potential energy derived from the boundaries conditions.

Depending on the acceptable computational cost, a larger QM region might be chosen.¹²⁴ Including some residues that interact with the ligand inside the QM region (from four to six angstrom around the ligand allocated inside the binding pocket), can give a better comprehension of the electronic

effects involved in the protein-ligand binding and of the polarization due to the residues present in the protein binding pocket.¹²⁴ Sampling the system with MD simulations and using several snapshots is also allowed to estimate the binding free energy. One has to keep in mind that when sampling through a classical MD simulation each ligand would need to be properly parameterized, which is computationally costly.¹²⁴ Another QM/MM-PBSA approach is to include only the ligand in the QM region, enabling the sampling of the protein and removing the mandatory parameterization of the ligand.¹²¹ Regarding the computational cost and the accuracy in the binding free energy calculations, QM/MM-PBSA method still has a considerable scope for improvement.

2.6 In vitro assays for LSD1 demethylases activity

In order to validate random or virtual screening results, *in vitro* testing are usually conducted, since they are easy to manage and allow high throughput. After that, *ex vivo* testing allows verifying the results in a cellular system, bringing them in a functional or phenotypic context and on the other hand, they allow avoiding systematic errors (bias) that can easily give false-positive results.^{125,126} Here are described briefly the selected LSD1 *in vitro* assays conducted during this work.

2.6.1 Peroxidase assay

From the demethylation mechanism of LSD1 (figure 3) two low molecular weight products are generated which are suitable for quantification: hydrogen peroxide (H₂O₂) and formaldehyde.

Horse radish peroxidase (HRP) is an enzyme that oxidises one molecule of hydrogen peroxide into oxygen with a simultaneous reduction of co-substrate. Now, choosing a co-substrate able to change its spectral properties according to its oxidation/reduction form might be useful for the detection of the resulting product. Two co-substrate have been described for the measurement of LSD1 activity: 4-aminoantipyrine and Amplex® Red.¹⁷ 4-Aminoantipyrine (**A**) is oxidized in the process, and upon reacting with a phenolic compound it generates a quinone structure (**C**), which has its maximum wavelength absorption at 500 nm (figure 16).¹⁷

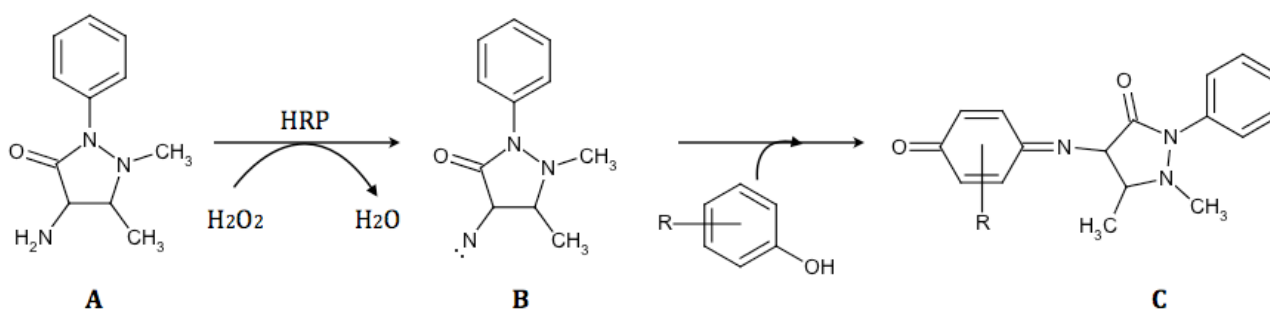


Figure 16: Peroxidase assay for LSD1 activity measurement using 4-aminoantipyrine and phenol as substrates of HRP. The enzymatic activity is measured through light absorption of the final quinone derivative.

Amplex® Red (**D**) is the co-substrate selected for the peroxidase assay conducted in this work. This molecule in presence of HRP and hydrogen peroxide is converted to resorufin (**E**) a fluorescent agent that has excitation/emission ($\lambda_{ex}/\lambda_{em}$) maxima at 570/585 nm, representing the most sensitive less error-prone method (figure 17).^{127,128}

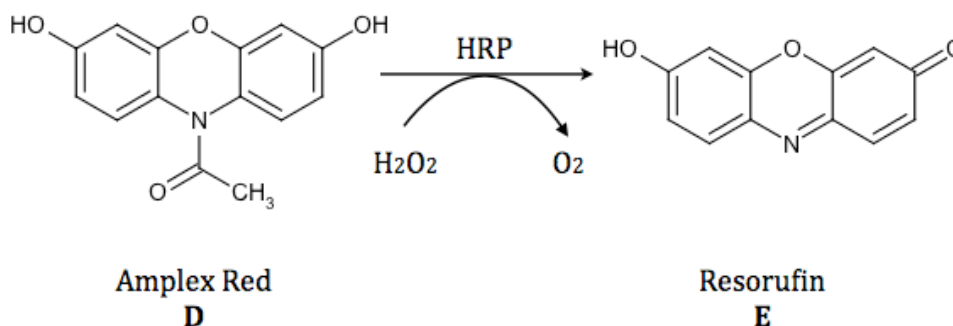


Figure 17: Peroxidase assay for LSD1 activity measurement using Amplex Red as substrate of HRP. The enzymatic activity is measured through fluorescence emission of the final product resorufin.

For the assays details see Appendix A.

2.6.2 Formaldehyde-dehydrogenase-dependent (FDH) assay

Formaldehyde is the second low molecular weight product formed during the active mechanism of LSD1 enzyme. This product in presence of NAD⁺ is oxidized by formaldehyde dehydrogenase forming formic acid while the co-substrate NAD⁺ is converted to its reduced form NADH. Now, NADH differs from NAD⁺ in to its fluorescence spectrum, therefore the final reduced product (NADH) might be measured fluorometrically ($\lambda_{ex} = 330 \text{ nm}$, $\lambda_{em} = 460 \text{ nm}$).¹²⁹

2.6.3 Antibody based assay

Another available approach to measure the LSD1 activity is using specific antibodies able to recognize the subtle difference between un-methylated and mono-methylated H3K4 peptide. Since LSD1 is also active on the C-terminal biotinylated H3K4me1 peptide (biot-H3K4me1) Perking Elmer Company took advantage of this feature to develop an AlphaLISA assay (figure 18).¹³⁰ Initially the biot-H3K4me0 is recognized and bound by a modified antibody carrying an acceptor bead. After that, the donor bead streptavidin is added to the system and binds the biotin structure in close proximity to the acceptor bead. Irradiating the system using light with a wavelength of 680 nm causes the donor bead to release a singlet oxygen, which reacts with the acceptor bead causing it to emit light with wavelength of 615 nm. If there is no acceptor bead in close proximity, the singlet oxygen reacts very quickly, due to its short half-life and no emitted light is observed.¹³⁰

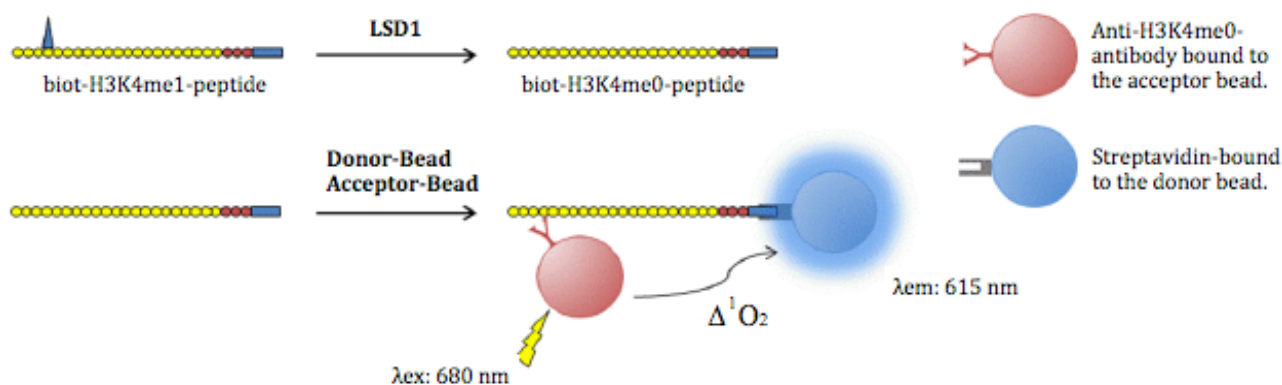


Figure 18: Mechanism of LSD1-AlphaLISA-assay developed by Perking Elmer Company. Picture adapted from Martin Leo Schmitt, Freiburg.

Perkin Elmer company developed another antibody assay called time resolved-fluorescence resonance energy transfer assay (TR-FRET). In principle the assay is very similar to AlphaLISA assay and it works with the same antibody that is chelated to an europium atom, which behaves as the acceptor bead. Once again the fluorophore donor bead streptavidin is added to the system to bind the biotinylated peptide. Now irradiating the system with light causes europium atom excitation, which transmits the radiated energy to the nearby fluorophore that emits light with wavelength of 665 nm.

The most used antibody assay for the detection of LSD1 activity during the work was a modified dissociation-enhanced lanthanide fluorescent immunoassay (DELFI[®]). The DELFI[®] assay is a modified sandwich ELISA assay (enzyme-linked immunosorbent assay). Here the peptide substrate (H3K4Me1) is immobilized to a streptavidin coated microtiter plate and the reaction is started by

the addition of the enzyme (LSD1). Alternatively, the reaction can also be started by adding the enzyme to the unbound substrate (H3K4Me1), and after the demethylation reaction is finished the demethylated product (H3K4Me0) is immobilized to a streptavidin coated microtiter plate (figure 19-B). Thereafter a modified antibody binds the modified substrate (H3K4Me0) exposing the fragment crystallizable region (F_c-region). This region now can be recognized by another antibody labelled with a chelating europium (Eu³⁺). Eu³⁺ is then released from the complex by means of an enhancement solution, and can be detected by time-resolved fluorescence (λ_{Ex} : 340nm, λ_{Em} : 615 nm).

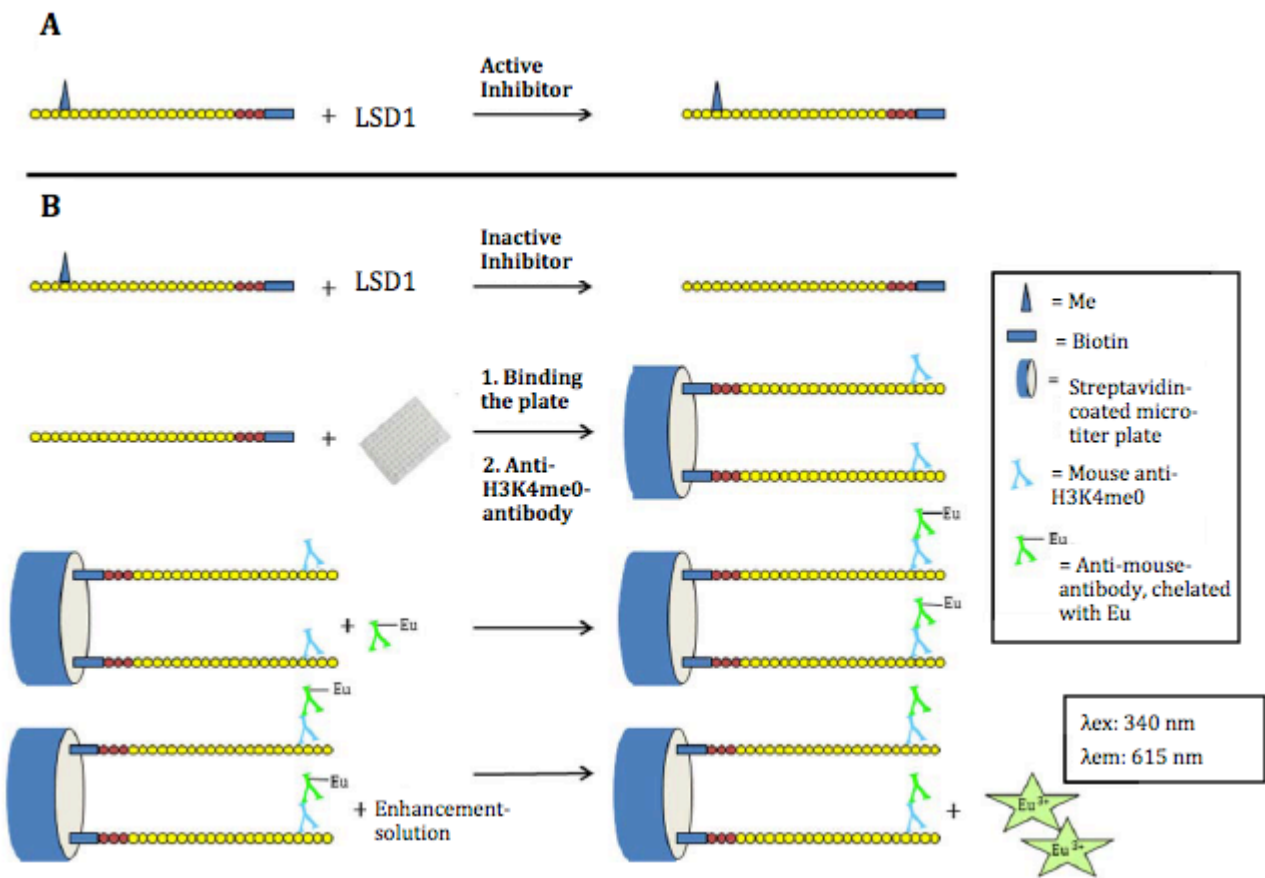


Figure 19: DELFIA assay design for a heterogeneous assay LSD1. **A:** The presence of an active inhibitor the enzyme reaction does not take place, thus there is no signal generated. **B** In case of inactive inhibitor the reaction takes place. The primary antibody recognize and bound the substrate fixed to the plate. Primary antibody F_c region is recognized by the secondary antibody and finally the europium dissolved out is detected. Picture adapted from Martin Leo Schmitt, Freiburg.

For the assay details see Appendix A.

2.6.4 Mass-spectrometry (MS) assay

This assay detects the occurrence of unmethylated H3K4 by measuring the mass of the molecule. Here the reaction mixture is embedded in a matrix. The molecules are ionized by matrix-associated laser desorption (MALDI) method and the mass of the molecules are determined repeatedly. Potent inhibitors of LSD1 result in a change of the H3K4me1/2 signals.³⁷

RESULTS AND DISCUSSION

3.1 Analysis of LSD1 crystal structures

3.1.1 LSD1 crystal structures

LSD1 X-ray crystal structure was resolved for the first time by Stavropoulos et al in 2006 (PDB: 2H94) as ‘apo’ structure.¹⁸ Eventhough the X-ray crystal structure was obtained with low resolution (2.9 Å), it was possible to carry out LSD1 structural analysis. Today 22 LSD1 crystal structures are deposited in the protein databank (table 1). According to their co-crystallized ligands, the structures might be grouped as: apo-form crystal structures (I), crystal structures binding the N-terminal H3 histone tail (II), and crystal structures binding covalent LSD1 inhibitors (III). There is no known three-dimensional structure showing LSD1 active pocket bound to non-covalent inhibitors. Here reported are the crystal structures deposited in the PDB.

Table 1: list of LSD1 crystal structures. Authors, co-crystallized structures and crystal resolutions data are also reported.

Group	PDB code	Resolution (Å)	Co-crystallized structures	References
I	2H94	2.90	//	(Stavropoulos et al.) ¹⁸ 2006
	2IW5	2.57	CoREST	(Yang et al.) ²⁴ 2006
	2HKO	2.80	//	(Chen et al.) ¹³¹ 2006
	2DW4	2.30	//	(Mimasu et al.) ¹³² 2007
II	2UXN	2.72	CoREST and 7 N-terminal H3 residues inside the active site, with K4 covalently bound to FAD cofactor.	(Yang et al) ¹³³ 2007
	2V1D	3.10	CoREST and 16 N-terminal H3 residues inside the active site, with K4 mutated into M4.	(Forneris et al) ²⁵ 2007
	2X0L	3.00	CoREST and 16 N-terminal H3 residues inside the active site, with K4 mutated into M4.	(Zibetti et al) ¹³⁴ 2010
	2Y48	3.00	CoREST and 9 N-terminal H3 residues inside the active site, with K4 mutated	(Baron R. et al) ¹³⁵ 2011

	3ZMU	3.20	into F4.	
	3ZMV	3.00	Short peptide reversible inhibitor.	(Tortorici M. et al) ¹³⁶ 2013
			Short peptide reversible inhibitor.	(Tortorici M. et al) ¹³⁶ 2013
III	2UXX	2.74	CoREST and tPCPA covalently bound to FAD cofactor.	(Yang et al) ¹³³ 2007
	2Z3Y	2.25	tPCPA covalently bound to FAD cofactor.	(Mimasu et al) ¹³² 2008
	2EJR	2.70	tPCPA covalently bound to FAD cofactor.	(Mimasu et al) ¹³² 2008
	2Z5U	2.25	tPCPA covalently bound to FAD cofactor.	(Mimasu et al) ¹³² 2008
	2XAJ	3.30	CoREST and tPCPA covalently bound to FAD cofactor.	(Binda et al) ⁴¹ 2010
	2XAS	3.20	CoREST and tPCPA analog covalently bound to FAD cofactor.	(Binda et al) ⁴¹ 2010
	2XAF	3.25	CoREST and tPCPA analog covalently bound to FAD cofactor.	(Binda et al) ⁴¹ 2010
	2XAG	3.10	CoREST and tPCPA analog covalently bound to FAD cofactor.	(Binda et al) ⁴¹ 2010
	2XAH	3.10	CoREST and tPCPA analog covalently bound to FAD cofactor.	(Binda et al) ⁴¹ 2010
	2XAQ	3.20	CoREST and tPCPA analog covalently bound to FAD cofactor.	(Binda et al) ⁴¹ 2010
	3ABT	3.20	Fluorine substituted tPCPA derivative covalently bound to the FAD cofactor.	(Mimasu et al) ⁶² 2010
	3ABU	3.10	Fluorine substituted tPCPA derivative covalently bound to the FAD cofactor.	(Mimasu et al) ⁶² 2010

3.1.2 Structural analysis

All the crystals show a specific folding, characterized by three important domains (figure 20):

1. **Swi3p; Rsc8p; Moira (SWIRM) domain:** in LSD1 the SWIRM domain, in contrast to the other DNA-binding SWIRM domains, does not show conserved residues.¹⁸ For this reason, it seems that in this case the SWIRM domain is not directly involved in DNA binding.¹⁸
2. **Amine oxidase domain (AOD),** which hosts the catalytic site. Two different lobes form the catalytic site, the first one recognizes and binds the substrate, whereas the second one binds the flavin adenine dinucleotide (FAD) cofactor.^{18,24} The lobe assigned to recognize and bind substrates

might be divided into four major invaginations: A) Catalytic Chamber: in this pocket the isoalloxazine ring of the cofactor is embedded, the involved residues are: Val317, Gly330, Ala331, Met332, Val333, Asn535, Phe538, Leu659, Asn660, Lys661, Trp695, Ser749, Ser760 and Tyr761. Available mutagenesis data showed that mutation of Asn535 and Tyr761 causes an activity reduction of LSD1.³⁰ These residues are important for positioning the cofactor and the lysine group of the substrate (N-terminal histone tail H3) in order to generate the optimal geometries for FAD nucleophilic attack.¹⁸ B) First side chain binding pocket: this site is next to the catalytic chamber and is formed by Val334, Thr335, Asn340, Met332, Tyr571, Thr810, Val811 and His812. C): Second side chain binding pocket: formed by Phe558, Glu559, Phe560, Asn8, Tyr807, Pro808.¹⁸ D): Third side chain binding pocket: formed by Asn540, Leu547, Trp552, Asp553, Gln554, Asp555, Asp556, Ser762, Tyr763, Val764 and Tyr773.¹⁸ Here mutagenesis data demonstrated that mutation of Asn540, Asp555, Asp556 and Trp552 resulted in a reduction of LSD1 activity.³⁰

3. Tower domain: this domain is formed by two long α -helices and is a sort of “binding platform” for interaction proteins such as Co-Rest (that stabilizes and enhances LSD1).^{25,24,18,26}

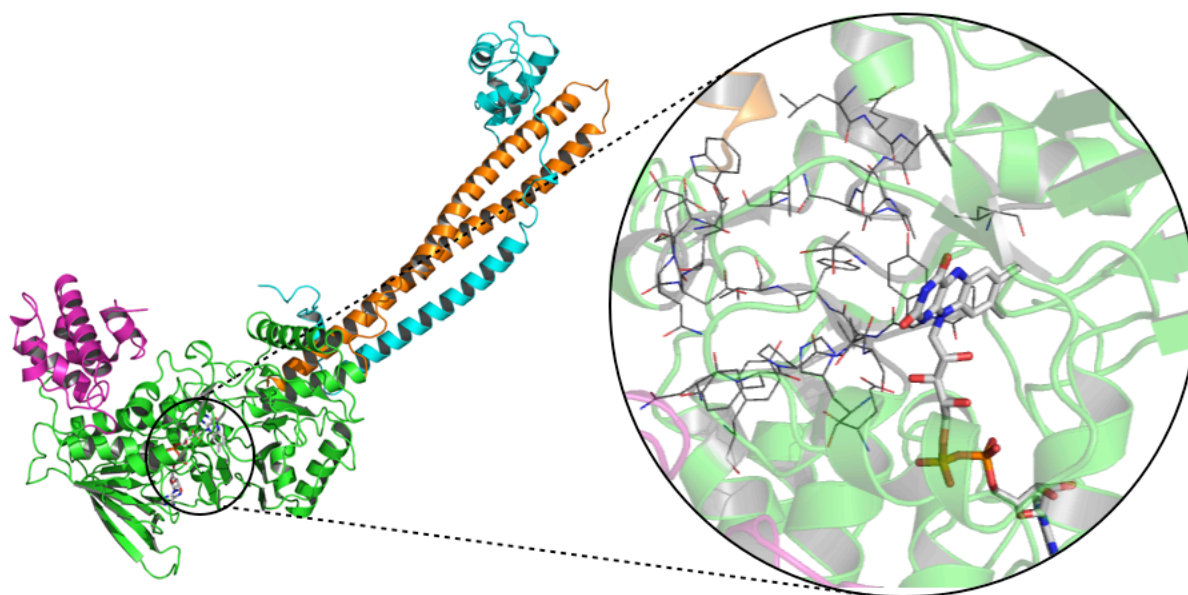


Figure 20: LSD1 crystal structure (PDB code 2IW5). The enzyme is coloured according to the domains present in the structure. SWIRM domain is coloured pink, amine oxidase domain is coloured green while the tower domain is coloured orange. In the picture is also shown the LSD1 binding site in proximity of FAD cofactor. Residues belonging to LSD1 active cavities are coloured grey, while FAD cofactor is coloured white.

From the crystal structures of LSD1 co-crystallized with histone H3 peptide (crystal structures that belong to group II in table 1), several interactions were described as important binding hotspots. In one of these peptides, encompassing the terminal 12 residues of H3, the exchange of Lys4 to a methionine resulted in a potent LSD1 inhibitor with a K_i of 0.05 μM .²⁵ In general, the X-ray structures of the peptide-LSD1 complexes showed that the peptide adopts three consecutive γ -turns

resulting in a perfect fit to the binding pocket (figure 21). The analysis further revealed that only three residues of the histone tail can be hosted in the binding pocket (figure 21). Major contributions to the peptide binding are coming from ionic interaction of H3-Arg8 with Asp375, and hydrogen-bond interaction of H3-Arg8 with Glu379 and the backbone of Cys360 as well as the interaction of H3-Arg2 with Asp553 and Asp556 of LSD1 (figure 21).^{25,30,135} Regarding the crystal structures with a covalent inhibitor the first resolved structures showed the well-known MAO tranlycypromine (PCPA) inhibitor and its analogues.^{41,61,62} Eleven crystal structures with this class of inhibitors were solved within the last few years (table 1). It was observed that PCPA, which is a suicide inhibitor, does not interact strongly with the active site of LSD1. The phenyl ring of the FAD-tranlycypromine adduct makes only weak interactions with the methyl groups of Thr335 and Thr810. Interestingly, the co-crystallization of PCPA with LSD1 resulted in two different reaction adducts. In one complex, a five-membered ring linking C-4a and N-5 of FAD is formed. The other reaction product shows a 3-phenylpropanol substituent attached to the N-5 atom of FAD.¹³² Increasing the size of the substituent at the phenyl ring of PCPA resulted in more bulky derivatives (PDB code 2XAS, figure 21) that could be co-crystallized with LSD1. The crystal structure shows that the “head” of the inhibitor is covalently attached to the FAD cofactor whereas the “tail” of the inhibitor fills the central part of the substrate-binding pocket. One of the aromatic rings present in the ligand is even facing out of the binding pocket. This shows that the fit between the inhibitor and LSD1 is far from being perfect.⁴¹ In order to fully understand the relevance of different binding modes and to design more active inhibitors, the resolution of new crystal structures with non-covalent LSD1 inhibitors is still mandatory.

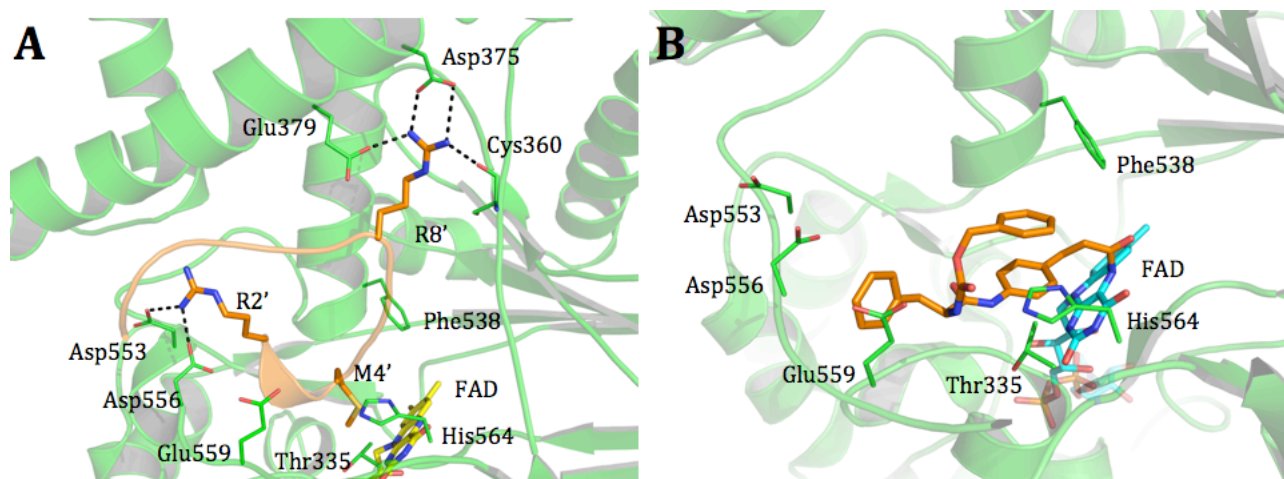


Figure 21: A) Crystal structure of LSD1-histone H3 peptide complex (PDB code 2V1D). The co-crystallized substrate analogue contains a methionine instead of the methylated Lys4 of H3. The peptide side chains are displayed as orange sticks, whereas the FAD cofactor is shown in yellow. Only relevant amino acids are displayed. Amino acids of the histone peptide are labelled using single letter code. B) Crystal structure of LSD1 in complex with a covalently bound tranlycypromine derivative (PDB code 2XAS). The inhibitor is coloured orange whereas the cofactor is coloured cyan. The protein backbone is displayed as green ribbon and only relevant amino acids are shown.

3.1.3 LSD1 similarity search

LSD1 is a flavin dependent amine oxidase enzyme and its FAD cofactor is the main actor in the active mechanism (Chapter 1). The first goal at the early stages of this work was to find sequence similarity between LSD1 and enzymes having FAD as cofactor (amine oxidase domains) in order to find differences and similarities among them. The general idea was to use the available structural and binding information of similar targets, especially those having AOD, and to apply them on LSD1 according to the paradigm “similar receptors bind similar ligands”¹³⁷.

Similarity studies taking into account only the residues belonging to the AOD were conducted by Forneris F. et al⁴⁴ based on DALI search¹³⁸ and demonstrated that LSD1 shows 26% sequence identity with maize polyamine oxidase enzyme (PAO) and yeast polyamine oxidase enzyme (FMS1). LSD1 showed also 20% sequence identity with human monoamine oxidase enzymes A and B (MAO-A and MAO-B),¹³⁸ also 19% similarity with the fungal monoamine oxidase enzyme N (MAO-N) and L-amino acid oxidase enzyme (from Bothrops Jararaca venom, LAAO).

Table 2: AOD of LSD1 similarity studies (adapted from Forneris F., et al.⁴⁴)

Protein	AOD Similarity	Biological Function
LSD1	//	Chromatin folding regulation
PAO	26%	Metabolism of polyamines
FMS1	26%	Vitamin B5 biosynthesis
MAO-A	20%	Metabolism of neurotransmitters
MAO-B	20%	Metabolism of neurotransmitters
MAO N	20%	Metabolism of amine
LAAO	20%	Amino acids metabolism

Despite similarities, LSD1 shows a binding pocket that is more than double the size observed for both MAO and PAO enzymes.⁴⁴ While the volume of MAO-B pocket is around 637 Å³ and the cavity volume of PAO is 601 Å³, the volume of LSD1 binding pocket is more than 1700 Å³.⁴⁴ This big active site renders the identification of putative binding mode of any small molecules inside LSD1 active pocket difficult.

However, several residues and a water molecule are described as conserved residues among MAO, PAO and LSD1 active sites.⁴⁴ A lysine residue (Lys661 in LSD1) is involved in the water molecule mediated binding of the FAD cofactor. This lysine residue has been demonstrated to be crucial

through mutagenesis data, causing the inactivation of LSD1 when replaced by an alanine residue.¹⁸ It seems indeed, that this residue is involved in the catalytic reaction of LSD1, activating an oxygen atom and allowing the turnover of FAD from its reduced form into the oxidised form, producing hydrogen peroxide (Chapter 1).¹³⁹ In amine oxidase domains there are two other conserved aromatic residues that together with the tricyclic isoalloxazine ring of FAD generate the well-known feature called “aromatic-cage”.¹⁴⁰ MAO-B mutagenesis data has demonstrated that these residues might have a steric role in the substrate binding, and they might be also involved in increasing the nucleophilicity of the substrate amine moiety.^{44,141} In PAO, Phe403 and Tyr439 are the residues designed for this role, while in MAO-A, MAO-B and MAO-N there are two tyrosine residues forming with FAD the aromatic cage. In LSD1 only one residue is conserved (Tyr761) and it is still considered to be involved in the recognition of methylated Lys4 amino group.⁴⁴ Meanwhile, the other residue is replaced by Thr810, causing the loss of the “aromatic-cage” feature (figure 22).

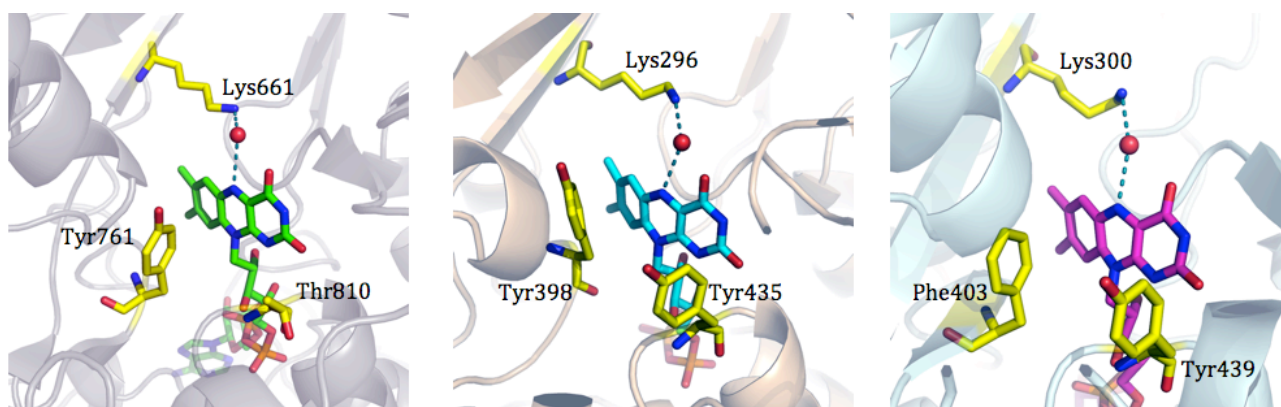


Figure 22: Active cavities of LSD1 (Left, PDB code 2DW4), MAO-B (Middle, PDB code 2C66) and PAO (Right, PDB code 1B5Q). In yellow are shown the considered conserved residues and the cofactor is represented in green, cyan and purple respectively. The conserved water molecule involved in the active mechanism is shown as a red ball, while the interactions are showed as cyan dashed lines.

3.1.4 Molecular dynamic simulations

At the beginning of the work, two 42 nanoseconds (ns) long molecular dynamic simulations (MDs) of LSD1 without substrate, were carried out using 15 Å radius of explicit solvent around the protein. The first system was set including the whole tower domain, while most of the tower domain was removed in the second system (residues were removed from Asp425 to Asp518, figure 23). These two simulations were carried out to understand the influence of the tower domain on the whole structure of LSD1.

In order to check the stability of the complexes and the differences among the systems in both MD simulations, system root mean square deviation (RMSD) values were calculated, compared to their respective initial structures. According to the obtained RMSD values, the protein carrying the TD appears to be much more flexible, due to the TD oscillations, than the system where the two alpha helices were removed. Visual inspection and RMSD calculations, considering only the residues involved in the AOD of LSD1, showed no grave structural differences between LSD1 with TD and LSD1 without TD during the MDs. The RMSD fluctuations' average, comparing LSD1 without and with TD, jumps from 1.7 Å to 2.7 Å respectively (table 3). On the other hand there is a huge gap between the simulations with respect to the computational time. The 42 ns MD simulation using LSD1 structure without TD and with 15 Å radius explicit solvent around the protein requires almost 92 hours using a GeForce GTX 580 with 3GB memory and 1.68 GHz frequency. Considering the same work conditions, LSD1 with TD requires almost 263 hours to reach the same MD length (table 3). In order to decrease the computational time, LSD1 protein was set without TD in all further molecular dynamics simulations.

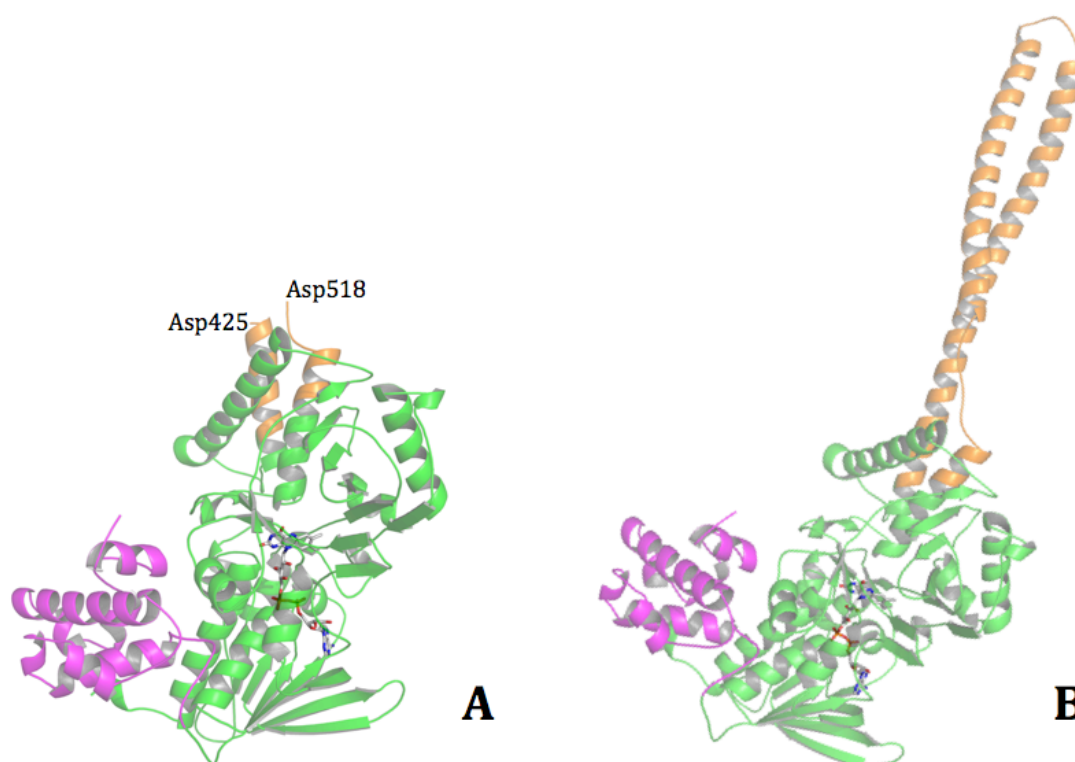


Figure 23: LSD1 crystal structure (PDB code 2V1D) used during the preliminary dynamics. A) LSD1 crystal structure without the TD. B) LSD1 crystal structure with the TD.

Table 3: 42 ns MD comparison of LSD1 set without and with the TD.

	Number of Atoms	RMSD AOD (Å)	Standard Deviation AOD (Å)	Computational Time (hours)
LSD1 without TD (A)	86522	1.693	0.224	~92
LSD1 with TD (B)	268726	2.667	0.508	~263

In order to analyse the stability of LSD1 with its natural substrate and evaluate the main interactions between LSD1 active cavity and the N-terminal histone number three tail, several exhaustive molecular dynamics simulations were carried out using the crystal structure of LSD1 co-crystallized with the histone H3 peptide (15 residues) bound to enzyme's active site (PDB code 2V1D). To assess the stability of the complexes in every MD simulation, systems root mean square deviation (RMSD) values were calculated, compared to their respective initial structures. The conserved water molecule interacting with FAD and Lys661 (figure 22) was extracted from 2DW4 crystal structure and added to the models.

All simulations were done (PDB code 2V1D) without the TD (from Asp525 to Asp518 residues) allowing fast and relatively “light” simulations from the computational point of view. The mutated methionine number four present in the co-crystallized N-terminal histone H3 tail was unbound from the FAD cofactor, and mutated into a lysine residue (LSD1 natural substrate). Finally, four different MD simulations were carried out, according to the different methylation levels found at Lys4' residue on histone tail number 3 (Lys4', Lys4'-Me1, Lys4'-Me2, Lys4'-Me3).

3.1.4.1 MD simulation of LSD1 complexed with the N-terminal histone tail and unmethylated Lys4'

The first simulation was done with unmethylated Lys4' side-chain. The RMSD plot shows that the protein reaches an equilibrium state after 5 nanoseconds (ns) and it oscillates around 2Å for the length of the MD (figure 24). Concerning the co-crystallized histone tail, the RMSD plot shows high RMSD value during the whole length of MD, due to the H3 tail C-terminal residues, which are partially exposed to the solvent. An RMSD plot calculated excluding the last three H3 tail C-terminal residue, shows lower oscillations. The H3 tail reaches the steady state after 5 ns of simulation and it oscillates around 2Å (figure 24).

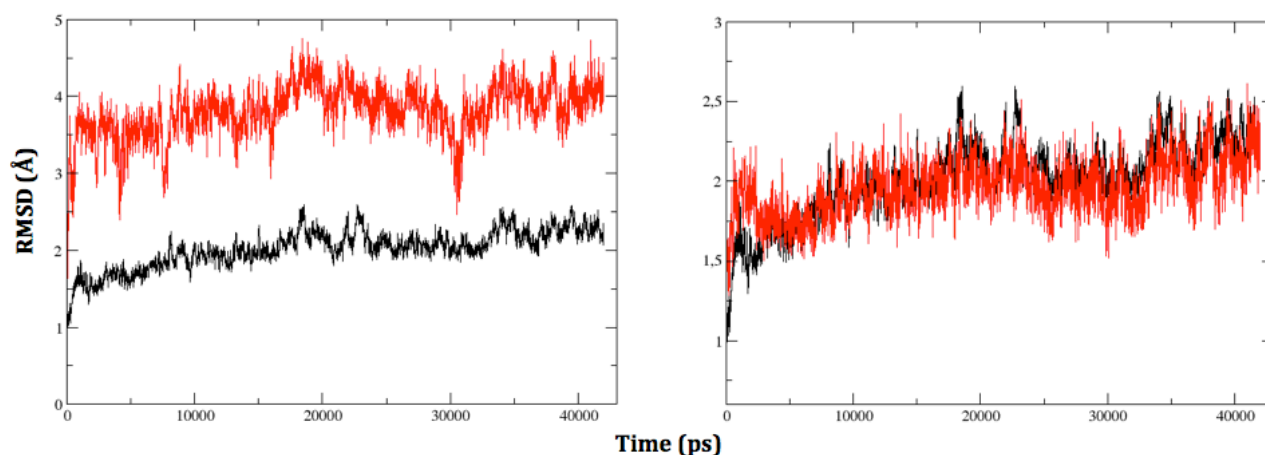


Figure 24: Root main square deviation (RMSD) plots representing LSD1 protein without TD (black trace) and H3 tail (red trace) during the MD with unmethylated H3-Lys4'. The plot on the right shows the RMSD values of H3 tail without considering the last three histone tail C-terminal residues oscillations.

Table 4: RMSD values of the most stable residues belonging to the H3 tail co-crystallized in the 2V1D crystal structure.

Structures	RMSD average (Å)	S.D. (Å)	RMSD Min. (Å)	RMSD Max. (Å)
FAD	1.11	0.16	0.44	1.61
H3-Ala1'	1.82	0.43	0.32	3.23
H3-Arg2'	2.06	0.28	1.01	3,17
H3-Thr3'	2.02	0.29	0.99	3.15
H3-Thr6'	1.34	0.36	0.26	2.60
H3-Lys4'	1.24	0.23	0.37	2.11
H3-Arg8'	1.98	0.51	0.43	4.80
H3-Lys9'	2.48	0.35	0.57	3.55

MD visual inspection revealed that LSD1 and the N-terminal H3 tail are very stable. RMSD values of the most involved residues of the histone tail calculated with respect to the initial structure were carried out (table 4). A hydrogen bond analysis on the trajectories obtained during the MD was also generated (table 5). A cut off distance of up to 3Å among the H-bond donor and the acceptor groups was used, while the bond angle cut off was set to 120° (default values from AMBER package⁸³).

The N-terminal residue H3-Ala1' undergoes hydrogen bond interactions with Asn540 side-chain of LSD1 active pocket. H3-Arg2' residue shows interactions with Asp556 side-chain and to a lesser extent Asp553 side-chain of LSD1. H3-Thr3' side-chain forms hydrogen bond interaction with Asp555 side-chain during the entire MD simulation. Unmethylated Lys4' forms hydrogen bond interaction with the backbone carboxyl group of Ala809 residue, while H3-Arg8' side-chain interacts through ionic interaction with Asp375 side-chain, and also through hydrogen bond

interactions with Glu379 side-chain and the backbone of Cys360 residue (figure 25). This arginine residue H3-Arg8' is strongly fixed in a highly negative electrostatic region of the LSD1 active cavity.

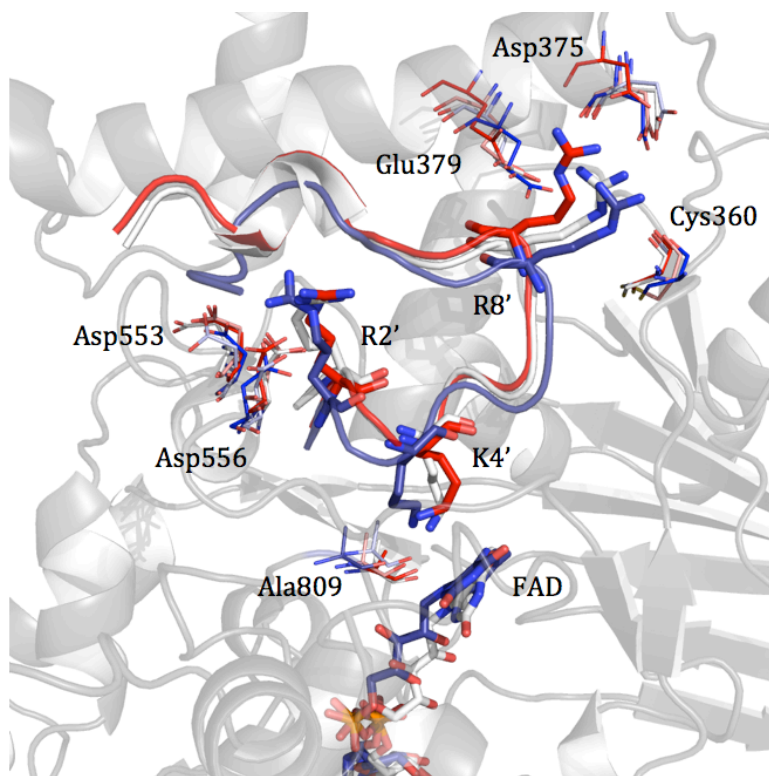


Figure 25: Three frames from the MD run on unmethylated H3-Lys4' N-terminal tail. The first frame (taken after the equilibration step) is coloured red, the second frame is taken after 16ns and is coloured white, while the third frame is taken after 40ns and is coloured blue. Only the initial LSD1 structure is shown and coloured grey. The most important residues of the tail and the enzyme are shown and coloured according to the respective MD frame. Amino acids belonging to the histone peptide are labelled using single letter code.

Table 5: Hydrogen bonds occupancy (%) for the histone peptide residues involved in binding with LSD1 active cavity during the MD setting with unmethylated H3-Lys4'.

Donor	Acceptor	Occupied	Distance
Asn540-OD1	Ala1'-N	97.14%	2.8
Asp556-OD1	Arg2'-NH1	94.76%	2.8
Asp556-OD2	Arg2'-NH1	34.29%	3.0
Asp555-OD2	Thr3'-OG1	98.88%	2.7
Ala809-O	Lys4'-NZ	94.43%	2.9
Asp375-OD1	Arg8'-NH1	52.44%	2.9
Asp375-OD1	Arg8'-NH2	54.93%	2.9
Asp375-OD2	Arg8'-NH1	56.42%	2.9
Asp375-OD2	Arg8'-NH2	40.08%	3.0

Glu379-OE1	Arg8'-NH2	38.9%	2.8
Cys360-O	Arg8'-NH1	34.23%	2.9
Cys360-O	Arg8'-NH2	26.46%	2.9
Cys360-O	Arg8'-NE	48.51%	2.9

3.1.4.2 MD simulation of LSD1 complexed with the N-terminal histone tail and mono-methylated Lys4'

The second molecular dynamic simulation was done with monomethylated H3-Lys4'. The RMSD plot (figure 26 on the right side) shows that both protein and histone tail reach the equilibrium after 5 ns of MD, as the previous run. Once the equilibrium is attained, LSD1 structure oscillates at around 2 Å for the whole MD. Also here large RMSD fluctuations of the histone peptide are due to the last three C-terminal histone tail residues that are partially exposed to the solvent (figure 26).

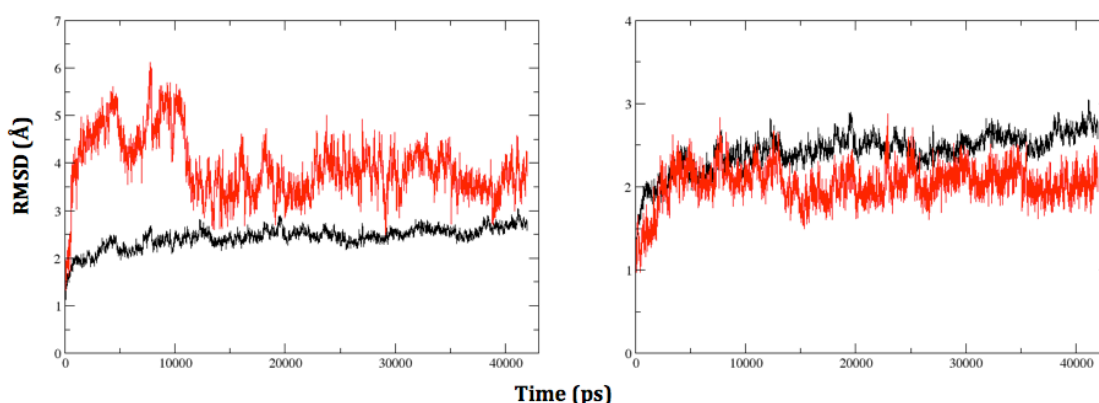


Figure 26: LSD1 (black) and histone tail (red) RMSD plot, setting H3-Lys4' as mono-methylated. On the right the RMSD is generated without the last three C-terminal histone tail residues.

From the visual inspection more interactions than the previous MD simulation were noticed. Using the same setting as previously described, a hydrogen bond analysis was carried out (table 6).

H3-Ala1' forms a hydrogen bond interaction with the amide group of Gln554, besides the hydrogen bond interaction with Asn540 side-chain described before. H3-Arg2' side-chain and H3-Thr3' side-chain undergo the same interactions previously described, while H3-Lys4' still forms hydrogen bond interaction with the backbone of Ala809 residue despite the presence of a methyl group on the nitrogen atom of Lys4'. H3-Arg8' side-chain as in the previous case, is well stabilized by Cys360, Asp375 and Glu379 amino acids. A new interaction was found: H3-Lys9' side-chain interacts through hydrogen bond interaction with Glu559 side-chain due to the movements of the entire H3 peptide (table 6).

Table 6: New hydrogen bonds occupancy (%) for the most involved histone peptide residues with LSD1 active cavity during the MD simulation of monomethylated H3-Lys4⁷.

Donor	Acceptor	Occupied	Distance
Asn540-OD1	Ala1 ⁷ -N	50.29%	2.8
Gln554-OE1	Ala1 ⁷ -N	85.36%	2.9
Asp555-OD2	Thr3 ⁷ -OG1	99.28%	2.7
Asp555-OD1	Thr3 ⁷ -OG1	35.85%	3.0
Glu559-OE1	Lys9 ⁷ -NZ	28.63%	2.9
Glu559-OE2	Lys9 ⁷ -NZ	28.21%	2.9

3.1.4.3 MD simulation of LSD1 complexed with the N-terminal histone tail and di-methylated Lys4⁷

In the third molecular dynamic simulation dimethylated H3-Lys4⁷ was studied. Like in the previous run, LSD1 reaches the equilibrium after 5 ns, meanwhile the histone peptide reaches the steady state only after 9 ns. Both maintain the equilibrium for the rest of the MD with an RMSD around 2.5 Å for LSD1 and 2 Å for the histone tail (figure 27).

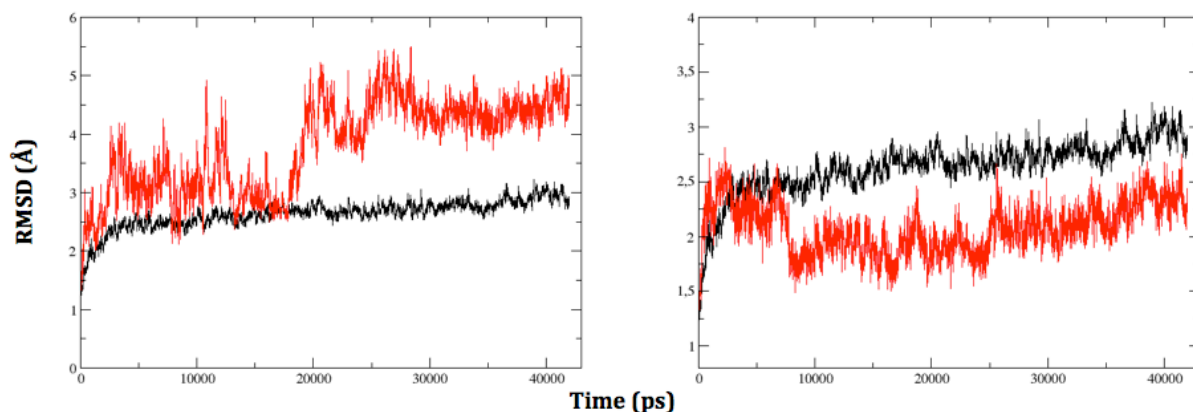


Figure 27: LSD1 (black) and histone tail (red) RMSD plot, setting H3-Lys4⁷ as di-methylated. On the right the RMSD is generated without the last three C-terminal histone tail residues.

The RMSD plots show that both the protein and the histone peptide are less stable compared to the previous runs. However, visual inspection and hydrogen bond analysis using the same setting described before showed that the interactions visualized in the previous two MDs are conserved. Furthermore, new interactions between H3 tail and LSD1 active cavity were found (table 7). H3-Lys4⁷ side-chain forms interactions with the backbone of Ala809 and the side-chain of Thr810. H3-Thr6⁷ side-chain undergoes during this simulation interactions with the imidazole group of

His564 residue, and H3-Lys9' side-chain strongly interacts through hydrogen bond with Glu559 side-chain (table 7).

Table 7: New hydrogen bonds occupancy (%) for the most involved histone peptide residues with LSD1 active cavity during the MD, setting H3-Lys4' side-chain as dimethylated.

Donor	Acceptor	Occupied	Distance
Ala809-O	Lys4'-NZ	67.18%	3.0
Thr810-OG1	Lys4'-NZ	45.64%	3.0
His564-ND1	Thr6'-OG1	67.49%	3.0
Glu559-OE2	Lys9'-NZ	64.34%	2.9
Glu559-OE1	Lys9'-NZ	50.14%	2.9

3.1.4.4 MD simulation of LSD1 complexed with the N-terminal histone tail and tri-methylated Lys4'

The last MD simulation was set with trimethylated H3-Lys4'. The RMSD plot clearly showed that the histone peptide reaches the steady state after 10 ns MD. Once the equilibrium is reached, the RMSD shows high oscillations around 4 Å. Beyond 30 ns the histone peptide RMSD jumps from 4 Å to 6 Å (figure 28), indicating larger conformational changes of the histone tail.

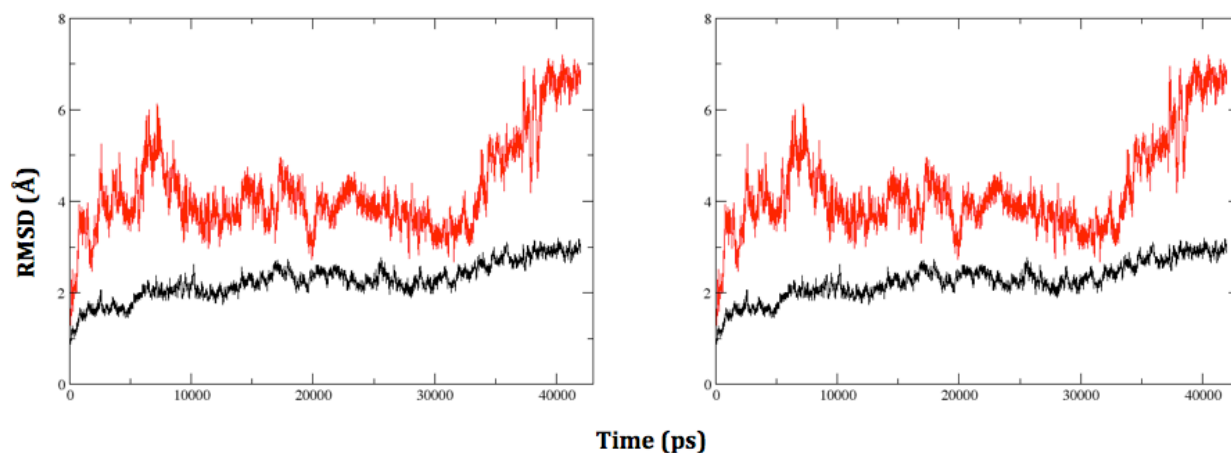


Figure 28: LSD1 (black) and histone tail (red) RMSD plot, with trimethylated H3-Lys4'. On the right the RMSD is generated without the last three C-terminal histone tail residues.

Through visual inspection it was possible to explain the interactions of the histone tail. The side-chain of the trimethylated H3-Lys4' does not have enough space inside the catalytic chamber of LSD1, due to the bulky methyl groups covalently bound to the quaternary ammonium. Lys4' projects its side-chain above the isoalloxazine ring of FAD causing the movement of the whole

histone tail backbone. Even if the interactions previously described are more or less conserved during the simulation, high differences occur regarding the amino acids belonging to the C-terminal histone tail. Here the residues that normally interact on LSD1 active cavity surface cannot reach the equilibrium, pulling the tail out of the LSD1 binding pocket, which is in agreement with the experimental data, which show no LSD1 activity on trimethylated H3-K4.^{142,22}

3.1.5 Discussion

Despite the sequence similarity with MAO-A/B (20%) and PAO (26%) enzymes, LSD1 shows a binding pocket that is more than double the size of the other enzymes.⁴⁴ Deep inside the active cavity in proximity of the FAD cofactor's isoalloxazine ring there are three residues that are considered conserved among these structures.⁴⁴ A lysine residue is involved in the catalytic mechanism of the enzyme,¹³⁹ and two aromatic residues form together with the isoalloxazine ring of FAD the so called "aromatic-cage".¹⁴⁰ LSD1 only partially displays the conserved aromatic feature, with a tyrosine residue (Tyr761) involved in the substrate recognition,⁴⁴ while the second aromatic residue is replaced by threonine (Thr810) in proximity of the isoalloxazine ring of the cofactor (figure 22).

LSD1 crystal structure (PDB code 2V1D) was used to explore the interactions of the histone tail through multiple MD simulations. Four different methylation states of H3-Lys4' side-chain (H3-K4'; H3-K4Me; H3-K4'Me2; H3-K4'Me3) were studied as LSD1 natural substrates. As shown by the performed MDs, the stability of the histone tail inside the active cavity of LSD1 is dependent on the degree of methylation on the Lys4' side-chain.

Hydrogen bond analysis and exhaustive MDs visual inspection showed conserved interactions, between H3 tail and AOD of LSD1.

The interactions found during the MD simulations, and the mutagenesis data available in literature^{24,30} confirmed the presence of some key-residues inside AOD of LSD1, which are crucial for the activity of the enzyme. The found key residues such as Cys360, Asp375, Glu379, Asn540, Gln554, Asp555, Asp556, His564, Tyr761, Ala809, and Thr810, might be helpful to further analyse and explain the putative binding modes of small LSD1 inhibitors.

3.2 Propargylamine derivatives

In 2006 Culhane et al.¹⁴³ discovered an oligopeptide by combining the inhibitory propargylamine group known from MAO inhibitors with the LSD1 substrate histone H3, which appeared to be a covalent modifier and thus irreversible inhibitor of LSD1 (figure 29). The found oligopeptide was used as a mechanistic tool in biochemical studies and is unlikely to have potential for drug development. In order to investigate the biological consequences of irreversible inhibition of LSD1, several inhibitors with different warheads (cyclopropylamines vs. propargylamines) were tested during this work.¹⁴⁴

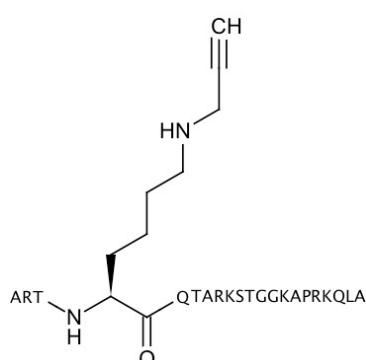
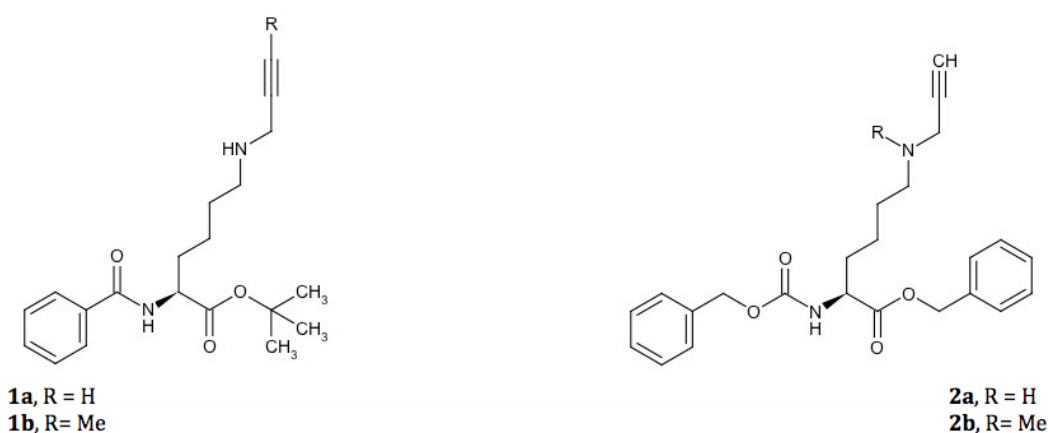


Figure 29: First oligopeptide as irreversible inhibitor of LSD1

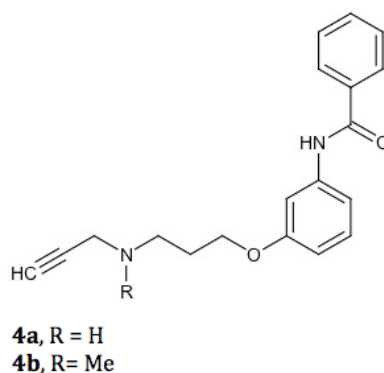
Synthesis of lysine-mimicking small molecules, carrying the propargyl warhead and resembling the LSD1 natural substrate (H3K4), was the preliminary step in order to generate new LSD1 inhibitors. Four compounds (figure 30) were first synthesised and tested in a biochemical *in vitro* peroxidase assay (Chapter 2).



Compound	IC ₅₀ ± SE [μM] or Inhib. @ conc. [μM]
Peroxidase assay	
1a	143.6 ± 16.1
1b	131.9 ± 15.7
2a	26% @ 100 μM
2b	42% @ 100 μM

Figure 30: Lysine-mimicking small molecules as LSD1 inhibitors and their biological activity in a biochemical *in-vitro* peroxidase assay.

After generating the first inhibitors which showed a potency around 100 μM (figure 30), new lysine-mimicking benzamide and anilide derivatives (**4a** and **4b** figure 31) were generated by substituting the amino acid core of compounds **2a** and **2b** by aromatic rings. The LSD1 inhibition *in vitro* peroxidase assay showed that inhibitors of the anilide series were active in the micromolar range (figure 31).



Compounds	IC ₅₀ ± SE [μM]
Peroxidase assay	
4a	184.2 ± 16
4b	93.1 ± 12.8

Figure 31: Anilide derivatives as LSD1 inhibitors and their biological activity in a biochemical *in-vitro* peroxidase assay.

3.2.1 Docking studies

The molecular structures of all the compounds analysed in the present study were generated using the MOE 2011.10 modeling software.¹⁴⁵ Initial ligand conformations were obtained from an energy minimization using the MMFF94x force field as implemented in MOE. The available crystal structures of LSD1 (Chapter 3.1) were used. For the subsequent docking studies, all water and ligand molecules were removed and all structures were protonated and minimized using the Amber99 force field. Docking studies were carried out to choose appropriate cap groups for lysine derivatives. Initially, the anilide derivatives (**4a**, **4b**) and the lysine-mimicking derivatives (**2a**, **2b**) were docked into the LSD1 substrate pocket. In the most favoured docking pose Tyr761 forms a hydrogen bond with the amine belonging to *N*-propargylamine warhead and the side chain of Asp555 interacts through a hydrogen bond with the amide nitrogen atom of the benzamide moiety (figure 32 and 33). The aromatic substituents form extensive hydrophobic and T-shaped aromatic interactions with Phe558, Phe560, Tyr807 and His812. The docking protocol was set keeping the warhead moiety (*N*-propargylamine group) constrained to the N5 nitrogen atom of the tricyclic isoalloxazine ring of FAD, according to the hypothesis of a covalent inhibitory mechanism.^{142,19} We used the Gold 4.1⁸⁷ docking program, GoldScore, and two LSD1 crystal structures (PDB code 2XAS⁴¹ (figure 32 A-B) and 2DW4¹³² (figure 33 A-B)).

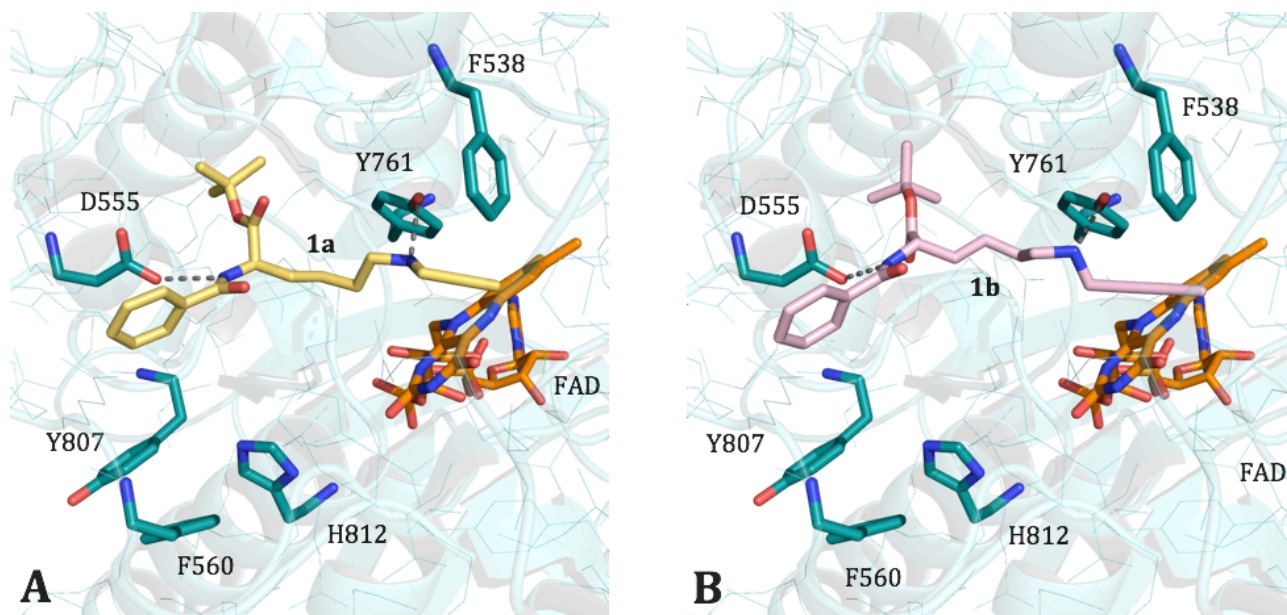


Figure 32: Most favourable docking solutions of lysine-mimicking small molecules inside LSD1 crystal structure (PDB code: 2XAS). The ligands are represented as yellow sticks (**1a**) and pink sticks (**1b**); respectively. FAD cofactor is shown as orange sticks and the most involved residues of LSD1 are shown as cyan sticks. Hydrogen atoms are omitted and hydrogen bond interactions are shown as grey dashed lines.

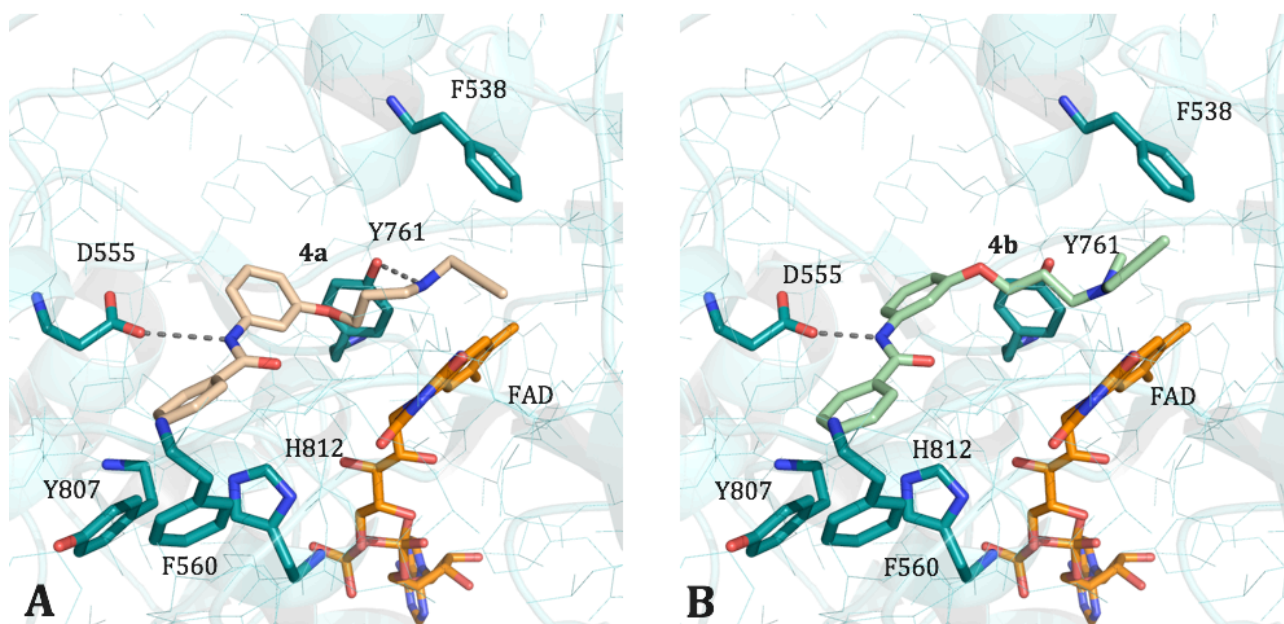


Figure 33: Most favourable docking solutions of the anilide derivatives inside LSD1 crystal structure (PDB code: 2DW4). The ligands are represented as light orange sticks (**4a**) and light green sticks (**4b**); respectively. FAD cofactor is shown as orange sticks and the most involved residues of LSD1 are shown as cyan sticks. Hydrogen atoms are omitted and hydrogen bond interactions are shown as grey dashed lines.

3.2.2 Virtual screening

In order to identify compounds for further *in vitro* testing, a substructure based virtual screening was conducted. The Enamine database comprising more than 750 000 compounds was chosen for this purpose and stored as an MOE database.¹⁴⁵ Since the small molecule inhibitors of LSD1 comprised tranylcypromine and analogues,^{41,55,61,62} two different moieties were selected for substructure search query: the *N*-propargylamine and the tranylcypromine moiety (both cases with the protonated nitrogen atom). Four compounds were identified (figure 34) and purchased from Enamine supplier but only two compounds containing the propargylamine group showed an inhibitory activity on LSD1 enzyme (**T5342129** (**5a**) and **T5342128**).

Results and Discussion: Propargylamine derivatives

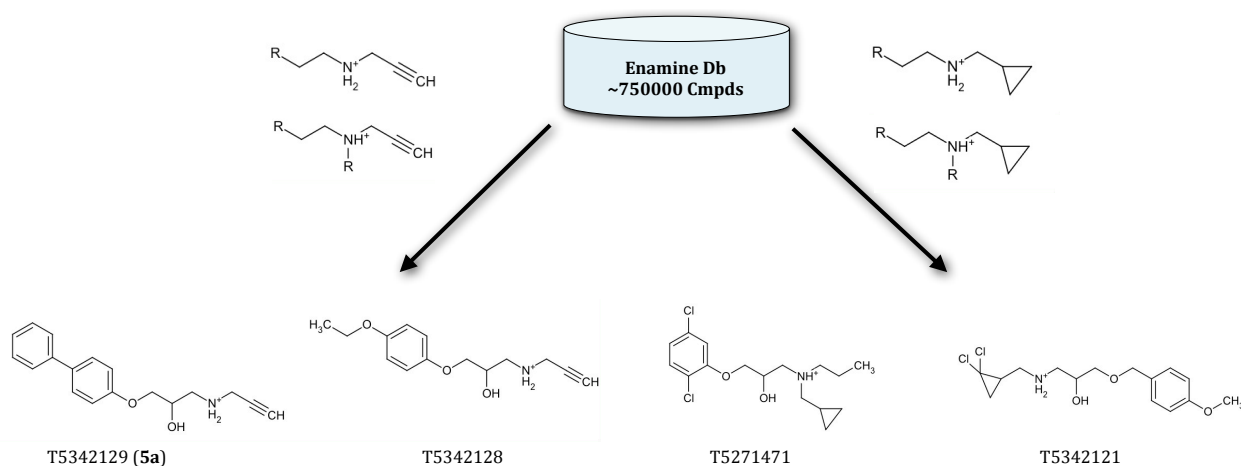


Figure 34: Substructure search protocol using the Enamine database and query compounds

The 3-aryloxy-2-hydroxy-propylamine **5a** with an *N*-propargyl group was identified as an LSD1 inhibitor with a potency of 44.0 μM (figure 35) in a peroxidase assay and 34.4 μM in a formaldehyde dehydrogenase assay (Chapter 2, data of FDH assay not shown).¹⁴⁴

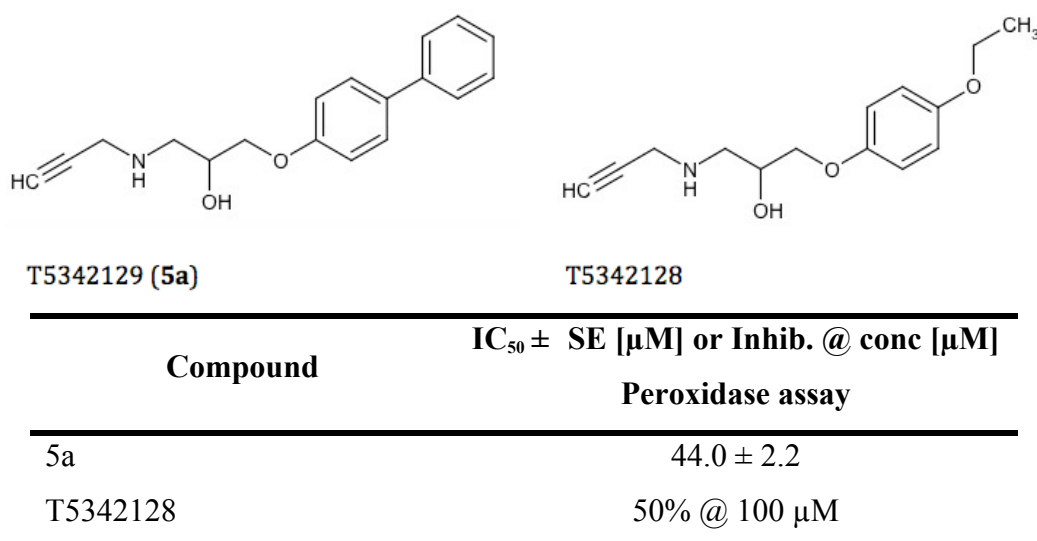


Figure 35: Active compounds obtained from virtual screening procedure, and their biological activity in a biochemical *in-vitro* peroxidase assay.

The docking results showed that the *N*-propargylamine group of **5a** is located near the reactive N5-nitrogen atom of the FAD cofactor, according to the covalent inhibitor hypothesis (figure 36). The hydroxyl group and the amine form hydrogen bonds with the backbone of Ala809 and the side-chain of Tyr761; respectively. The hydrophobic biaryl group interacts via van der Waals interactions with Phe560 and Tyr807.

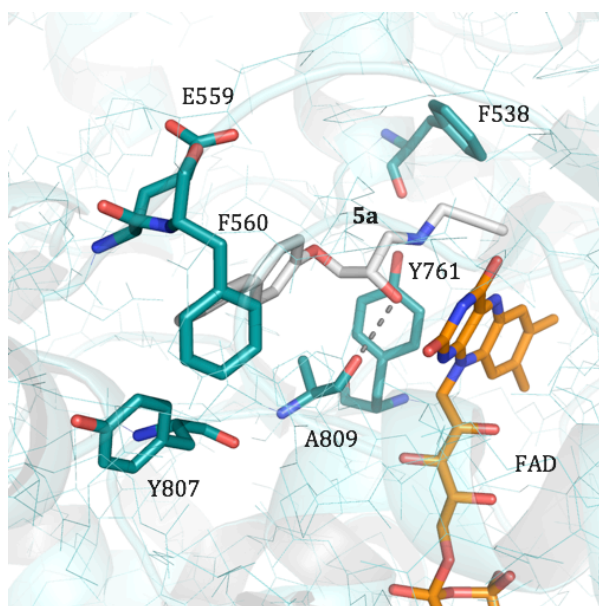


Figure 36: Most favourable docking solution of the biphenyl derivative **5a** obtained through virtual screening inside LSD1 crystal structure (PDB code: 2XAJ). The ligand is represented as white sticks. FAD cofactor is shown as orange sticks and the most involved residues of LSD1 are shown as cyan sticks. Hydrogen atoms are omitted while hydrogen bond interactions are shown as grey dashed lines.

3.2.3 New propargylamine derivatives

Using MOE suite,¹⁴⁵ interaction possibilities inside the LSD1 active cavity were identified by calculating the hydrophobic contact preferences (figure 37). This step allowed understanding the preferred position of the hydrophobic ligand atoms using the 3D coordinates of LSD1 (figure 37). Through the visual inspection of the found binding mode it was possible to identify a further sub-pocket formed by Phe560, Tyr807 and His812, able to host a bulkier lipophilic group than a biphenyl moiety (figure 37).

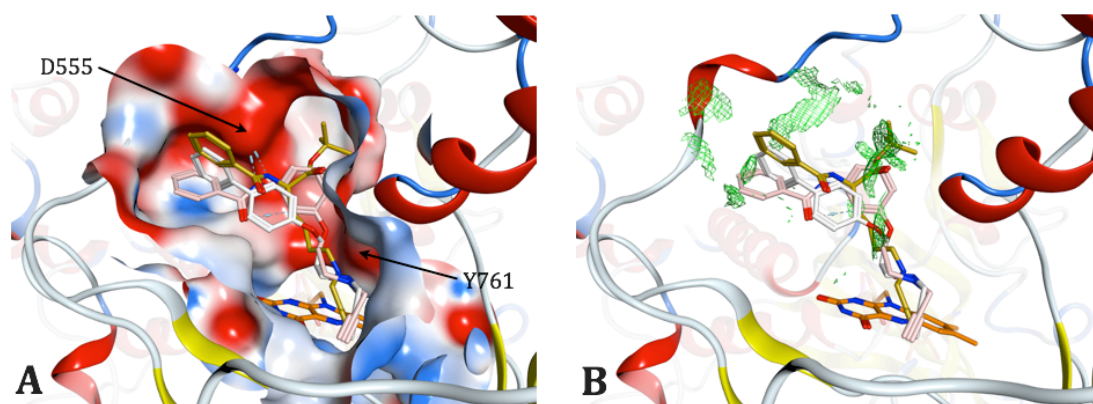
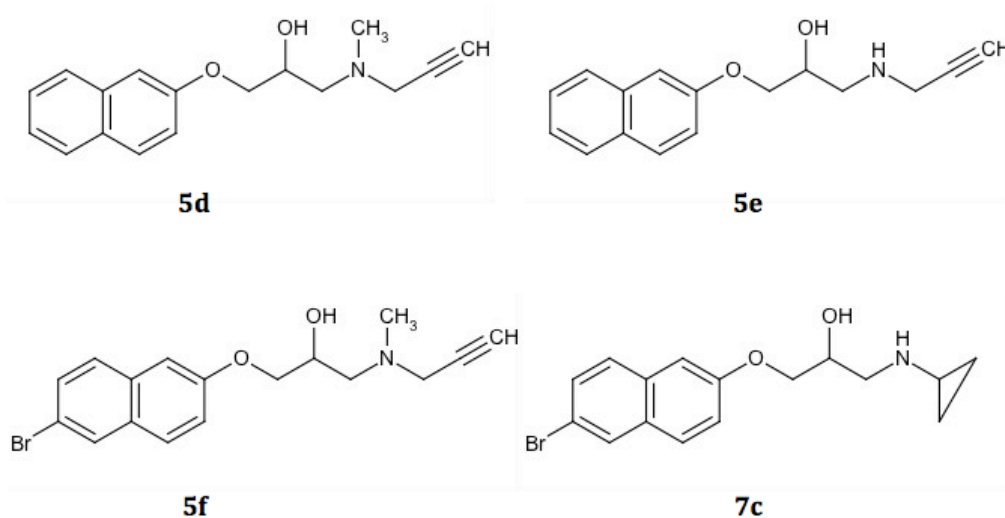


Figure 37: A: Docking solutions of compound **4b** and **5a** inside LSD1 (PDB code 2DW4). The active pocket is shown as surface and coloured according the electrostatic potential (blue = positive potential, white = neutral potential, red = negative potential). B: Interaction possibilities at the binding pocket obtained by calculating the contact preference using the MOE software. The *N*-propargylamine atoms were not considered in the calculation. The favourable hydrophobic interaction preferences are colored green. Hydrogen atoms are omitted in both pictures.

In order to increase the LSD1 inhibition, four different series of compounds with bulky lipophilic aryl groups were synthesized as suggested by docking studies. The synthesis focused on annellated aryl and biaryl structures: 1-aryloxy-3-(prop-2-ynylamino)-propan-2-ols (**5a-5r**, for structures see Appendix B table 1),¹⁴⁴ aza-analogues (**6a-6b**, for structures see Appendix B table 1)¹⁴⁴ and cyclopropylamine warhead (from tranylcypromine) analogues (**7a-7d**, for structures see appendix B table 1)¹⁴⁴. The most potent inhibitor, containing a 6-bromo-2-naphthyl-substituent and *N*-methylation was found (**5f**) with a potency of 22.0 μM when tested in a peroxide assay (figure 38). Two potent inhibitors belonging to different series (**4b** and **5a**) were also tested in a cellular assay, showing that the inhibitor is affecting the cells and increasing the levels of H3K4me2 (western blot analysis).¹⁴⁴



Compound	IC ₅₀ ± SE [μM] Peroxidase assay
5d	51.7 ± 2.4
5e	92.2 ± 2.3
5f	22.2 ± 2.1
7c	111.4 ± 15.5

Figure 38: Naphthyl derivatives as LSD1 inhibitor and their biological activity in a biochemical *in-vitro* hydrogen peroxide dependent assay.

The synthesized compounds (**5b-5r**, **6a-6b** and **7a-7d**) were also docked to LSD1 using the same docking settings as previously described. The putative binding mode found for the most active compounds (**5d**, **5e**, **5f** and **7c**) shows the *N*-propargylamine group (or *N*-cyclopropylgroup for **7c**) located near the active nitrogen of the cofactor (N5) while the naphthyl moiety is located in a

hydrophobic cavity (figure 39). The hydroxyl and the amine groups of the compounds undergo hydrogen bond interactions with Ala809 and Tyr761; respectively. Meanwhile, the naphthyl rings shows van der Waals interactions with residues Glu559, Phe560 and Tyr807. The bromonaphthyl moiety of **5f** fills the sub-pocket more deeply resulting in the best binding interaction and biological activity.

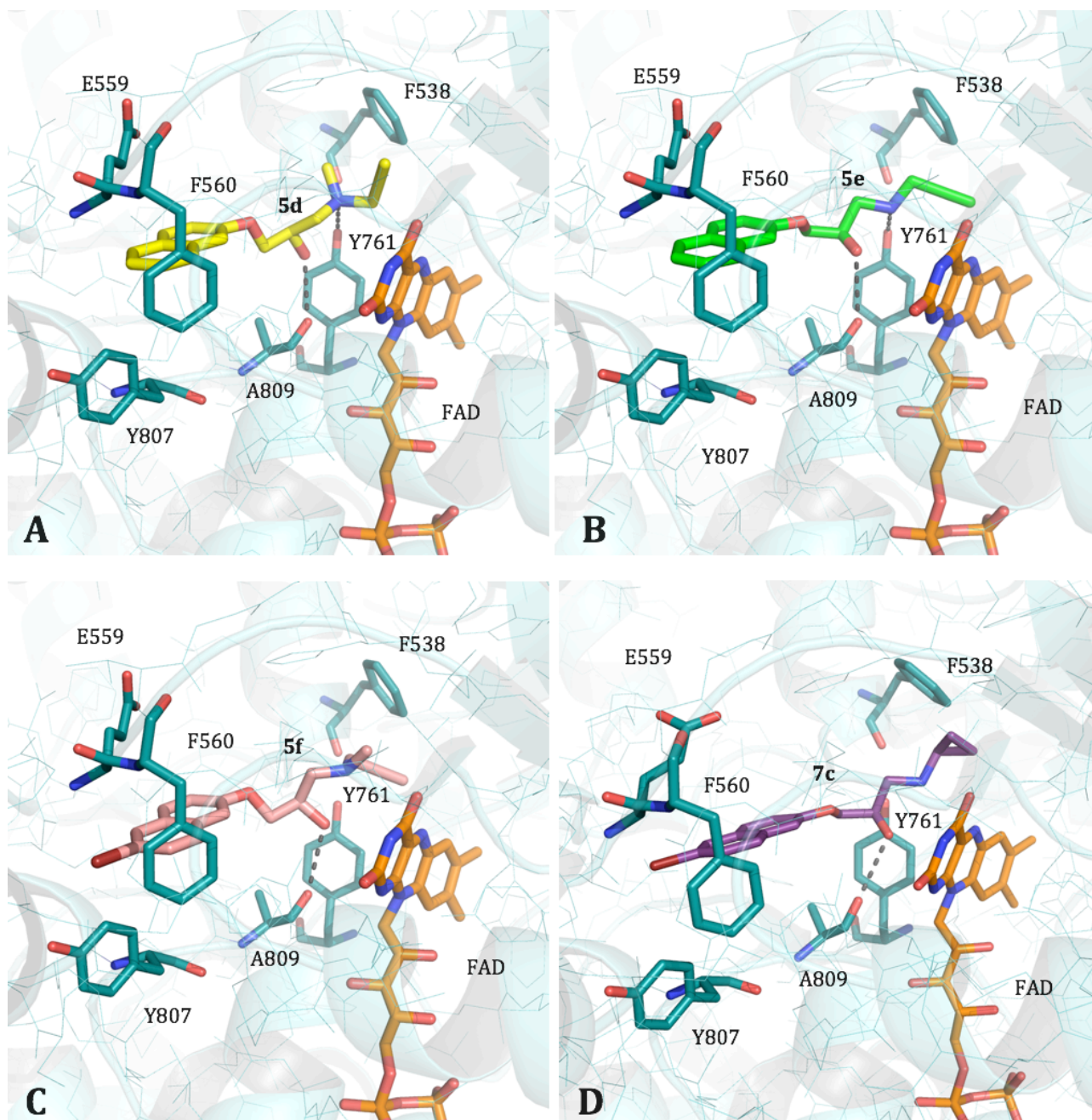


Figure 39: Most favourable docking solution of the naphthyl derivatives inside LSD1 crystal structure. A: **5d** inside 2UXN LSD1 crystal structure. B: **5e** inside 2UXN LSD1 crystal structure. C: **5f** inside 2XAJ LSD1 crystal structure. D: **7c** inside 2XAJ LSD1 crystal structure. Ligands are represented as yellow, green, pink and purple sticks; respectively, while the FAD cofactor is shown as orange sticks. The most involved residues of LSD1 are shown as cyan sticks. Hydrogen atoms are omitted while hydrogen bond interactions are shown as grey dashed lines.

3.2.4 Discussion

Synthesis of lysine-mimicking small molecules instead of oligopeptides, allowed developing new LSD1 covalent inhibitors active in the micromolar range in *in vitro* peroxidase assay.¹⁴⁶

The docking studies allocate the benzamide moiety of compounds **1a** and **1b** in the so-called “second side-chain binding pocket” (Chapter 3.1), and the tert-butyloxy group inside the “third side-chain binding pocket” (Chapter 3.1). This resulted in an important hydrogen bond interaction between the benzamide nitrogen atom and the negatively charged region (figure 32A) well-known as an important “hotspot” for the stabilization of H3K4 tail inside LSD1 pocket (Chapter 3.1). Compared to the first synthesized ligand (**1a**, **1b**), replacement of the amino acid core by an anilide moiety (**4a** and **4b**) did not decrease the activity, when tested for *in vitro* LSD1 inhibition. According to the binding mode found, compound **4a** and **4b** do not interact inside the “third side chain binding pocket” as compound **1a** or **1b**, but they interact using the anilide nitrogen atom with the electrostatically negative region present in the binding cavity (figure 37A).

Structure-based virtual screening procedure using the Enamine database permitted the selection of two new active compounds (T5342129 (**5a**) and **T5342128**). The derivatives selected (figures 36) occupy the same region inside LSD1 binding pocket when compared to the previous docking studies. In contrast to the previously found binding mode, compound **5a** does not form any hydrogen bond with Asp555, and it does not show any lipophilic moiety interacting inside the “third side chain binding pocket”. However, the ligand shows an IC₅₀ value around 44 μM, which renders it the second most active LSD1 inhibitor of the studied dataset. Its activity is due to the presence of a hydroxyl group (generating two different enantiomers not separated during the biological testing) that interacts with the backbone of Ala809 (figure 36). This new interaction might allow a better positioning of the *N*-propargylamine group in order to reach the best geometry for the nucleophilic attack on the isoalloxazine (N5) of the cofactor.

New compounds were synthesized considering the hydroxyl feature present in ligand **5a** as an important hydrogen bond donor and taking into account the possibility to better fill the “second side chain binding pocket” in order to increase the shape complementarity. The compound bearing a 6-bromo-2-naphthyl group and an *N*-propargylamine moiety (**5f**) allocates the bromine atom deep inside the “second side chain binding pocket” and proved to be the most active LSD1 inhibitor belonging to the propargylamine dataset.

3.3 Basic polyamino and guanidino derivatives

The aim of this part of the work was to find new LSD1 inhibitors through exploiting the similarities of lysine demethylase enzyme with polyamine oxidase enzymes (PAOs).

3.3.1 Docking studies

Several reported guanidino^{65,70,71,72} and cyclic guanidino¹⁴⁷ ZmPAO inhibitors were used for docking studies on LSD1.

All the derivatives used in this study were generated and minimized (MMFF94x force field) using the MOE 2011.10 modelling software.¹⁴⁵ The guanidine groups and the amine nitrogen atoms were protonated according to their pK_b values, while all the crystal structures used were protonated with a specific module called “protonation 3D” implemented in MOE suite, and subsequently minimized using Amber99 force field.

Guazatine (compound **17**, figure 42) was co-crystallized with ZmPAO (PDB code 1H82¹⁴⁸) and for this reason it was used as reference structure to derive the best docking setup able to reproduce the same binding mode of guazatine inside PAO enzyme. The docking setup obtained during the validation step was then used for the docking of the polyamine ligands.

ZmPAO crystal structure shows specific interactions between guazatine and PAO '30Å U-shaped' binding pocket (figure 40 and 44).⁶⁸ The flexibility of the co-crystallized ligand allows a perfect fit with the residues of the binding pocket (figure 44). The large entrance of the cavity shows a wide negative electrostatic potential formed by charged amino acids (Asp117, Glu120, Glu121, Glu124, Asp194 and Asp195)⁶⁸ (figure 40). The guanidine groups present in guazatine interact with specific residues inside PAO's tunnel. One guanidine moiety interacts through a cation- π interaction with the sidechain of Phe189 located close to the main entrance of the pocket, while the second guanidine moiety interacts through hydrogen bond with Asn405 and Asn437, which are located at the opposite region of the tunnel (figure 40). The central amine moiety of the ligand is positioned close to the cofactor and might interact through a hydrogen bond with the side-chain of Tyr298 (figure 40). The considered conserved residues Tyr439 and Phe403 (Chapter 3.1) forming the “aromatic cage” located in the inner region of the ZmPAO cavity accommodate the aliphatic side-chain of the guazatine ligand (figure 44).

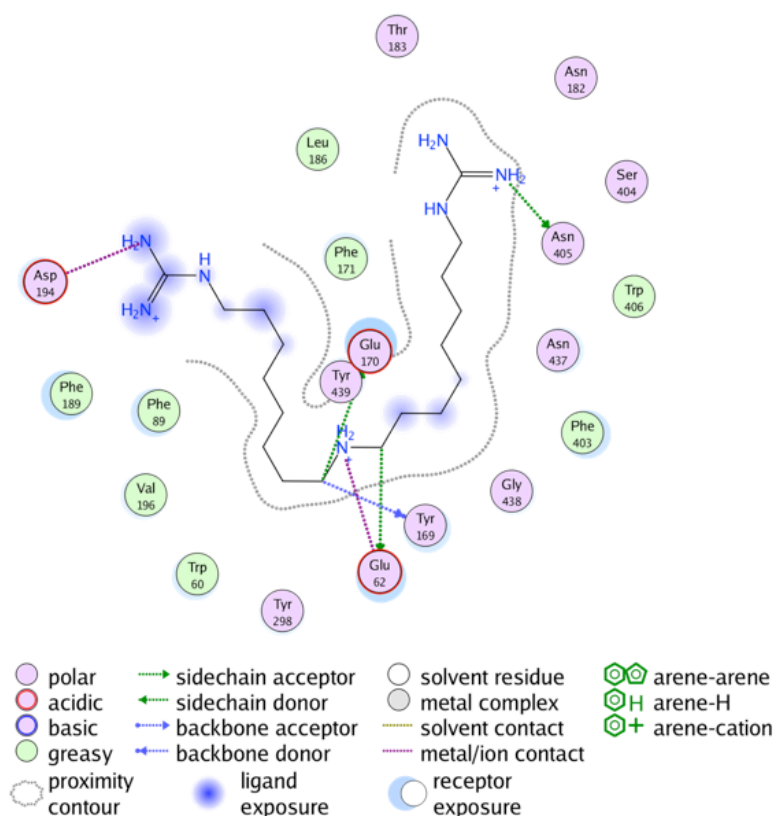


Figure 40: 2D representations of guazatine inside ZmPAO crystal structure (PDB code 1H82).

Gold 4.1⁸⁷ (goldscore as scoring function) and Glide⁹⁰ (glidescore as scoring function) were validated using guazatine co-crystallized with 1H82 ZmPAO crystal structure. Both docking programs could allocate the ligand in the same position as shown in the original crystal structure, obtaining a root-mean-square-deviation (RMSD) value of 1.25Å with Gold and 1.35Å using Glide (table 8), when compared to the original crystal structure.

Table 8: Scoring values and RMSD values of the docking softwares used during the validation step.

Compound	Glide (GlideScore)	RMSD Glide (Å)	Gold (GoldScore)	RMSD Gold (Å)
Guazatine	-4.65	1.35	98.33	1.25

After the validation step using PAO crystal structure, a docking study, using 24 guanidino⁶⁵ and 5 cyclic guanidino¹⁴⁷ PAO inhibitors found in literature¹⁴⁷, was carried out using nine LSD1 crystal structures (PDB codes: 2DW4, 2EJR, 2H94, 2HKO, 2UXN, 2V1D, 2Z3Y, and 2Z5U). The setting of Gold (goldscore) and Glide (glidescore) obtained during the validation step was used for the LSD1 docking studies. The compounds were now selected according to their scoring values and

their capability to reproduce inside the pocket of LSD1, the same binding mode shown by guazatine inside the ZmPAO crystal structure (figure 41).

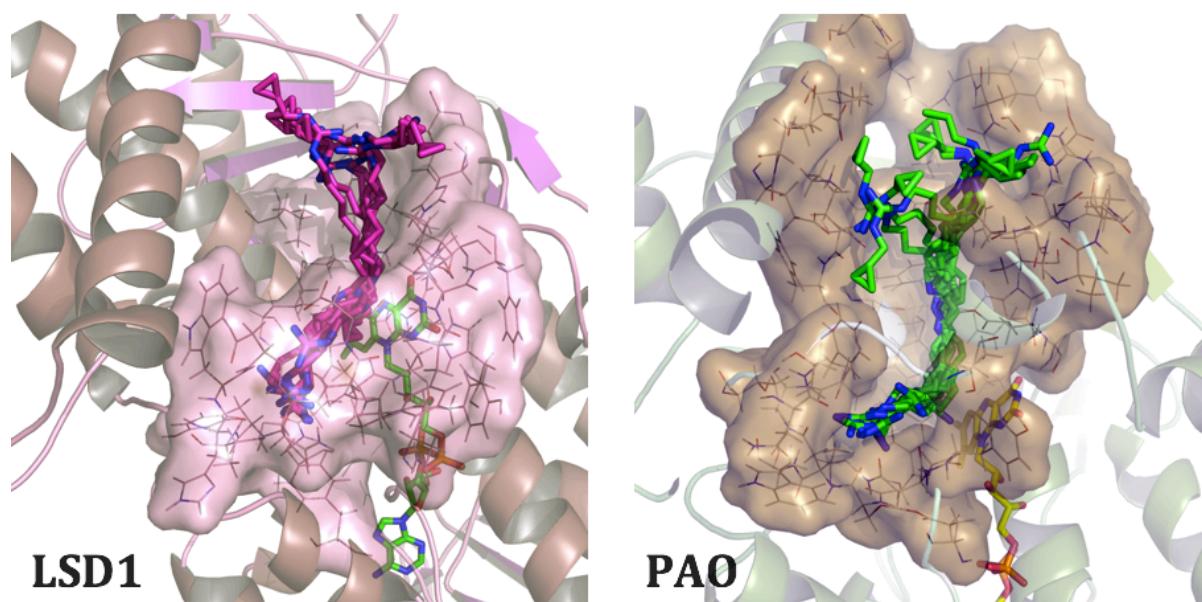


Figure 41: Docking results of ligand 19 for LSD1 and ZmPAO crystal. The top ranked docking poses for the inhibitor are displayed showing that the compound can adopt different conformation. Ligand **19** occupies the same region of the active cavity inside LSD1 and PAO. (PDB code of LSD1: 2UXN; PDB code of PAO: 1H82).

Five compounds were selected and tested in an *in vitro* peroxidase assay using LSD1 as enzyme and H3K4me2 (aa 1-20) as natural substrate (figure 42). Three of them showed an inhibitory activity (compound **17**, **23**, and **22**, table 9), while, one of them showed an inhibitory potency in the low micromolar range (compound **19**, table 9).

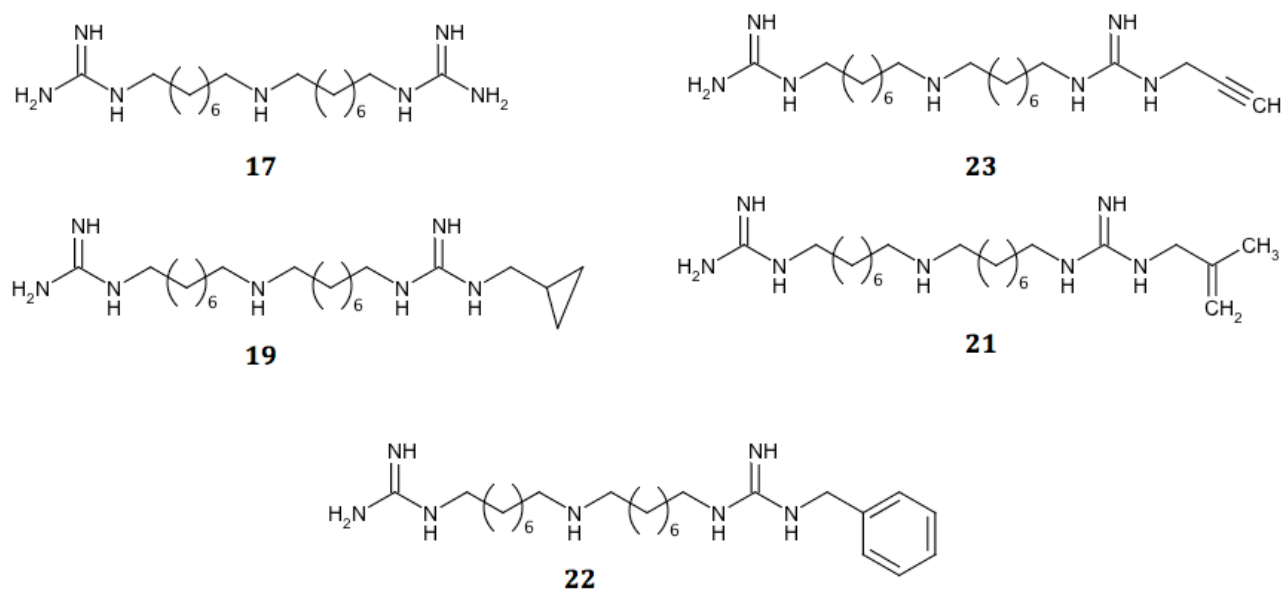


Figure 42: Compounds selected after the preliminary docking studies as potentially new inhibitors of LSD1.

Results and Discussion: Basic polyamino and guanidino derivatives

Table 9: ZmPAO binding constant and LSD1 inhibition values for the five guanidine derivatives selected.

Compound	ZmPAO K_i [nM]	LSD1 $IC_{50} \pm SE$ [μ M]	
		Peroxidase assay	Peroxidase assay
17	7.5		39% @ 25.5
19	0.08	12.3 \pm 2.2	
21	0.5	Inactive	Inactive
22	1.0		29% @ 28.3
23	0.7		73% @ 28.8

For LSD1 two putative binding modes were found for compound **19** (figure 43), **21**, **22**, and compound **23**. However, the scoring functions used were not able to rank the compounds according to their experimental activity at LSD1 (table 10).

Table 10: Docking studies results reveals two putative bindings. Scoring values are reported for every active inhibitor and for both putative binding modes. According to the scoring values it was not possible to decide which is the most probable binding mode.

Compounds	LSD1 (2DW4)		PAO (1H82)	
	Glidescore	Goldscore	Glidescore	Goldscore
17	-6.26	72.35	-4.65	98.33
19 BM1	Not Found	75.00	Not Found	99.82
19 BM2	-4.00	66.88	//	//
21 BM1	Not Found	76.41	-5.61	95.52
21 BM2	-2.83	Not found	//	//
22 BM1	-4.73	87.50	Not Found	99.69
22 BM2	-4.35	61.94	//	//
23 BM1	-4.93	75.39	-5.82	94.06
23 BM2	-4.02	64.93	//	//

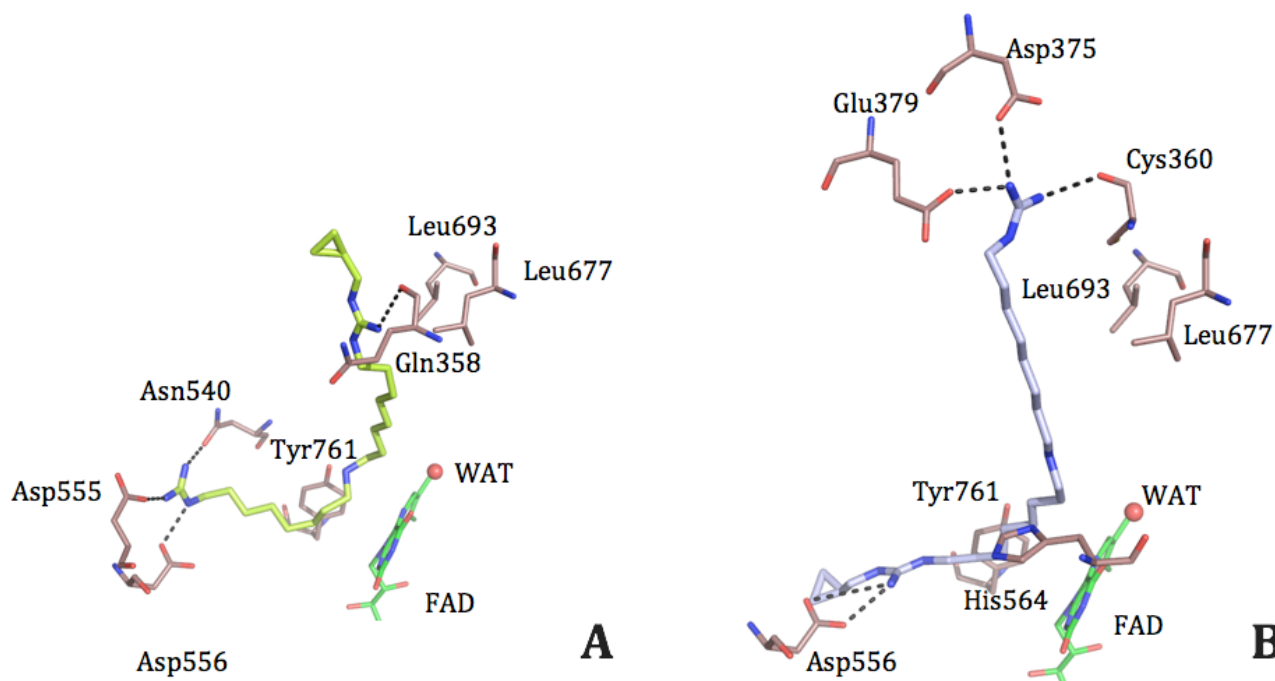


Figure 43: Binding mode hypothesis of the most active LSD1 inhibitor picked from the PAO inhibitor datasets present in literature. A: binding mode hypothesis one, ligand **19** is shown as yellow sticks, while FAD is shown as green sticks and the most involved residues of LSD1 are represented as sticks and coloured brown. The water molecule is shown as a red ball and H-bond interactions between ligand and LSD1 active cavity are shown as black dashed lines. B: binding mode hypothesis two, ligand **19** is shown as light blue sticks, FAD and the most involved residues of LSD1 are shown as green and brown sticks; respectively. The water molecule is shown as a red ball while the interactions involved in the binding mode are shown as black dashed lines.

The hypothesized binding modes of the docked compounds show the same ligand-geometries that have been reported for the co-crystallized ligand guazatine in ZmPAO (figures 41 and 43). The main difference between the two proposed binding modes is the location of the terminal moiety (for instance cyclopropane for compound **19**, figure 43).

The first proposed binding mode shows the terminal moiety exposed to the surface of the cavity and the guanidine group is interacting through a hydrogen bond with the carbonyl of Gln358 (figure 43A). The second guanidine group is interacting with the negatively charged Asp555 and Asp556 side-chains, and through a hydrogen bond with Asn540 side-chain located deep inside in the active cavity. Whereas the second binding mode is showing the terminal moiety deeper inside in the cavity, more precisely in the “third side chain binding pocket” (Chapter 3.1). Here the positive charge of the guanidine moiety is stabilized by the negative charge of Asp555 side-chain, while the second guanidine group interacts with Glu379 and Asp375 side chains and with the backbone of Cys360 residue through salt interactions and hydrogen bonds; respectively (figure 43B).

Binda and co-workers¹⁴⁸ demonstrated that some of the basic poly-amino compounds (such as CHENS pm, figure 44) are able to covalently bind to the isoalloxazine ring of FAD in PAO (figure 44). We did not exclude the possibility that all these poly-amino and guanidino derivatives found to

be active on LSD1 might show an irreversible binding mode according to the reaction leading to the formation of PAO-CHENS pm covalent adduct proposed by Binda et al (figure 45).¹⁴⁸

Based on the docking poses and their corresponding scores, no final decision can be made which binding hypothesis is more likely. Further biochemical studies are necessary to derive the real mode of action of this class of compounds.

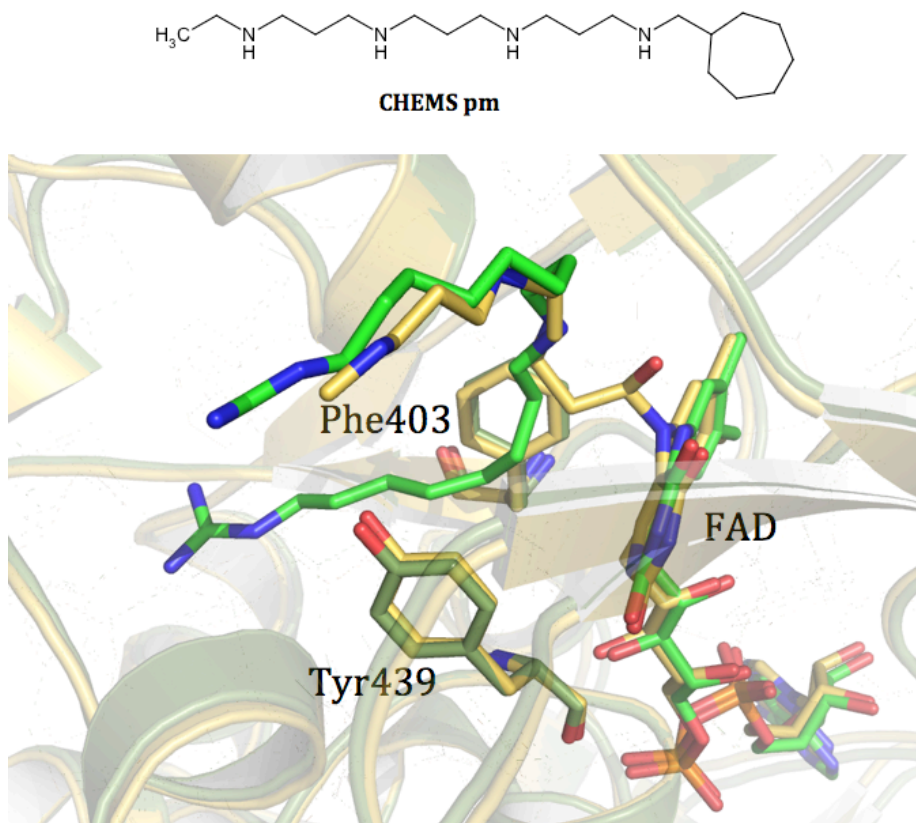


Figure 44: Top: CHEMS pm 2D structure. Bottom: PAO crystal structure superimposition (PDB code 1H82 and 1H84). Crystal structure 1H82 is dark green coloured, the considered conserved residues are shown as dark green sticks and named. Guazatine and FAD are shown as green sticks. Crystal structure 1H84 is coloured yellow and the considered conserved residues are shown as yellow sticks and labelled. CHEMS pm covalently bound to the FAD cofactor is shown as yellow sticks.

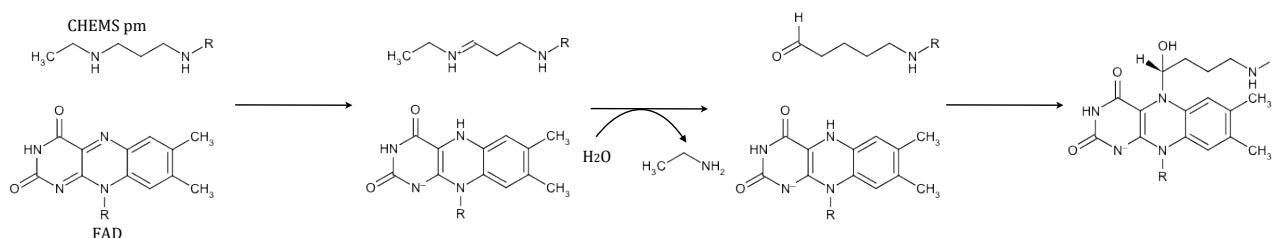


Figure 45: Proposed reaction for the formation of PAO-CHEMSpm covalent adduct. Picture adapted from Binda et al.¹⁴⁸

3.3.2 Molecular Dynamics simulations

To fully analyse the interactions of the inhibitors inside LSD1, two molecular dynamic simulations (22 nanoseconds) of the most active compound (**19**) were performed (Appendix A). In both simulations the tower domain of LSD1 enzyme was not included (Chapter 3.1).

To assess the stability of the LSD1-polyamine complexes and evaluate the differences between the two binding hypotheses, the root mean square deviation (RMSD) values of the ligands, the proteins and the cofactors were generated and plotted compared to their initial structures respectively (figures 46 and 47).

Concerning the first hypothesized binding mode, the ligand shows high flexibility during the simulation. The RMSD value fluctuates around ~ 5 Å in the first 10 ns of the MD afterwards it decreases to ~ 4.5 Å staying stable for the rest of the MD (figure 46). The RMSD values for the protein and the FAD cofactor are stable around ~ 2 Å and ~ 0.8 Å respectively, for the whole simulation (figure 46). The high RMSD value of the ligand is due to the rearrangement of the polyamine inside the cavity of LSD1. The terminal moiety with the cyclopropane loses the interaction with the backbone of Gln358 residue and the whole ligand moves more deeply inside the cavity (figure 46). While the second guanidine group keeps the interaction with Asp555 and Asp556 side-chain as found in the docking studies (figure 46).

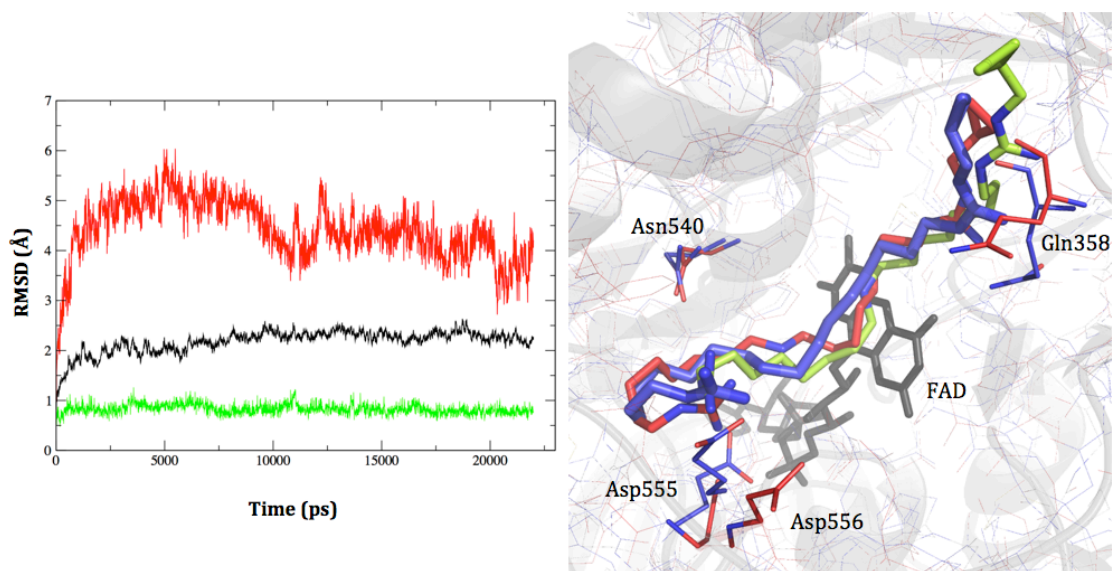


Figure 46: MD analysis of the first binding mode. The left side of the picture shows the RMSD value of the ligand (red) the protein (black) and the cofactor (green) during the simulation. On the right side of the pictures are superimposed three snapshots of the MD. Ligand **19** before the MD simulation is represented as yellow sticks. Compound **19** after 800ps and 20ns is shown as red and blue sticks; respectively. The most involved residues of LSD1 at 800ps and 20ns are shown as red and blue sticks; respectively. FAD cofactor is represented as one single structure and coloured black.

The second binding mode hypothesis still shows high ligand flexibility but the ligand is more stable inside the LSD1 cavity. The RMSD value increases to ~ 2.5 Å, remaining stable for the whole MD (figure 47). The RMSD values of the enzyme and the FAD cofactor are showing the same deviation as the previous simulation. The smaller RMSD value of the ligand is due to the stabilization of the LSD1-polyamine complex. The terminal moiety with the cyclopropane group is interacting for the whole MD with the “third side-chain binding pocket” (Chapter 3.1) with the guanidine group that is stabilized by the negatively charged Asp556 side chain. The second guanidine group interacts for the entire simulation with Asp375, Glu379 side chain and Cys360 backbone through salt interactions and hydrogen bonds, which is in accordance with the binding mode hypothesis (figure 47).

Considering the higher stability showed by the second proposed binding mode during the molecular dynamic, due to the lack of structure rearrangement, we argue that the second binding mode is the putative binding mode of compound **19** (and most probably its derivatives) inside LSD1 crystal structure.

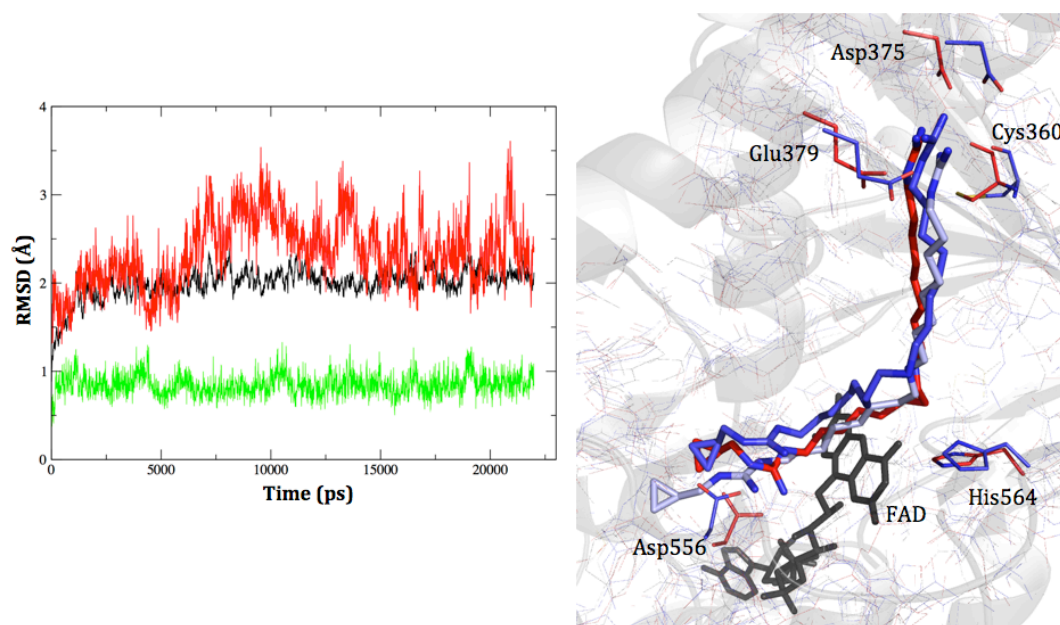


Figure 47: MD analysis of the second binding mode. The left side of the picture shows the RMSD value of the ligand (red) the protein (black) and the cofactor (green) during the simulation. On the right side of the pictures are superimposed three snapshots of the MD. Ligand **19** before the MD simulation is represented as light blue sticks. Compound **19** after 800ps and 20ns is shown as red and blue sticks; respectively. The most involved residues of LSD1 at 800ps and 20ns are shown as red and blue sticks; respectively. FAD cofactor is represented as one single structure and coloured black.

3.3.3 Multi-Step target-based virtual screening

Both proposed binding modes showed one guanidine moiety well positioned nearby the charged residues (Asp556, Asp555). This observation was used as hypothesis for a virtual screening taking into account the position of this specific feature (figure 49).

The multi-step virtual screening procedure was carried out using the ZINC Drug-Like database¹⁴⁹ with more than two million compounds. The first step was a database pre-filtering where only the compounds having a guanidine group were selected for the docking step. This allowed reducing the number of the compounds from 2184729 to 2660, considerably reducing the computational time required for the docking runs. The docking step was conducted on four minimized crystal structures of LSD1 (PDB codes: 2H94, 2UXN, 2V1D, 2Z3Y) showing an average RMSD value among them of 10.74 Å (all backbone atoms, figure 48).

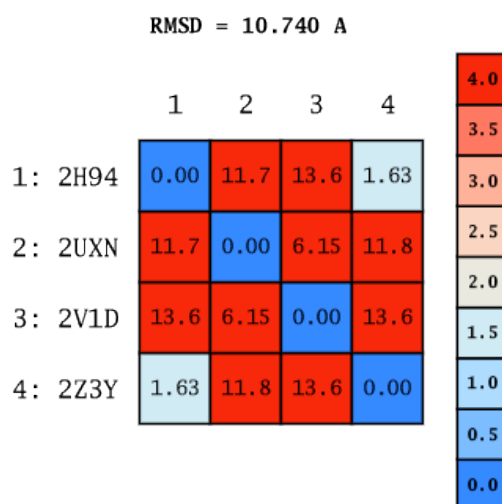


Figure 48: RMSD matrix of the LSD1 crystal structures used for the docking step during the stepwise virtual screening procedure. The values are calculated with MOE software after the optimization of the structures.

Gold and Glide were the selected docking softwares while goldscore, glidescore and chemscore were the scoring functions used in order to rank the docked compounds. For every crystal structure and for every scoring function fifty top ranked compounds were selected. 150 compounds were selected for each LSD1 crystal structure. Only the same compounds picked up for each crystal structure and for each scoring functions were chosen for the next step. 26 compounds were obtained (12 from goldscore, 7 from glidescore and 7 from chemscore). For the selected compounds physico-chemical properties were calculated: lipophilicity (logP), exposed surface to the solvent (TPSA), and lead-like properties. Compounds violating the Lipinski rules (Chapter 2) were discarded. The last step was a visual inspection of the compounds selected inside LSD1 pocket. 12 compounds were

Results and Discussion: Basic polyamino and guanidino derivatives

selected (figure 50) and only three of them were commercially available (**ZINC01607827**, **ZINC02149680**, and **ZINC01730680**). At the end none of the compounds showed any activity on LSD1 in the peroxidase assay (Chapter 2).

DRUG LIKE ZINC Database: 2184729 compounds

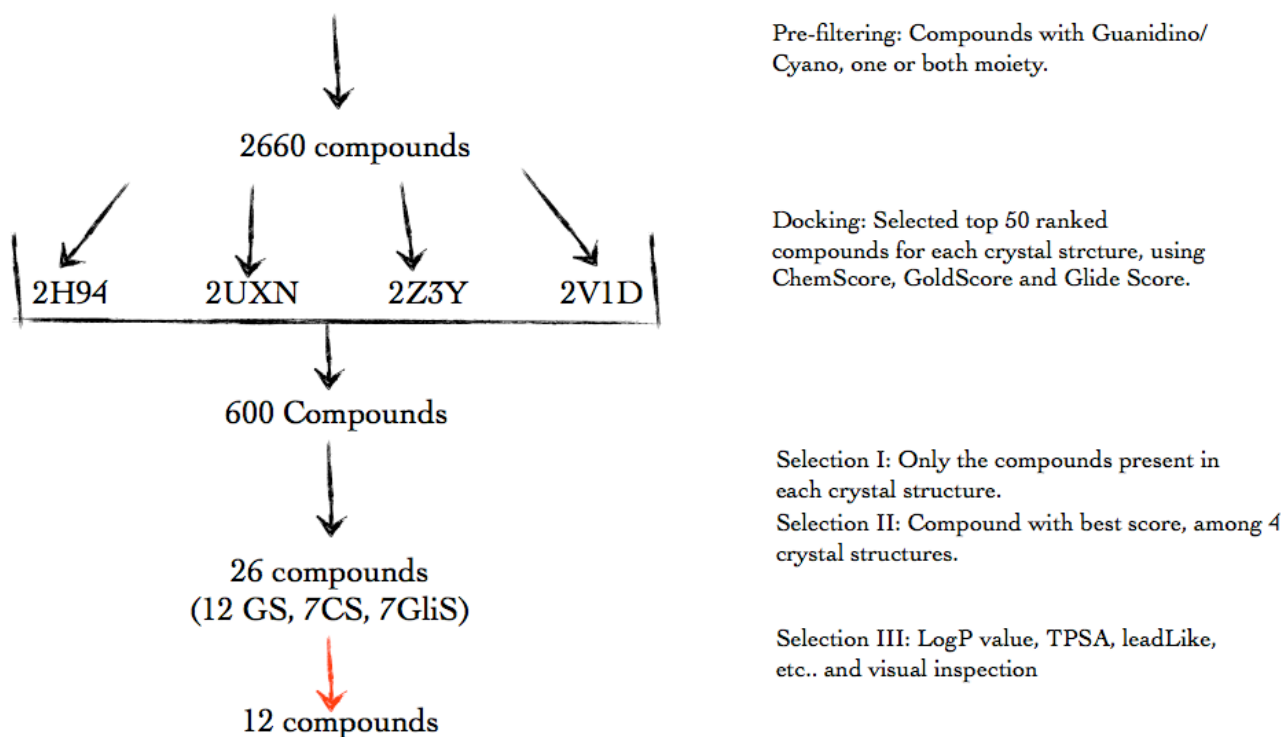


Figure 49: Schematic representation of the conducted multi-step virtual screening.

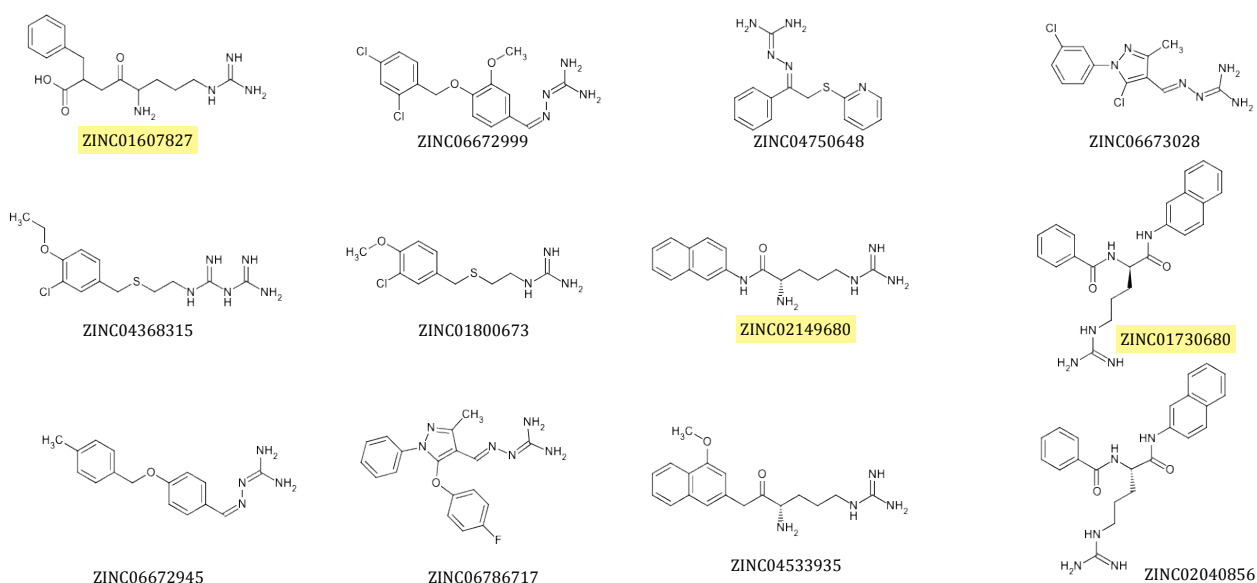


Figure 50: Compound structures obtained after the virtual screening procedure. Highlighted are the purchased ones.

In order to understand the inactivity of compounds **ZINC01607827**, **ZINC02149680** and **ZINC01730680**, the putative binding mode obtained through the virtual screening procedure was analysed (figure 51).

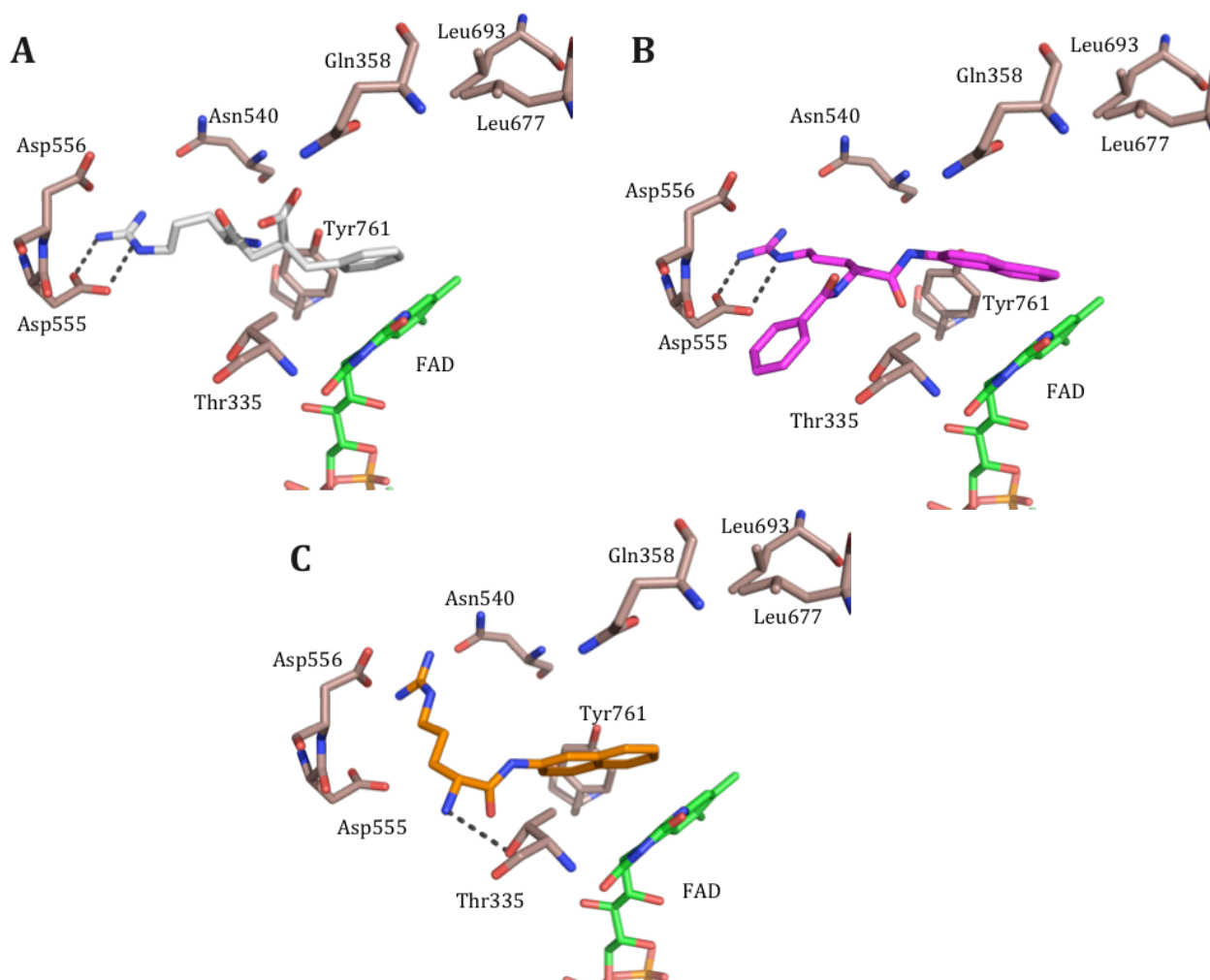


Figure 51: Binding mode representation of the three derivatives, obtained from virtual screening, using LSD1 crystal structure (PDB code 2V1D). In picture A **ZINC01607827** in B **ZINC01730680** and in C **ZINC02149680** are shown and they are represented as white, purple and orange sticks, respectively. The cofactor is represented as green sticks while the most involved residues of LSD1 are named and showed as brown sticks. Interactions are shown as black dashed lines while the hydrogens are omitted.

The derived binding mode showed for the three compounds an ionic interaction between the guanidine group and the side chain of Asp555. However, for all the derivatives the rest of the ligands structure is interacting mainly through hydrophobic interactions with the residues delimitating the cavity. Only **ZINC01607827** can find a hydrogen bond interaction with Thr335 (figure 51). All the molecules show a very flexible structure linked to an aromatic bulky moiety (benzyl group for **ZINC01607827** or naphthyl group for **ZINC02149680** and **ZINC017330680**) that in all the cases is projected into the region of the cavity most exposed to the solvent. The lack

of strong interactions between LSD1 and the bulky groups of the compounds, that show high flexibility, might be responsible for the inactivity.

3.3.4 Discussion

In order to identify new putative LSD1 inhibitors, PAO inhibitor datasets found in literature were used for docking studies. Five compounds were chosen and tested in peroxidase assay and one of them showed an inhibitory activity in the micromolar range (ligand **19**). A putative binding mode was derived using molecular dynamic simulations of ligand **19** inside LSD1. From this binding mode, a specific feature namely a guanidine group was taken into consideration for a stepwise virtual screening (VS) procedure using the ZINC Drug Like database.¹⁴⁹ From the VS procedure, 12 putatively active inhibitors were selected: three of them were commercially available, and none of them showed inhibitory activity on LSD1. From the docking poses obtained (figure 51), it was clear that the lack of interactions between the bulky group of the ligand and the residues inside the cavity of LSD1 was the reason for their inactivity. For this reason increasing the shape complementarity and the interactions between the aromatic bulky region of the compounds and the active cavity of LSD1 might be useful to define further inhibitors for LSD1.

3.4 4H-chromen-4-one derivatives

Since γ -pyrones derivatives have been demonstrated to be a novel class of reversible MAO-A and MAO-B inhibitors,¹⁵⁰ a γ -pyrone library containing 705 compounds was biologically screened against LSD1 enzyme in a peroxidase assay.^{45,54} From the screening two compounds were identified: 3-chloro-6-nitro-2-(trifluoromethyl)-4H-chromen-4-one (named **Namoline**) and 3-chloro-6-nitro-2-(pentafluoroethyl)-4H-chromen-4-one (figure 52). The second derivative inhibits the LSD1 demethylase activity with an IC_{50} of 250 μ M, while **Namoline** inhibits the LSD1 activity with an IC_{50} of 51 μ M.⁴⁵ Both derivatives inhibit LSD1 enzyme in a reversible manner, and **Namoline** causes cellular changes in the histone methylation levels at a concentration more than 20 μ M in dose-dependent manner, indicating its inhibitory potential on LSD1 also in cell-based assays.⁴⁵

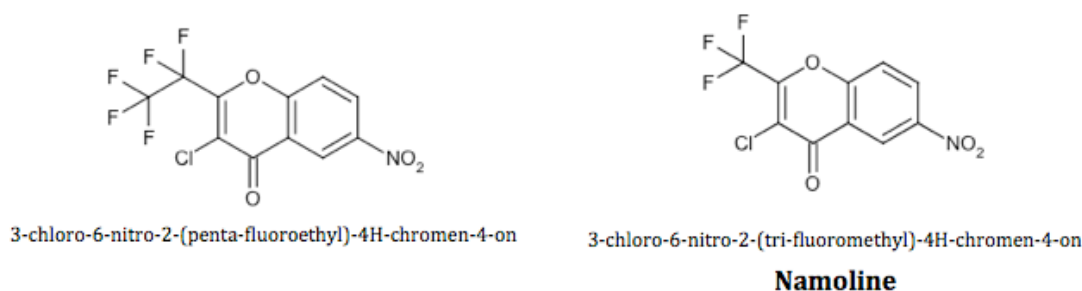


Figure 52: New non-covalent LSD1 inhibitors found through high throughput screening (HTS).

In order to optimize the hit **Namoline** a chemical synthesis coupled with a similarity search was carried out. To identify *in silico* models able to describe the binding mode inside LSD1 enzyme and to differentiate between active and inactive inhibitors a structure-based modelling was carried out. Starting from the Namoline structure, an *in silico* similarity search screening was conducted. The 3-chloro-2-(trifluoromethyl)-4H-chromen-4-one was selected as a substructure for similarity search using the ZINC database. Tanimoto similarity index was set from 60% up to 90% similarity, and a dataset of 26 compounds was obtained. Seven compounds were commercially available (figure 53). Three of them showed an inhibitory activity on LSD1 (**35a**, also named Shaolin; **32e**; and **33e**. For a complete name list, see Appendix B table 2) when tested in *in vitro* peroxidase assay (Chapter 2).

Results and Discussion: 4H-chromen-4-one derivatives

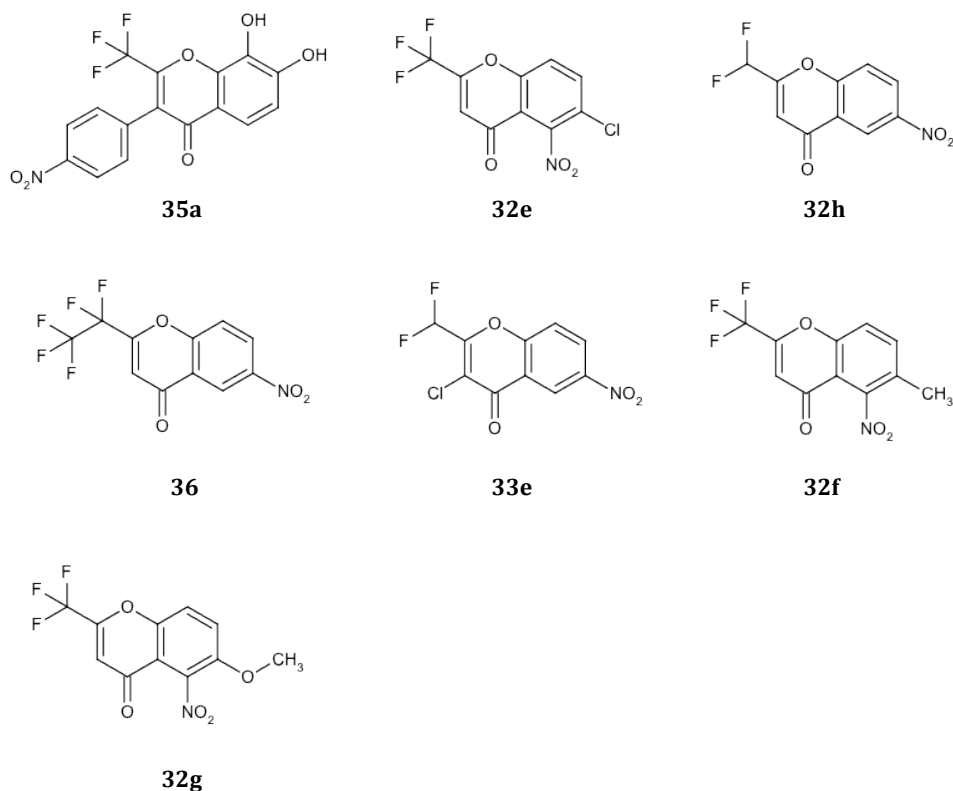


Figure 53: Commercially available Namoline analogues obtained by a similarity search.

Through synthesis carried out by Prof. Jung's group, 13 new compounds were generated and three of them were found to be active in the micromolar range when tested in the peroxidase assay (**33k**, **33l**, **33g** table 11 and Appendix B table 2). A new chromane dataset formed by the so far obtained seven LSD1 active inhibitors (table 11) and 14 inactive compounds was generated (figure 53 and 54, for a complete name list of the derivatives see Appendix B table 2). A mass spectrometry assay (data not shown) and an *in vitro* DELFIA® assay were also conducted for some of the compounds belonging to the dataset in order to confirm the activity of these new active inhibitors (table 11 and Appendix B table 2).

Results and Discussion: 4H-chromen-4-one derivatives

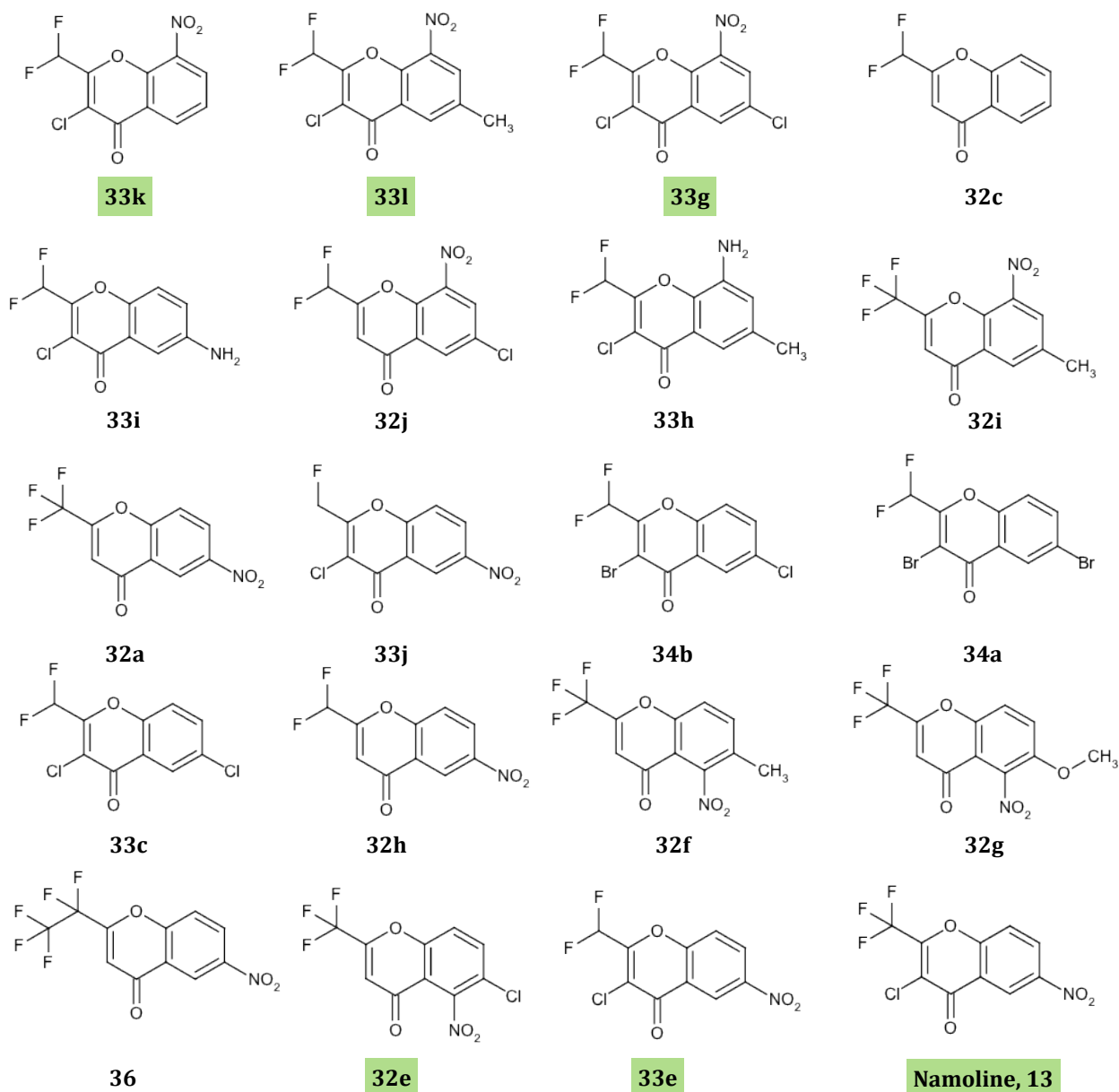


Figure 54: Dataset of six active and 14 inactive Namoline derivatives. Actives are highlighted.

Table 11: IC₅₀ values obtained by peroxide assay and DELFIA assay of the new non-covalent LSD1 active inhibitors.

Compounds	IC ₅₀ ± SE [μM] or Inhibition @ conc [μM]	
	Peroxidase assay	DELFLIA assay
Namoline	56.03 ± 1.62	3.80 ± 0.35
33k	1.43 ± 0.41	44% @ 0.2
331	3.93 ± 0.57	0.413
33g	7.06 ± 1.07	0.859
33e	26.66 ± 3.90	1.98 ± 0.18

32e	24.55 ± 2.98	Inactive @ 20
35a	2.47 ± 0.38	Inactive
33i	Inactive @ 50	//
33h	25% @ 50	//

DELFI[®] assay confirmed the inhibitory activity for all the new active inhibitors except for **32e**, which has been discarded from subsequent studies. Compound **33h** has been also discarded from the subsequent studies since no IC₅₀ value was determined.

3.4.1 Structure-Based approach

The aim of the structure-based approach was to generate a suitable model able to describe the binding mode of the active compounds inside LSD1 and then differentiate between active and inactive chromane derivatives. Docking studies and binding free energy calculation methods (Chapter 2) were the techniques used to reach this goal.

3.4.1.1 Docking studies

The molecular structures of all compounds analysed in the present study were generated using the MOE 2011.10 modelling software.¹⁴⁵ Initial ligand conformations resulted from an energy minimization using the MMFF94x force field implemented in MOE. Two available crystal structures of LSD1 (PDB codes: 2DW4 and 2UXN) were chosen according to their co-crystallized ligand and to their crystal structure resolution (Chapter 3.1). For the subsequent docking studies all water and ligand molecules were removed except in crystal structure 2DW4, where the conserved water molecule (Chapter 3.1) was kept inside the LSD1 active cavity for the whole study. Then all the structures were protonated using the “protonation 3D module” implemented in MOE and minimized using the Amber99 force field. Two docking softwares were used: Gold⁸⁷ and Glide⁹⁰ packages while the scoring functions selected were Goldscore⁸⁷ and Glidescore⁹⁰ standard precision (SP). Preliminary docking studies were conducted using three active compounds (**Namoline**, **33g** and **33i**). The best docking results, obtained with Gold, were analysed and the suitable poses inside the LSD1 active cavity were selected. Regarding **Namoline**, 35 different poses were selected using 2UXN crystal structure and 17 poses were selected using 2DW4 crystal structure. Concerning **33i**,

45 suitable poses inside 2UXN and 21 poses inside 2DW4 crystal structure were chosen. Whereas for **33g**, 44 poses inside 2UXN and 27 suitable poses using 2DW4 crystal structure were selected (table 12). It was impossible to identify a common putative binding mode from the obtained docking poses, based only on their dockings cores.

Table 12: Putative binding poses selected for three of the active inhibitors inside two LSD1 crystal structures.

Compounds	2UXN	2DW4
Namoline	35 poses	17 poses
331	45 poses	21 poses
33g	44 poses	27 poses

3.4.1.2 Binding free energy calculation studies

Since the docking scores did not allow to find a common binding model, we decided to estimate the binding free energy values of the obtained docking poses (rescoring process), using MM-PB(GB)SA and QM/MM-GBSA methods as described in chapter 2. AMBER 11⁸³ was chosen, while a single trajectory approach, generated after a minimization step, was the setup selected for the whole study.

These methods usually provide a more accurate description (taking into account a continuum solvent electrostatic to calculate the solvation free energy) of the ligand-protein interactions than the empirical Gold scoring function. In order to have a better discrimination among all the poses selected inside LSD1 crystal structures, two binding free energy calculation methods (MM-PBSA and MM-GBSA see Chapter 2) were applied for every ligand-crystal structure obtained from the docking studies. Taking into account also the nature of the active inhibitors, comprising two aromatic fused rings with a strong electron-withdrawing group in position number five or six or eight and several electron donor moieties; a strong role concerning the electrostatic energy of binding and the presence of polarizable effects during the interactions between chromane derivatives and LSD1 binding cavity was postulated. For these reasons a third protocol able to recognize small differences in the electrostatic profiles of the new active inhibitors (QM/MM-GBSA method) was also considered.

For all QM/MM-GBSA calculations only the ligand was chosen as quantum-mechanic region (QM region) of the system, and the semi-empirical potential Austin Model 1 (AM1)¹⁵¹ was selected. A semi-empirical potential with a small QM region (only the active ligand), roughly approximates the

binding free energy values, but on the other hand, it allows to drastically reduce the computational time of the rescoring calculations and permits to have a better discrimination among all the poses generated by the docking.

At the end of the first rescoring step, five putative binding modes were selected using 2UXN crystal structure (table 13), while only three putative binding modes were selected using 2DW4 crystal structure (table 14). The selection of the putative binding modes was based on considering the enthalpy values calculated with QM/MM-GBSA method and the enthalpy values calculated with MM-PB(GB)SA methods.

Table 13: Putative binding mode selection using 2UXN crystal structure, after MM-PBSA, MM-GBSA and QM/MM-GBSA binding free energy rescoring. ΔH_{tot} is the enthalpy value of the binding calculated for every method.

BM	Compound	GoldScore	ΔH_{tot}	ΔH_{tot}	ΔH_{tot}
		(Gold Ranking)	(MM-GBSA)	(MM-PBSA)	(QM/MM-GBSA)
1	Namoline	31.25 (16)	-20.28	-10.26	-24.92
	33g	34.56 (08)	-22.48	-9.60	-26.28
	33l	34.68 (06)	-24.53	-15.82	-34.15
2	Namoline	33.82 (05)	-28.54	-15.72	-31.55
	33g	31.72 (04)	-25.40	-12.29	-27.60
	33l	39.46 (01)	-27.29	-14.12	-28.26
3	Namoline	33.84 (04)	-20.12	-11.32	-18.97
	33g	40.81 (01)	-27.32	-18.07	-36.78
	33l	37.58 (04)	-25.26	-14.62	-32.78
4	Namoline	31.13 (18)	-18.30	-9.90	-21.51
	33g	31.25 (34)	-25.96	-11.15	-28.72
	33l	31.19 (30)	-21.59	-11.75	-24.43
5	Namoline	32.02 (12)	-15.33	-6.65	-20.97
	33g	33.34 (18)	-21.75	-15.73	-28.09
	33l	34.17 (09)	-18.23	-7.15	-26.59

Table 14: Putative binding mode selection using 2DW4 crystal structure, after MM-PBSA, MM-GBSA and QM/MM-GBSA binding free energy rescoring. ΔH_{tot} is the enthalpy value of the binding calculated for every method.

BM	Compound	GoldScore	ΔH_{tot}	ΔH_{tot}	ΔH_{tot}
		(Gold Ranking)	(MM-GBSA)	(MM-PBSA)	(QM/MM-GBSA)
1	Namoline	29.75 (15)	-21.77	-12.12	-23.61
	33g	35.07 (12)	-22.05	-15.85	-26.57
	33l	34.77 (11)	-21.19	-15.57	-25.36
2	Namoline	43.94 (01)	-26.33	-21.28	-30.02
	33g	42.42 (01)	-25.80	-20.45	-30.25
	33l	44.14 (01)	-26.59	-19.90	-29.84
3	Namoline	34.21 (08)	-21.93	-8.23	-22.51
	33g	37.27 (05)	-25.77	-8.43	-26.16
	33l	38.62 (03)	-24.57	-10.38	-28.55

Based on the *in vitro* data of the peroxidase assay a training set, formed by four active inhibitors (**33l**, **33g**, **33k**, and **Namoline**) and seven inactive compounds (**32i**, **32j**, **32h**, **32c**, **32a**, **34b**, and **33j**) was picked up. The selected compounds were docked inside LSD1 crystal structures (PDB codes: 2UXN and 2DW4). Then the MM-PB(GB)SA and QM/MM-GBSA protocols previously described were applied generating the binding free energy value for all the compounds belonging to the training set, according to the binding modes found (table 13 and 14).

In order to increase the accuracy of QM/MM-GBSA calculations, an increased QM region, taking into account the ligand and the residues within 4.5 Å around it, was selected for all the calculations. Through this approach we could hypothesize a reasonable binding mode, able to distinguish between active and inactive compounds, inside LSD1 (table 15).

The binding free energy values obtained with the previously described three methods (MM-PB(GB)SA and QM/MM-GBSA) for the compounds belonging to the training set using the five selected putative binding modes inside 2UXN crystal structure did not show any discrimination between active and inactive inhibitors (data not shown). On the other hand, the binding mode hypothesis 1 inside 2DW4 crystal structure (table 14) showed discrimination between active and inactive inhibitors using QM/MM-GBSA rescoring method (table 15). For this reason this hypothesis was selected as putative binding mode of the new chromane derivatives (figure 55).

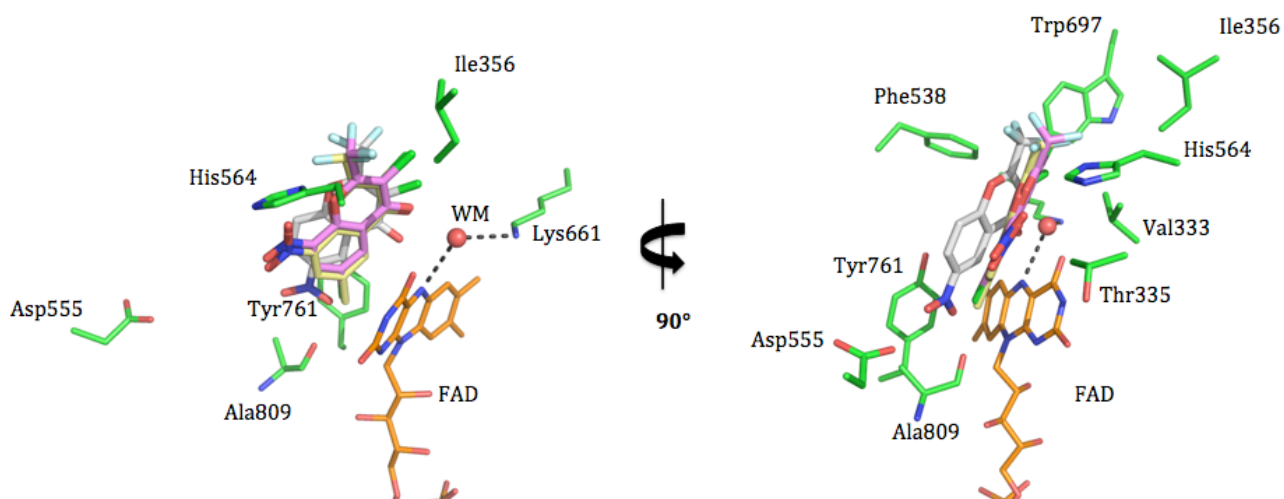


Figure 55: Superposition of three active compounds inside LSD1 crystal structure (PDB code: 2DW4), according to the putative binding mode. **Namoline** is represented as white sticks, **33g** is shown as yellow sticks, **33l** is represented as pink sticks. The most involved residues are named with three letters code and shown as green sticks, while FAD cofactor is represented as orange sticks. The water molecule is represented as red ball while the main interactions are shown as black dashed lines. Hydrogen atoms are omitted.

Table 15: Final hypothesized binding mode. Enthalpy contributions were obtained by MM-PBSA, MM-GBSA and QM/MM-GBSA (considering all the residues within 4.5 Å around the ligand as QM region) of the chromane training set (pdb code 2DW4). ΔH_{tot} is the enthalpy value of binding for every method. In addition the Goldscore is shown. Actives are listed as the first four compounds in the table.

Compound	GoldScore (Gold Ranking)	ΔH_{tot} (MM-GBSA)	ΔH_{tot} (MM-PBSA)	ΔH_{tot} (QM/MM-GBSA)
Namoline	29.75	-21.77	-12.12	-0.59
33g	35.07	-22.05	-15.85	-4.62
33l	34.77	-21.19	-15.57	-3.50
33k	31.61	-21.03	-10.15	-2.34
32a	33.28	-20.42	-9.20	2.60
32i	33.77	-19.12	-10.07	0.27
32c	30.95	-16.21	-8.75	4.11
32j	35.14	-19.16	-10.83	-0.90
32h	36.15	-21.82	-16.08	0.63
34b	35.84	-21.89	-15.07	-1.80
33j	33.41	-23.43	-17.19	0.05

The binding mode shows the chromane ring allocated directly in front of the FAD cofactor isoalloxazine ring (figure 55). Here the carbonyl group present in the chromane structure interacts through a hydrogen bond with the water molecule present in the crystal structure. The other

residues delimiting the binding area (His564, Thr335, Val333, Ile356, and Tyr761) interact through hydrophobic and electrostatic interactions (figure 55). The calculated enthalpy obtained with QM/MM-GBSA calculations showed favourable energy values for all the active inhibitors while it rescored the inactive compounds with unfavourable energy values except for compound **32j** and compound **34b** (table 15).

The last stage of the study was to rescore the docking poses of the entire dataset using the QM/MM-GBSA method. Two main QM/MM-GBSA settings were used in this step. In the first case the QM region included the ligand and all the residues within 4.5 Å around it as calculated for the training set (data not shown), meanwhile in the second case the calculations were carried out considering the ligand and all the residues within 6 Å around the ligand as QM region (table 16). This allowed increasing the accuracy of the binding free energy estimation even if it slowed down the calculation time for every complex (the time needed for the generation of binding free energy was about 15 minutes per complex). The AM1 pseudo potential was selected for the rescoring of all complexes (figure 56).

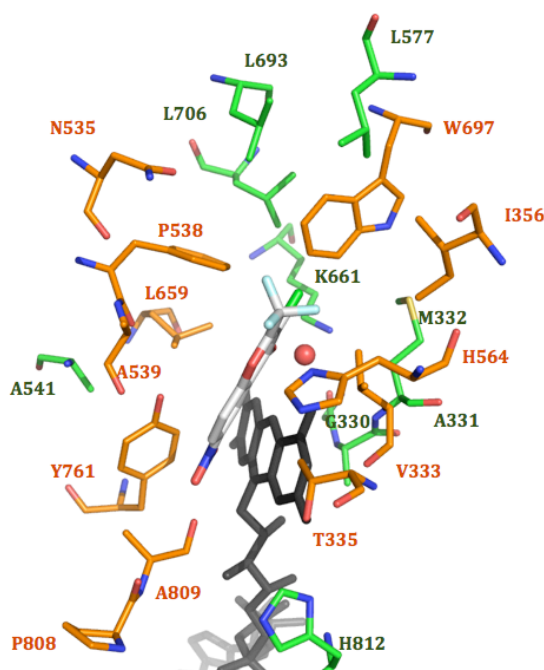


Figure 56: Residues selected as QM region during the QM/MM-GBSA calculations. The residues located up to 4.5 Å around namoline are shown as orange sticks and named using one letter code (orange). The residues located up to 6 Å around namoline are shown as green sticks and named using one letter code in green. **Namoline** is shown as white sticks, FAD as black sticks, while the water molecule is shown as red ball.

Table 16: Energy contributions to the free energy of binding of the whole chromane dataset inside LSD1 crystal structure (PDB code 2DW4) using QM/MM-GBSA rescoring method (6 Å around the ligand as QM region) ΔE_{vdw} are the van der Waals energies of binding; ΔG_{solv} are the contributions of the solvation free energy; ESCF are the self-consistent field method energies obtained from the QM region of the system; ΔH_{tot} are the enthalpy values of the binding, while ΔG_{calc} are the binding free energy values taking into account enthalpy and entropy contributions. Bonds are the number of rotatable bonds present in the ligand structure.

Compounds	ΔE_{vdw}	ESCF	ΔG_{solv}	ΔH_{tot}	Bonds	ΔG_{calc}	IC ₅₀ [μM]
Namoline	-10.85	-7.42	20.65	2.38	0	2.38	51
33g	-9.28	3.99	3.52	-1.80	1	-2.80	7.06 ± 1.07
33l	-9.65	6.44	2.78	-0.46	1	-1.46	3.93 ± 0.57
33k	-8.01	1.20	7.77	0.94	1	-1.94	1.43 ± 0.41
33e	-10.94	-11.08	24.99	2.97	1	1.97	26.66 ± 3.90
32a	-10.02	-0.44	16.14	5.65	0	5.65	Inactive
32i	-8.73	15.03	-2.86	3.45	0	3.45	Inactive
32c	-6.63	-8.64	22.69	7.39	1	6.39	Inactive
32j	-8.31	9.20	1.01	1.90	1	0.90	Inactive
32h	-10.28	-13.07	26.90	3.55	1	2.55	Inactive
34b	-9.05	-14.10	24.49	1.40	1	0.40	Inactive
33j	-10.73	-18.49	32.30	3.11	1	2.11	Inactive
33c	-8.52	-4.24	14.48	1.69	1	0.69	Inactive
34a	-9.30	-14.95	26.49	2.18	1	1.18	Inactive
33i	-9.64	-20.68	28.82	-1.49	1	-2.49	Inactive
32f	-7.21	-18.45	31.94	6.29	0	6.29	Inactive
36	-10.14	0.01	15.48	5.41	1	4.41	Inactive
32g	-7.74	-22.81	37.07	6.51	1	5.51	Inactive

After the ΔH_{tot} was calculated for each LSD1-ligand complex, the binding free energy values (ΔG_{calc}) were also approximated taking into account the entropy (TΔS) data. As reported in literature there is a linear correlation between the entropy values and the number of rotatable bonds present in a ligand structure.¹²⁴ Several authors also argued that including the entropy values approximated from the number of the ligand rotatable bonds might increase the correlation between the approximated binding free energy value calculated and the experimental activity data.^{124,152,153} For this reason we decided to approximate the entropy value from the number of rotatable bonds present in each active inhibitor structure (1 kcal/mol for 1 rotatable bond).¹²⁴

Summing then the enthalpy values (ΔH) and the approximated entropy values ($T\Delta S$) we could estimate the binding free energy for every LSD1-chromane derivative complex (ΔG , Chapter 2), with a reasonable computational time.

3.4.1.3 QM/MM-GBSA model and new putative in silico generated LSD1 inhibitors activity prediction.

Through structure-based approach several descriptive models were generated according to the binding mode selected. The binding free energy values obtained for every model (MM-PB(GB)SA and QM/MM-GBSA) were correlated to the activity values in order to check the quality of the models obtained. It is important to note that since *in vitro* assay data are used instead of thermodynamic values (as by calorimetric studies), the correlation of the binding free energies with the activity values is in any case an approximation. However neither MM-PBSA nor MM-GBSA model were able to rank the new inhibitors according to their *in vitro* activity. Both models indeed rescore all compounds with favourable binding free energies. For example, the ΔG_{calc} values calculated with the MM-PBSA approach for the active compounds lie between -11.15 Kcal/mol (for **33k**) and -16.85 Kcal/mol (for **33g**). Concerning the inactive compounds the ΔG_{calc} values range from -9.20 Kcal/mol (for **32a**) to -19.97 Kcal/mol (for **33i**). Active and inactive inhibitors are ranked in the same range of values, and so no differentiation was achieved between active and inactive compounds.

The energy data obtained from the QM/MM-GBSA model was also examined. The generated model can roughly discriminate active and inactive inhibitors (figure 57 and table 16), even if it is not ranking the compounds according to their experimental activity (there is no correlation between experimental data and *in silico* binding free energy).

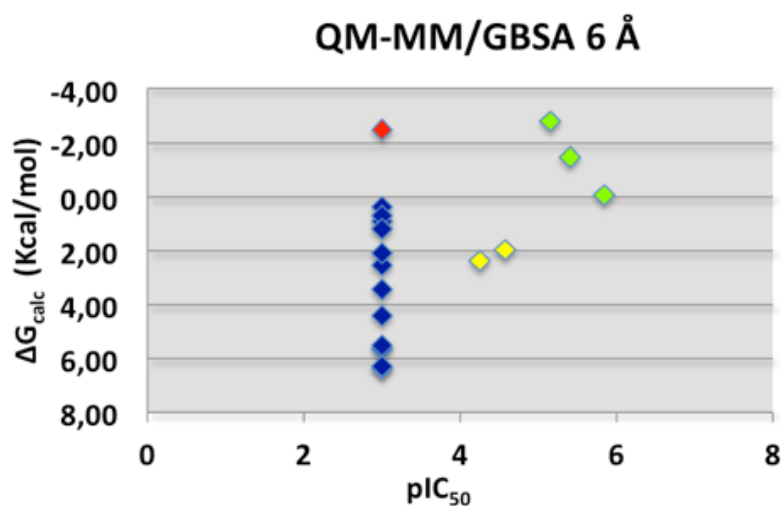


Figure 57: Correlation between the binding free energy values obtained by QM/MM-GBSA method and the activity data. Active inhibitors are represented as green dots, while inactive inhibitors are represented as blue dots. **Namoline** and **33e** are represented as yellow dots, and compound **33i** is represented as red dot. pIC_{50} is the negative logarithm form of the IC_{50} normalized to the ligand IC_{50} ($pIC_{50} = -\log_{10}(IC_{50} \cdot 10^{-6})$).

The active inhibitors show ΔG_{calc} values ranging from -2.80 Kcal/mol (**33g**) to 2.38 Kcal/mol (**Namoline**), while the inactives' ΔG_{calc} values are from -2.49 Kcal/mol (**33i**) to 6.39 Kcal/mol (**32c**). The model obtained is able to discriminate between highly active and inactive compounds giving them an unfavourable binding free energy value (all the energy scores are included from zero to positive values) except for **33i** that shows negative ΔG_{calc} (-2.49 Kcal/mol). The negative binding free energy of **33i** is mainly due to a hydrogen bond interaction between the amine moiety present in compound **33i** and the backbone of Ala809 residue present in the active cavity of LSD1 (figure 55) which increases the enthalpy of binding. Concerning the active compounds, the model generated cannot properly classify **Namoline** and **33e** active inhibitors with an IC_{50} of 51 and 26.66 μM ; respectively.

In order to investigate the binding role of the nitro group at the chromane scaffold and in order to understand the role of the polarizable effects inside LSD1 active cavity, new proposals for **Namoline** analogues were generated *in silico* (figure 58), docked inside LSD1 active cavity according to the selected putative binding mode, and their binding free energies were calculated using the selected QM/MM-GBSA rescoring model.

Results and Discussion: 4H-chromen-4-one derivatives

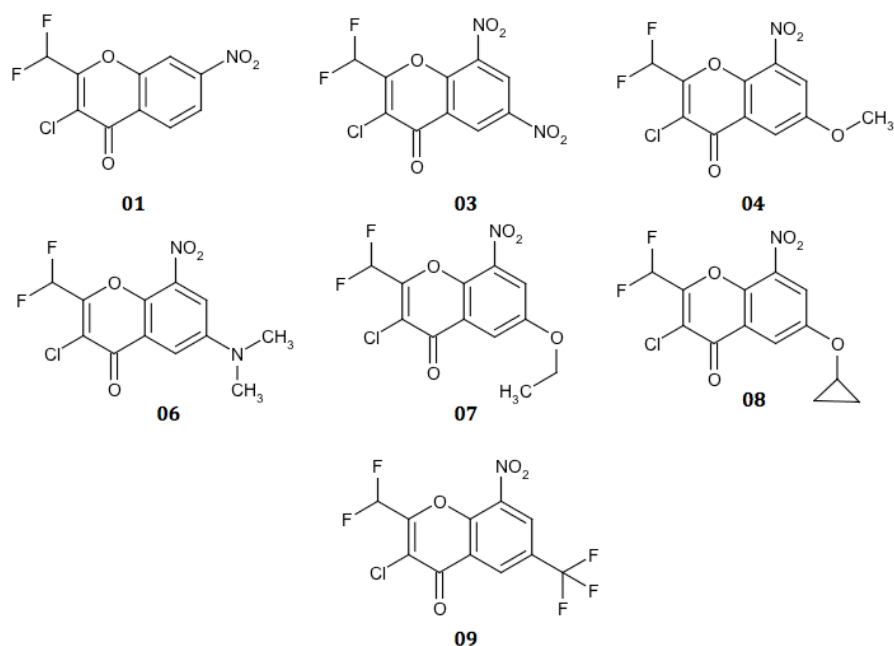


Figure 58: Proposal for novel LSD1 inhibitors.

The generated structures showed a common 3-chloro-8-nitro-2-(difluoromethyl)-4H-chromen-4-one scaffold, like the most active inhibitors present in the dataset (with different electron withdrawing or electron donor groups in position 6 of the chromane scaffold) except for compound **01** which is unsubstituted at position six.

Table 17: Binding free energies of the *in silico* generated putative LSD1 inhibitors using QM/MM-GBSA rescoring method (6 Å around the ligand as QM region). ΔE_{vdw} are the van der Waals energies of binding; ΔG_{solv} are the contributions of the solvation free energy; ESCF are the self-consistent field method energies obtained from the QM region of the system; ΔH_{tot} are the enthalpy values of the binding, while ΔG_{calc} are the binding free energy values taking into account enthalpy and entropy contributions. Bonds are the number of rotatable bonds present in the ligand structure.

Compounds	ΔE_{vdw}	ESCF	ΔG_{solv}	ΔH_{tot}	Bonds	ΔG_{calc}
01	-9.77	-0.72	10.28	-0.17	1	-1.17
03	-10.90	4.13	8.81	2.01	1	1.01
04	-9.51	8.18	0.88	-0.44	2	-2.44
06	-9.48	4.33	5.20	0.10	2	-1.90
07	-7.40	3.43	4.54	0.57	3	-2.43
08	-9.28	11.84	-2.88	-0.38	2	-2.38
09	-9.57	11.53	-0.67	1.20	1	0.20

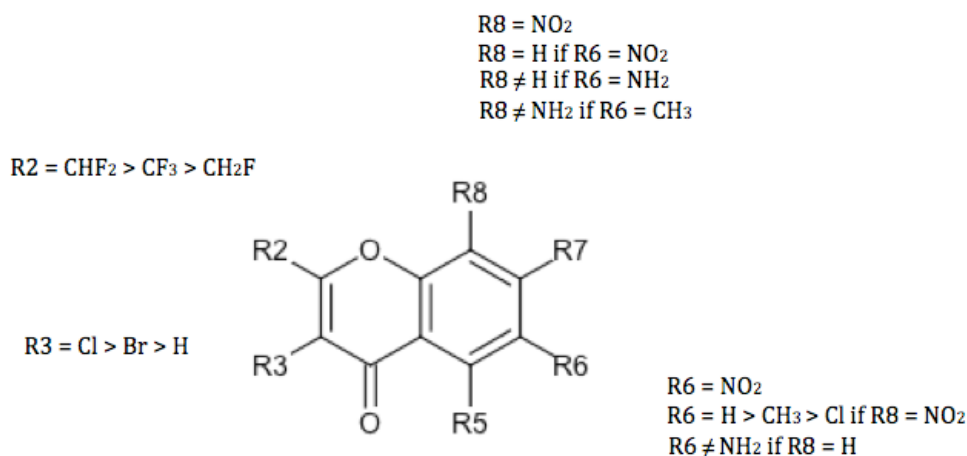
The calculated ΔG_{calc} values were favourable for all the new *in silico*-generated compounds except for compound number **03** and compound number **09**. This indicates that compounds bearing a

strong electron withdrawing groups like nitro and trifluoromethyl in position 6 of the 3-chloro-8-nitro-2-(difluoromethyl)-4H-chromen-4-one scaffold might be not active as inhibitors on LSD1. While compounds with a strong or medium electron donor in position six (compounds **04**, **06**, **07**, **08**) might result as new LSD1 active inhibitors. Moreover, compound **01** also showed a favourable ΔG_{calc} value indicating a putative inhibition on LSD1.

3.4.2 Structure activity relationships (SAR) of chromane derivatives

According to the models generated by the structure-based approach we were able to describe some preliminary SAR of the new LSD1 non-covalent chromane inhibitors.

In position number two of the chromane scaffold a difluoromethyl group (**33e**, **33g**, **33k** and **33l**) is more favourable than a trifluoromethyl (**Namoline**) or a monofluoromethyl (**33j**). While in position number three the chlorine group is preferred (**Namoline**, **33e**, **33g**, **33k** and **33l**) to a hydrogen atom (**32a**, **32h**, **32i**, and **32j**). At position number six of the chromane scaffold a nitro group is a more favourable substitution (**Namoline**) than the corresponding compound without nitro group (**33c**). With the concomitant presence of a nitro group in position number eight of the chromane scaffold (**33g**, **33l** and **33k**), in position number six a hydrogen atom is more favourable (**33k**) than a bulky weak electron donor as a methyl (**33l**) or a bulky weak electron withdrawing group as a chlorine (**33g**). Again an amino group in position number eight with a concomitant presence of an electron donor as a methyl group in position number six (**33h**) and an amino group in position number six with a concomitant presence of a hydrogen atom in position number eight (**33i**) indicates inactivity for LSD1 inhibition.



3.4.3 Discussion

Through the combination of computational methods and experimental *in vitro* testing five new non-covalent chromane scaffold inhibitors active in the micromolar range were found.

In order to obtain more information about the LSD1-ligand complex binding affinity Gold and Glide docking solutions were rescored by three main binding free energy calculation methods (MM-PBSA, MM-GBSA and QM/MM-GBSA), based on a single protein-ligand complex trajectory obtained after minimization. The mixed quantum mechanics/molecular mechanics method (QM/MM-GBSA) generated the most reliable structure-based model, since it takes into account polarizable effects, providing a putative binding mode able to discriminate between active and inactive compounds inside LSD1 active cavity.

Nine putatively active non-covalent chromane-scaffold-based LSD1 inhibitors were generated *in silico* and the calculated binding free energy values according to the generated model showed favourable energy scores for most of them.

However, one should keep in mind that QM/MM-GBSA score is not the absolute binding free energy value and it should not be used as a quantitative value because the method definitely depends on the initial docking pose, on the dataset and on the setting parameters chosen (force fields, dimensions QM region and MM region, etc.).

Here was showed that for the chromane inhibitors a structure-based approach might result partially reliable approach. However the model is not validated yet and the characterization of new LSD1 inhibitors chromane-scaffold based is still mandatory in order to improve the reliability and predictively of the model.

3.5 Coumarine derivatives

Coumarine derivatives are well-known MAOs inhibitors.¹⁵⁴ In this chapter a similarity screening approach is proposed in order to find new putative LSD1 inhibitors based on a coumarine scaffold.

3.5.1 Preliminary similarity search

A coumarine derivative (**HR47**) has been synthesized in Prof Jung's group and showed micromolar inhibitory activity on LSD1 (table 18) when tested in an *in vitro* peroxidase assay (Chapter 2). A similarity search procedure using **HR47** as query structure, MACCS fingerprints and the Chembridge database¹⁵⁵ was carried out. 19 molecules were selected (Tanimoto coefficient 0.80) and purchased (for a complete list of the purchased compounds see Appendix B table 4). Three of them showed an inhibitory activity in the low micromolar range when tested on LSD1 in the *in vitro* peroxidase assay (figure 59, table 18, and Appendix B table 4 for the complete list).

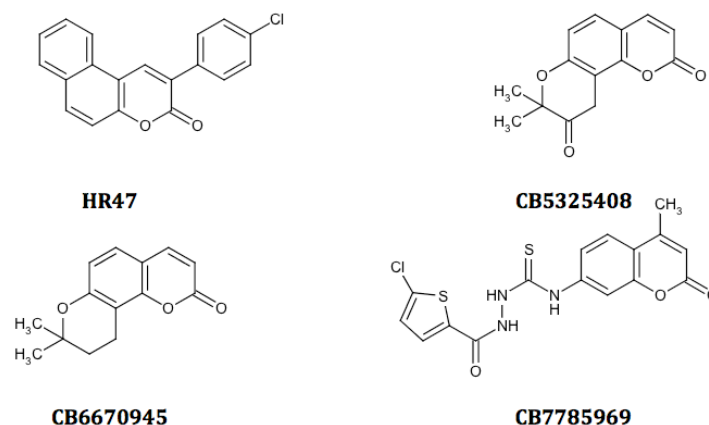


Figure 59: identified Coumarine derivatives showing inhibitory activity at LSD1.

Table 18: *in vitro* peroxidase assay activity data of the new coumarine-scaffold based LSD1 inhibitors.

Compounds	$IC_{50} \pm SE$ [μM] or Inhibition @ conc [μM]
	Peroxidase Assay
HR47	90% @ 50 μM
CB5325408	~ 5 μM
CB6670945	100% @ 50 μM
CB7785969	~ 3 μM

3.5.2 Docking studies

The available crystal structures of LSD1 with PDB code 2DW4¹³² and 2UXN¹³³ were chosen for the docking studies. For the subsequent docking studies, only the protein and one conserved water molecule were retained. The protein structures were protonated using the “protonation 3D module” implemented in MOE and minimized using the Amber99 force field. Two docking software were used: Gold⁸⁷ and Glide⁹⁰ while Goldscore⁸⁷ and Glidescore⁹⁰ were selected as scoring functions.

The new coumarine derivatives were also docked to MAO-B. The preliminary cross-docking step using the co-crystallized MAO-B non-covalent inhibitors (PDB structure 1OJA,¹⁵⁶ 2V60¹⁵⁷ and 2V61¹⁵⁷) allowed to choose the best docking setting to use for the docking of the obtained coumarine inhibitors. Cross-docking studies using four water molecules in the MAO-B active site and Gold⁸⁷ software showed the best performance (tables 19, 20 and figure 60).

Table 19: Cross-validation step using Gold docking software and three MAO-B crystal structures (PDB codes 1OJA, 2V60 and 2V61). The rank of the poses with the lowest RMSD is given in brackets.

Gold Score RMSD Matrix (Å)			
Ligand/Protein	1OJA	2V60	2V61
1OJA	0.54 (1)	3.22 (6)	4.53 (13)
2V60	5.54 (3)	0.55 (1)	1.30 (2)
2V61	5.93 (10)	1.24 (1)	2.47 (11)

Table 20: Cross-validation step using Glide docking software and three MAO-B crystal structures (PDB codes 1OJA, 2V60 and 2V61). The rank of the poses with the lowest RMSD is given in brackets.

Glide Score RMSD Matrix (Å)			
Ligand/Protein	1OJA	2V60	2V61
1OJA	0.43 (1)	2.38 (1)	0.95 (1)
2V60	Not Found	0.76 (3)	2.49 (Last)
2V61	Not Found	1.22 (1)	2.69 (2)

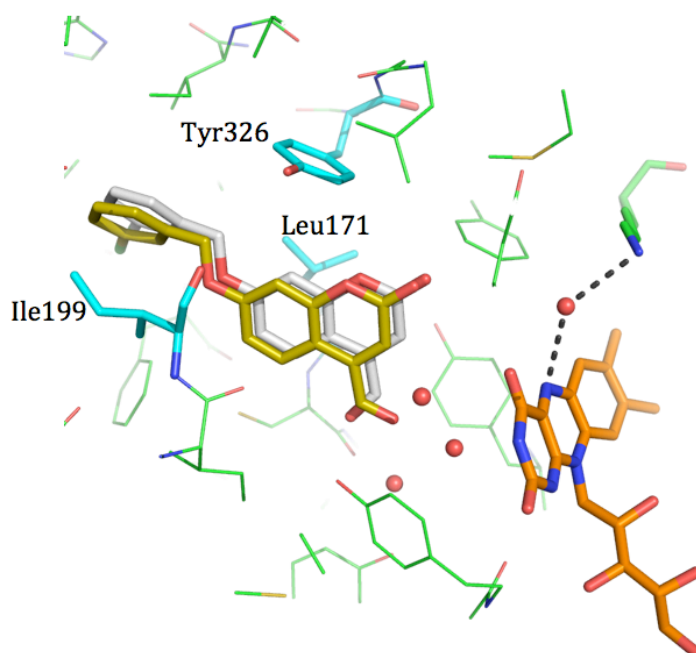


Figure 60: Best docking pose overlapped with the co-crystallized ligand inside MAO-B (PDB code 2V60), obtained with Gold software. The FAD cofactor is represented as orange sticks. The best docking pose is shown as yellow sticks while the co-crystallized ligand is shown as white sticks. Some of the main-involved residues are represented as cyan sticks and the water molecules are shown as red balls. Hydrogen atoms are omitted.

After the cross-docking step, the coumarine scaffold-based inhibitors (figure 59) were docked to MAO-B. A putative binding mode inside MAO-B was found for the most active compound (**CB7785969**) of the dataset. In the hypothesis compound **CB7785969** adopts the same conformation of the coumarin as in the co-crystallized ligand present in the PDB crystal structure 2V60 (figure 61). Compound **CB7785969** interacts through electrostatic and van der Waals interactions with Tyr326, Leu171 and Gln206 side chains, and with the backbone of Ile199 and Leu164 like the co-crystallized ligand (figure 61A). The chlorine atom present in the structure might also find a halogen bond interaction with the backbone carbonyl group of Leu164 residue (figure 61C). Also this last interaction is found with the co-crystallized ligand of the crystal structure 2V60 and for these reasons the found binding mode has been considered as putative binding mode of compound **CB7785969** at MAO-B. Moreover compound **CB7785969**, due to its molecular length (it is longer than the co-crystallized ligand present in MAO-B crystal structure 2V60), might find new water molecule mediated hydrogen bond interaction with the N5 atom of FAD cofactor (figure 61B).

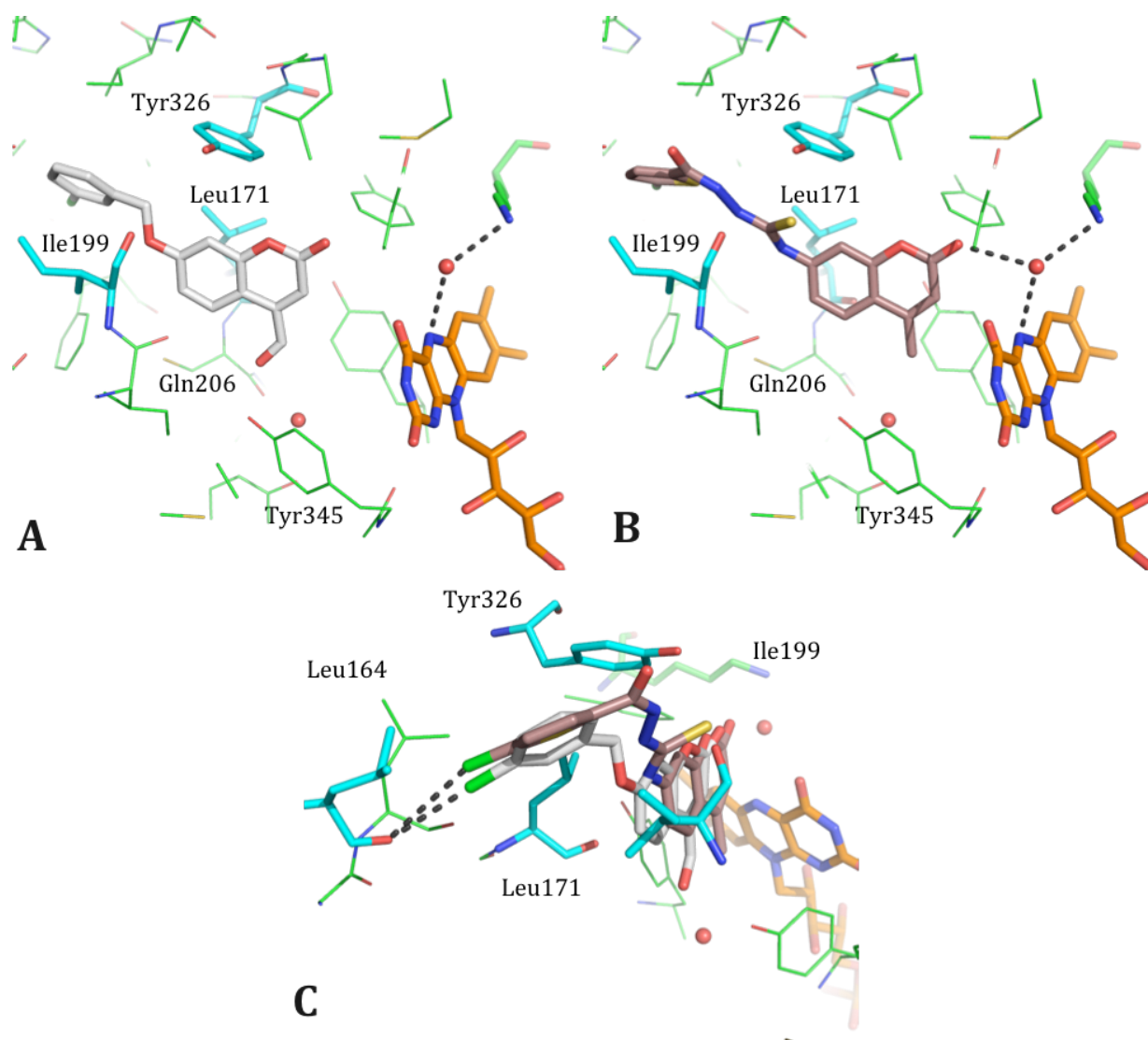


Figure 61: **A** Binding mode of the co-crystallized ligand with PDB code 2V60. **B** Putative binding mode of compound **7785969** inside MAOB active cavity (PDB code 2V60), found through docking studies. **C** Halogen bond interaction with the backbone carbonyl oxygen of Leu164, which both ligands might find. FAD cofactor is coloured as orange sticks, compound **7785969** is shown as brown sticks. The co-crystallized ligand is coloured as white sticks. Some of the most involved residues of MAO-B are represented as cyan sticks while the water molecules are shown as red balls. Hydrogen bonds are shown as black dashed lines, while hydrogen atoms are omitted.

In case of LSD1 both docking softwares were able to identify a common putative binding mode even if no scoring function correctly ranked the derivatives according to their experimental activity obtained from *in vitro* peroxidase assay. The putative binding mode identified for LSD1 (figure 62) showed a similar orientation of the coumarine ring when compared to the binding mode of compound **CB7785969** in MAO-B.

The hypothesis shows the same water-mediated interaction between the N5 atom of the FAD isoalloxazine ring and the carbonyl oxygen of the coumarine ring (figure 62). Other interactions are mainly electrostatic and van der Waals interactions with Phe538, Ala539, Asp555, Trp695, and

Ala809 residues (figure 62). However, compound **CB7785969** shows an additional hydrogen bond interaction with the backbone of Ala809 residue (figure 62D).

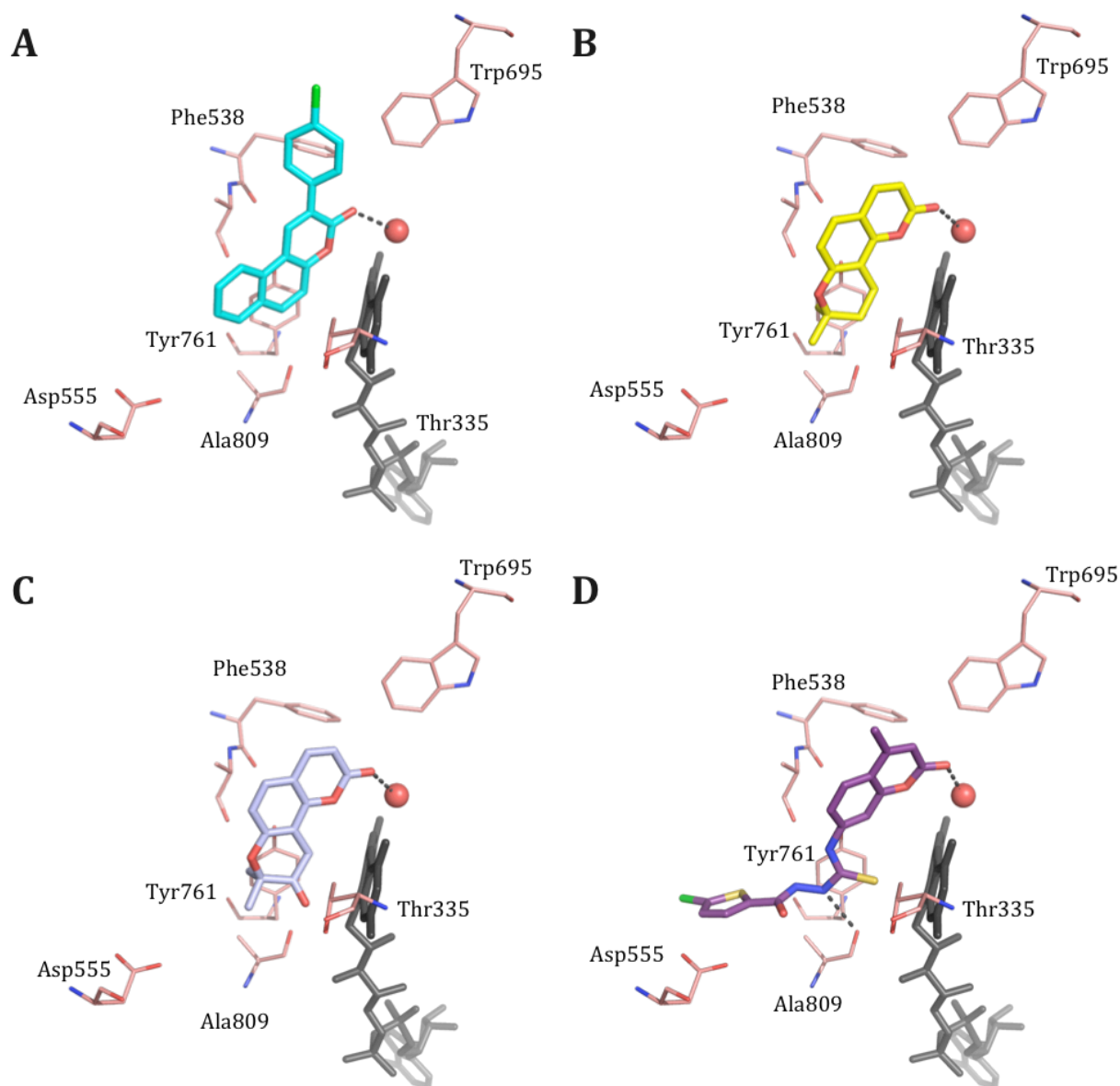


Figure 62: Hypothesis of binding for the new coumarine derivatives. A: **HR47**, B: **CB6670945**, C: **CB5325408** D: **CB7785969**.

FAD is represented as black sticks; the coumarine derivatives are represented as cyan, yellow, light blue and purple sticks respectively. The most involved residues of LSD1 are named and shown as pink sticks. Water molecules are shown as red spheres, while hydrogen bond interactions are represented as black dashed lines.

3.5.3 Similarity search

In order to identify further coumarine-based LSD1 inhibitors a second similarity search virtual screening was carried out. Derivative **CB7785969** was selected as a query structure (figure 63) while the virtual screening was conducted using MACCS fingerprints within MOE2011.10.¹⁴⁵

Results and Discussion: Coumarine derivatives

Chembridge compound collection (590029 compounds) was selected and a Tanimoto coefficient of 0.85 was taken as cutoff value.

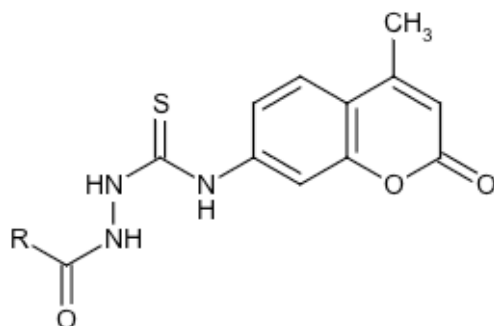
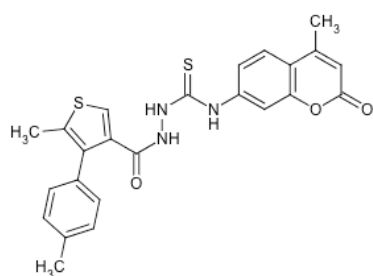
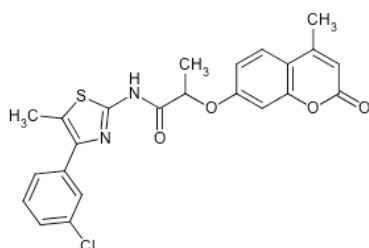


Figure 63: Coumarine derivative query structure used for the similarity search virtual screening.

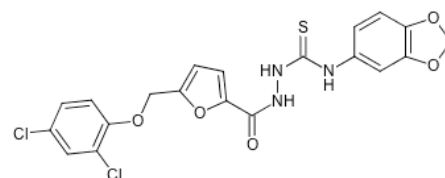
24 hits (figure 64) were identified and docked into MAO-B (PDB code 2V60) and LSD1 (PDB code 2DW4) binding pockets. For docking Gold software and GoldScore was used and finally the top ranked compounds, which showed the same interactions (calculated with the MOE protein-ligand interaction fingerprint, data not showed) as compound **CB7785969** were retrieved and purchased (figure 64).



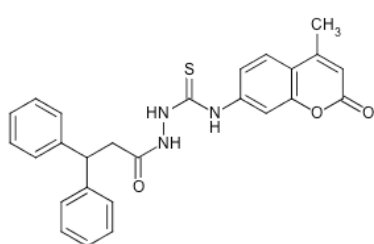
CB7799868



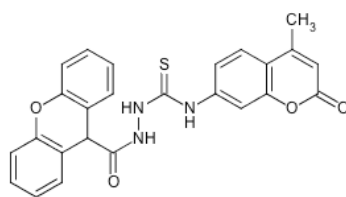
CB7930465



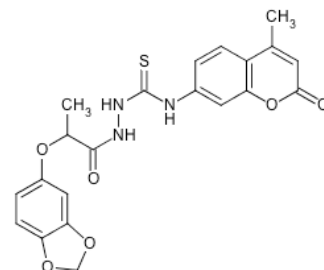
CB7788596



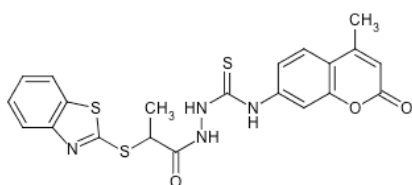
CB7821436



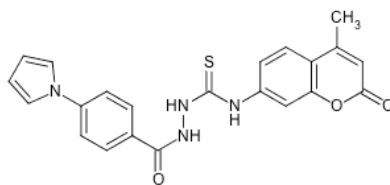
CB7800770



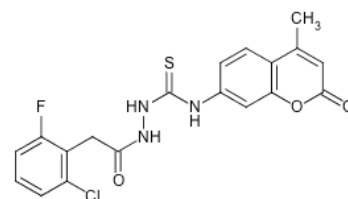
CB7798329



CB7824193



CB7799353



CB7833564

Results and Discussion: Coumarine derivatives

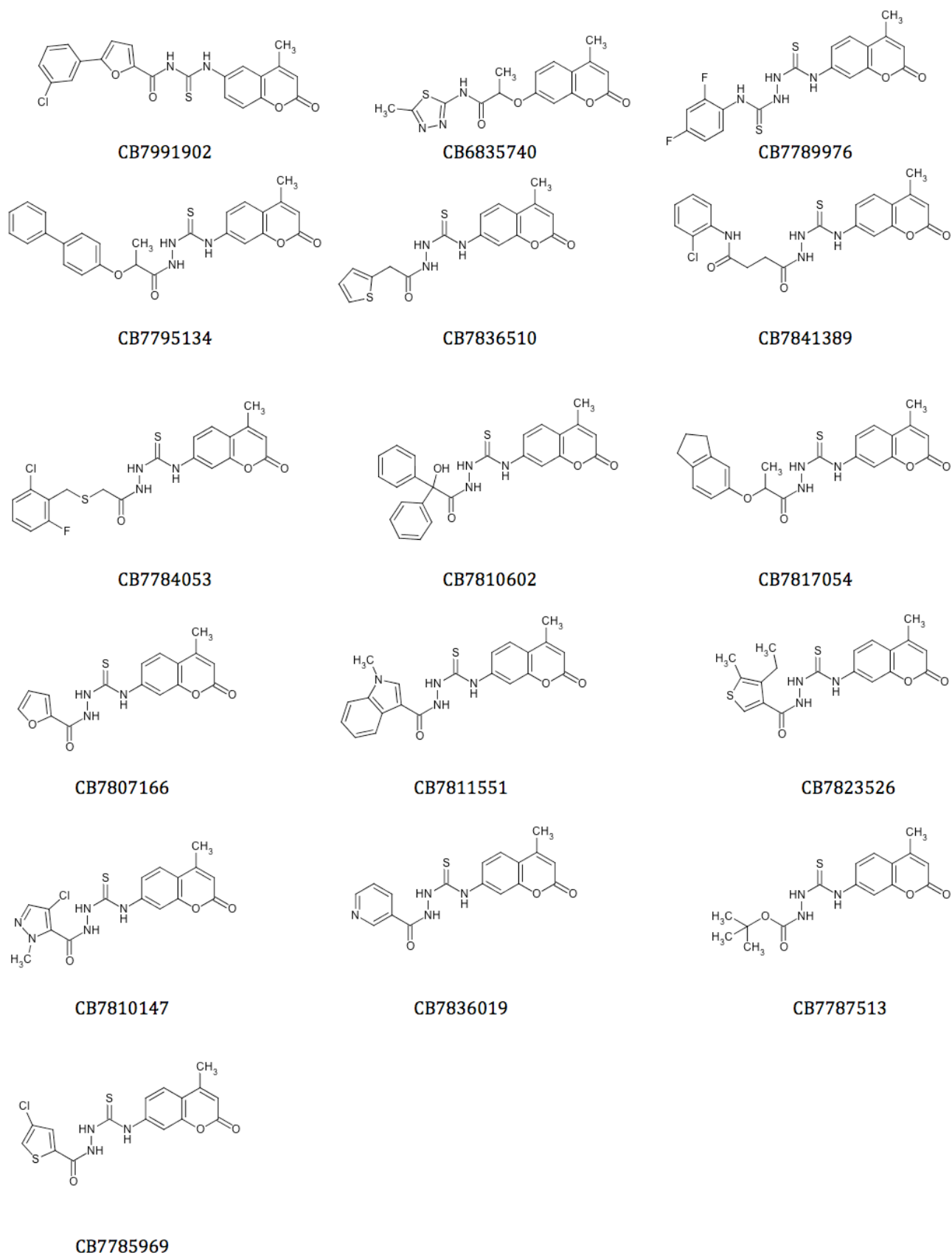


Figure 64: Coumarine derivatives obtained through the similarity search virtual screening using compound **CB7785969** as query structure.

The selected compounds were tested in an *in vitro* peroxidase assay and showed an inhibitory activity on LSD1 (table 21). However some of the compounds showed an abnormal LSD1 inhibition of over 200% in the peroxidase assay (in figure 65A is reported as example the inhibition values of compound **CB7785969** and compound **CB7787513**). The reactivity of those compounds was proved to be mainly towards hydrogen peroxide itself (H₂O₂) (in figure 65B is reported as example the reactivity of compound **CB7785969** and compound **CB7787513**). Therefore the compounds were incubated with only H₂O₂ (5 μM) and the peroxide concentration was detected. A concentration dependent change in signal was observed suggesting that these compounds react with hydrogen peroxide, which obfuscates potential LSD1 inhibition. When compounds **CB7785969** and **CB7787513** were tested in the DELFIA® assay no LSD1 inhibition was detected, which shows that these compounds have no effect on LSD1 demethylation at a concentration of 20 μM. Also compounds **CB5325408**, **CB6670945**, and some of the new derivatives found through the second virtual screening (table 21) were tested in the DELFIA® assay, showing in the best case (compound **CB7991902** and **CB6835740**) an inhibitory activity of 20% at a concentration of 20 μM, which is not enough to consider the derivatives as promising LSD1 inhibitors.

Table 21: *in vitro* peroxidase assay activity for the new coumarine derivatives obtained with the similarity search virtual screening procedure. For some of them also the activity values concerning the *in vitro* DELFIA assay are reported.

Compound	Inhibition @ conc [μM]	
	Peroxidase Assay	DELFIA Assay
CB7799868	124.1% @ 50 μM	//
CB7930465	32.7% @ 50 μM	//
CB7788596	118.5% @ 50 μM	//
CB7821436	105.6% @ 50 μM	//
CB7800770	123% @ 50 μM	//
CB7798329	97.2% @ 50 μM	//
CB7824193	116.5% @ 50 μM	//
CB7799353	89.1% @ 50 μM	//
CB7833564	112.2% @ 50 μM	//
CB7991902	61.4 % @ 50 μM	~20% @ 20 μM
CB6835740	65.6% @ 50 μM	~20% @ 20 μM
CB7789976	90.1% @ 5 μM	//
CB7795134	125.9% @ 50 μM	//

Results and Discussion: Coumarine derivatives

CB7836510	113.5% @ 50 μ M	//
CB7841389	137.2% @ 50 μ M	//
CB7784053	138.4% @ 50 μ M	//
CB7810602	152% @ 50 μ M	//
CB7817054	57.6% @ 5 μ M	//
CB7807166	173.5% @ 50 μ M	//
CB7785969	188.8% @ 50 μ M	//
CB7811551	90.2% @ 5 μ M	Inactive @ 20 μ M
CB7823526	160.8% @ 50 μ M	//
CB7810147	163.0% @ 50 μ M	//
CB7836019	145.2% @ 50 μ M	//
CB7787513	83.7% @ 5 μ M	Inactive @ 20 μ M

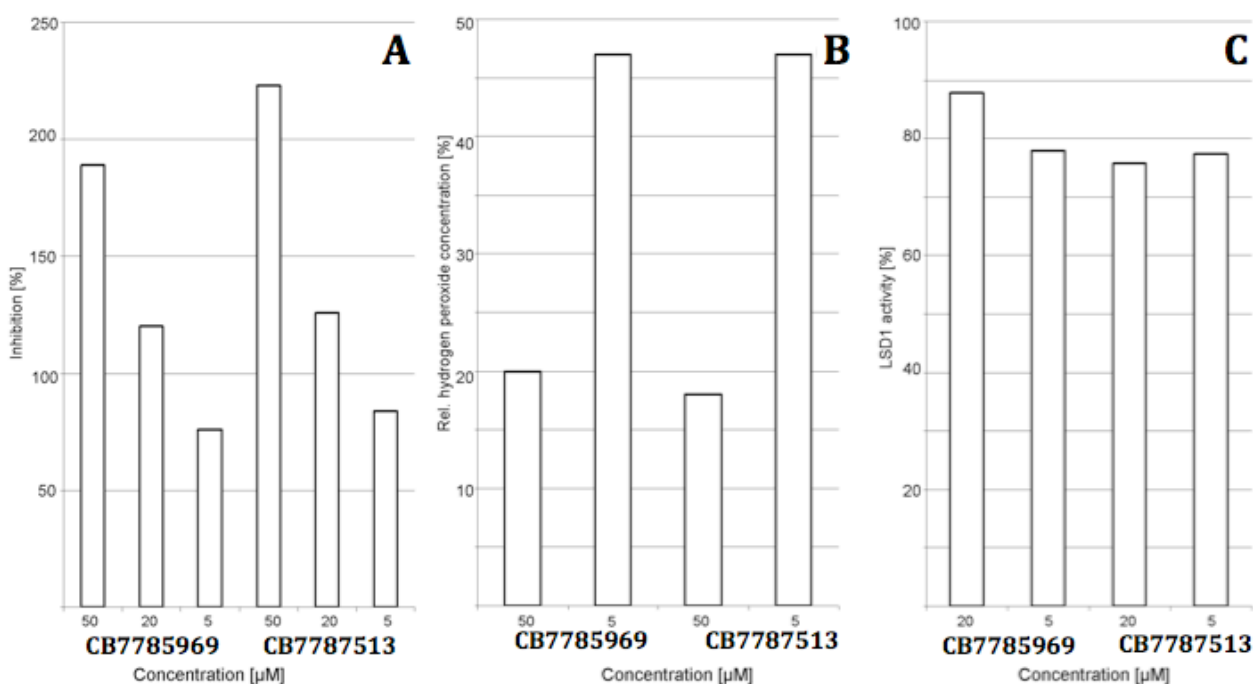


Figure 65: A Apparent inhibition potency of compounds **CB7785969** and **CB7787513** against LSD1 using the peroxidase assay. B H_2O_2 consumption of compounds **CB7785969** and **CB7787513**; 100% represents no consumption; 0% full consumption. C Potency of **CB7785969** and **CB7787513** against LSD1 in DELFIA assay.

3.5.4 Discussion

Since namoline (Chapter 3.4) has been identified as LSD1 inhibitor based on a protein similarity analysis followed by focussed library screening of a MAO based chromane library, we applied a similarity based screening to look for new potential LSD1 inhibitors among coumarin derivatives.

The analysis of the available X-ray structure of MAO-B co-crystallized with a coumarine derivative and the docking poses obtained for the virtual screening hit **CB7785969** and its derivatives (figures 66 and 67) at LSD1 showed similar interactions. These compounds interact with the FAD cofactor as well as with conserved residues of the substrate binding pocket. A similar orientation of the coumarine ring within the binding pocket was observed (figures 66 and 67). Due to the structural similarity and the conserved binding mode it was suggested that compound **CB7785969** and its derivatives might be active as LSD1 inhibitors. However, this was not supported by the newly developed DELFIA® assay. Due to the larger binding cavity of LSD1 (1700 Å) compared to MAO-B (637 Å) (Chapter 3.1), it is possible that these new compounds adopt another orientation, which is not sufficient to block the enzymatic activity. Another reason for the inactivity of the identified coumarine derivatives compared to the known MAO-B inhibitor might be the higher polarity of the VS hits (logP differences between coumarin co-crystallized into MAO-B active cavity (4.26 logP(o/w)) and VS hits **CB7785969** (3.33 logP(o/w)) and **CB7787513** (2.83 logP(o/w))).

Finally it has been demonstrated that DELFIA® assay (Chapter 2) can be used to test reference inhibitors and potential new LSD1 inhibitors that were identified from a virtual screening campaign. The developed assay showed that the identified coumarins were false positives.

CONCLUSION AND OUTLOOK

4

As a very new anticancer potential target LSD1 did not carry any information about binding small inhibitors at the beginning of the work. For this reason the strong connection between *in vitro* and *in silico* studies conducted for the whole thesis was crucial for the identification of novel active LSD1 irreversible and reversible inhibitors.

The computational method mainly conducted during the work has been the docking procedure. By means of docking studies suitable binding modes of the active inhibitors on LSD1 were generated using several datasets. Docking approaches are the fastest computational methods able to allocate small molecules inside a receptor binding cavity according to suitable molecular interactions between the environment of the protein and the functional groups present in the active compounds. During the work Gold,⁸⁷ Glide⁸⁷ and ParaDockS¹⁰² were the docking softwares used, even if none of their scoring function was able to properly rank the docking solutions according to the *in vitro* activity data on LSD1. This was mainly caused by the type of interaction inside LSD1 binding cavity: small inhibitors inside an extremely huge cavity namely the AOD of LSD1, which is 1700 Å³,⁴⁴ resulting in a multitude of poses ranked as putative binding mode in every docking run (due to the limitation of the scoring function). To overcome this problem docking studies with constraints resulted in a good compromise for the dataset formed by covalent inhibitors carrying a well-known reactive group able to covalently bind the isoalloxazine ring of the LSD1 cofactor, for instance the propargylamine derivatives (Chapter 3.2).

Through ligand-based virtual screening procedures two new LSD1 active propargylamine derivatives and three new LSD1 active chromane derivatives have been identified. Concerning the performed structure-based VS procedure (Chapter 3.3) no suitable results were found; the compounds obtained were inactive when tested at LSD1. Even if some of the compounds of the obtained dataset still have to be tested on the epigenetic target, the main limitation of this VS approach was due to the uncertainty of the binding mode found. This uncertainty was also coupled with the uncertainty of the mode of action of the known active compounds, which are generally considered as reversible inhibitors, however the hypothesis that these compounds might covalently bound LSD1 FAD cofactor is also suggested (Chapter 3.3)). This is considered as a weakness of a

structure-based virtual screening procedure. For the derivatives obtained through VS, the presence of highly flexible compounds, which are hardly stabilized inside a huge, mostly neutral region of LSD1 binding cavity (Chapter 3.3) might be the reason of their inactivity, even if the *in silico* binding mode showed a strong ionic interaction between the guanidine group present in the compound and a negatively charged region present in the LSD1 binding pocket. A molecular dynamic simulation might be useful to understand the behaviour of the new compounds inside the LSD1 binding cavity.

Other techniques have been also explored during this work. Molecular dynamic (MD) simulations were widely used for the whole work. Through this technique most of the involved protein residues in substrate-binding were identified and their importance was confirmed by mutagenesis data found in literature (Chapter 3.1). MD methodology also allowed selecting between putative binding modes of polyamine inhibitors (Chapter 3.3), but classical MD showed also some restrictions in model generation. For instance using the chromane dataset (Chapter 3.4) it was not possible to assess the stability of the suggested binding mode, apparently due to the presence of polarizable and strong electrostatic effects that characterize the way of binding for these compounds. To overcome this restriction, more computational demanding, mixed quantum mechanics/molecular mechanics (QM/MM) MDs might prove to be a better methodology to assess the stability of the proposed binding mode.

Binding free energy calculations allowed carrying out a partially descriptive structure-based model using chromane preliminary dataset (Chapter 3.4). The calculations were carried out using the minimized LSD1-chromane inhibitor complex generated from docking studies (single trajectory approach, see Chapter 2). These allowed a good compromise between accuracy and calculation time required for every LSD1-chromane inhibitor complex (Chapter 3.4). MM-PBSA and MM-GBSA calculations showed no reliable results, most probably due to the fact that they are not taking polarizable effects into consideration, and on the other hand, we should consider that a single trajectory approach might not be the best set up for this type of calculations. QM/MM-GBSA calculation with QM region increased to 6 Å around the ligand using AM1 as pseudopotential was found to be the most reliable approach for the differentiation of putative binding modes inside LSD1 and for the discrimination between highly active and inactive chromane inhibitors. The generated QM/MM-GBSA model was also used to predict the activity for some new *in silico* generated chromane derivatives. They showed a calculated binding free energy in the same order of magnitude as the most active compounds present in the initial chromane dataset, which might be

indicative for putative LSD1 inhibition. However, here the identification of new chromane scaffold-based inhibitors with activity data spanning around two or more orders of magnitude still remains the first mandatory step in order to generate more descriptive and predictive models. Then more accurate and computational time demanding binding free energy calculation methods (like LIE or TI) might be the best way to improve the model. Once a suitable dataset has been generated, also a ligand-based approach (for example a 3D-QSAR technique) might represent a good strategy for the identification of new non-covalent chromane-scaffold-based inhibitors.

During the work, the selection of a suitable assay for the activity detection of novel LSD1 inhibitors was a key step in order to avoid false positive/negative results. We faced the problem when two entire dataset of compounds (coumarine derivatives (Chapter 3.5) and a second dataset of chromane derivatives obtained from a similarity search campaign using compound **35a** as query structure (Appendix B table 3) were identified as positive under peroxidase assay condition, but when tested in *in vitro* DELFIA® assay they did not show any inhibitory activity on LSD1. More deep observations showed that these new derivatives might react with H₂O₂ (the secondary product generated by the LSD1 enzymatic activity, see Chapter 1) when tested in *in vitro* peroxidase assay, obfuscating the potential LSD1 inhibition (Chapter 3.4, Chapter 3.5 and Appendix B table 3).

Finally, it has been demonstrated that even if nowadays there are only little information about interactions between LSD1 binding cavity and small non-covalent inhibitors (no LSD1 crystal structures with small non-covalent inhibitor has been deposited in the protein crystal structure database yet), a structure-based approach might be the best approach to identify new LSD1 active inhibitors. With our work we mainly focused on the development of new LSD1 *in silico* leads which, through future optimization, might become potential drug candidates for the treatment of hormone-dependent cancer targeting on epigenetic mechanism.

Conclusion and Outlook

It has been demonstrated that the epigenetic code might be involved in several types of human diseases and that targeting the proteins involved in the generation/regulation of the code may represent a new pharmacological strategy against some of the most important diseases of the XXI century, such as cancer. For instance, some inhibitors of DNA methylation, or histone deacetylation (HADC) are already approved for the clinical use of cancer therapy.^{2,9,10}

Lysine histone demethylase one (LSD1) is one of the proteins that are involved in the regulation of the epigenetic code, it belongs to the flavin oxidase enzyme family and it shares sequence similarity with other FAD-dependent amine oxidases such as MAOs and PAO. The role of LSD1 is to remove methyl flags on mono- or dimethylated lysine on histone number three (H3K n Me1/2). According to the targeted lysine, LSD1 shows a dual role in gene regulations: when it acts on the active mark H3K4me1/2, it causes inactivation of the gene transcription, while the concomitant interaction with AR causes LSD1 substrate switching, demethylating the inactive mark H3K9me1/2 and promoting hormone-induced gene activation, especially in those tissues where AR has a key physiological role. Up to date several data also confirmed the activity of LSD1 on non-epigenetic targets, such as p53 protein and the involvement of LSD1 in several cancer diseases. The main aim of the work was to assess the structural information available on LSD1 and develop new leads able to show inhibitory activity on the epigenetic target.

Initially MDs and mutagenesis data present in literature, allowed identifying some key residues involved in the binding of the H3K4me1/2 substrate (such as Cys360, Asp375, Glu379, Asn540, Gln554, Asp555, Asp556, His564, Tyr761, Ala809, and Thr810). Then several molecular dataset carrying preliminary inhibitory data on LSD1 were used as a starting point for the generation of new putative lead compounds.

We described four putative binding modes (each one specific for the molecular dataset used) inside LSD1 active cavity by structural studies, docking studies and MDs. Concerning the chromane derivatives dataset, a QM/MM-GBSA binding free energy protocol, showed a rough differentiation

Summary

between active and inactive derivatives inside LSD1. This allowed the *in silico* identification of seven new chromane scaffolds as putatively active inhibitors on LSD1.

Using ligand-based VS methods two new LSD1 inhibitors, propargylamine derivatives with activity values in the micromolar range were identified. While nine propargylamine derivatives were found as putative LSD1 inhibitors by a multistep structure-based VS procedure.

All the computational studies conducted during this work allowed identifying new LSD1 features and new active leads, confirming the potentiality of CAMD in medicinal chemistry as potent tool for the identification of new active compounds.

BIBLIOGRAPHY

1. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33**, 245–54 (2003).
2. Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K. & Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Discov.* **11**, 384–400 (2012).
3. Meaney, M. J. Epigenetics and the biological definition of gene x environment interactions. *Child Dev.* **81**, 41–79 (2010).
4. Kelly, T., Carvalho, D. De & Jones, P. Epigenetic modifications as therapeutic targets. *Nat. Biotechnol.* 1069–1078 (2010).
5. Meaney, M. & Ferguson-Smith, A. Epigenetic regulation of the neural transcriptome: the meaning of the marks. *Nat. Neurosci.* **13**, 1313–1318 (2010).
6. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat. Biotechnol.* **28**, 1057–1068 (2010).
7. Smith, S. S. Biological implications of the mechanism of action of human DNA (cytosine-5)methyltransferase. *Prog. Nucleic Acid Res. Mol. Biol.* **49**, 65–111 (1994).
8. Jinek, M. & Doudna, J. a. A three-dimensional view of the molecular machinery of RNA interference. *Nature* **457**, 405–12 (2009).
9. Das, P. M. & Singal, R. DNA methylation and cancer. *J. Clin. Oncol.* **22**, 4632–42 (2004).
10. Dokmanovic, M., Clarke, C. & Marks, P. a. Histone deacetylase inhibitors: overview and perspectives. *Mol. Cancer Res.* **5**, 981–9 (2007).
11. Luger, K., Mäder, a W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–60 (1997).
12. C. David Allis; Thomas Jenuwein; Danny Reinberg. *Epigenetics*. (Cold Spring Harbor Laboratory Press, 2006).
13. Heinke, R., Carlino, L., Kannan, S., Jung, M. & Sippl, W. Computer- and structure-based lead design for epigenetic targets. *Bioorg. Med. Chem.* **19**, 3605–15 (2011).
14. Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, A. P., Casero, R. A., Shi, Y. Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **119**, 941–953 (2004).

Bibliography

15. Metzger, E., Wissmann, M., Yin, N., Müller, J. M., Schneider R., Peter, A. H. F. M., Günther, T., Buettner, R., Schüle, R. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**, 436–9 (2005).
16. Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., Zhang, Y. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811–6 (2006).
17. Forneris, F., Binda, C., Vanoni, M. A., Mattevi, A. & Battaglioli, E. Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. *FEBS Lett.* **579**, 2203–7 (2005).
18. Stavropoulos, P., Blobel, G. & Hoelz, A. Crystal structure and mechanism of human lysine-specific demethylase-1. *Nat. Struct. Mol. Biol.* **13**, 626–32 (2006).
19. Fraaije, M. W. & Mattevi, a. Flavoenzymes: diverse catalysts with recurrent features. *Trends Biochem. Sci.* **25**, 126–32 (2000).
20. Nakamura, Y. Umehara, T., Nakano, K., Jang, M. K., Shirouzu, M., Morita, S., Uda-Tochio, H., Hamana, H., Terada, T., Adachi, N., Matsumoto, T., Tanaka, A., Horikoshi, M., Ozato, K., Padmanabhan, B., Yokoyama, S. Crystal structure of the human BRD2 bromodomain: insights into dimerization and recognition of acetylated histone H4. *J. Biol. Chem.* **282**, 4193–201 (2007).
21. Bannister, A. J., Schneider, R. & Kouzarides, T. Histone methylation: dynamic or static? *Cell* **109**, 801–6 (2002).
22. Forneris, F., Binda, C., Battaglioli, E. & Mattevi, A. LSD1: oxidative chemistry for multifaceted functions in chromatin regulation. *Trends Biochem. Sci.* **33**, 181–9 (2008).
23. Giorgio, M., Trinei, M., Migliaccio, E. & Pelicci, P. G. Hydrogen peroxide: a metabolic by-product or a common mediator of ageing signals? *Nat. Rev. Mol. Cell Biol.* **8**, 722–8 (2007).
24. Yang, M., Gocke, C. B., Luo, X., Borek, D., Tomchick, D. R., Machius, M., Otwinowski, Z., Yu, H. Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. *Mol. Cell* **23**, 377–87 (2006).
25. Forneris, F., Binda, C., Adamo, A., Battaglioli, E. & Mattevi, A. Structural basis of LSD1-CoREST selectivity in histone H3 recognition. *J. Biol. Chem.* **282**, 20070–4 (2007).
26. Lee, M. G., Wynder, C., Cooch, N. & Shiekhhattar, R. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* **437**, 432–5 (2005).
27. Tian, X. & Fang, J. Current Perspectives on Histone Demethylases. *Acta Biochim. Biophys. Sin. (Shanghai)*. **39**, 81–88 (2007).
28. Wysocka, J. & Milne, T. A. Taking LSD1 to a New High. *Cell* **122**, 653–8 (2005).
29. Klose, R. J. & Zhang, Y. Regulation of histone methylation by demethylination and demethylation. *Nat. Rev. Mol. Cell Biol.* **8**, 307–18 (2007).

Bibliography

30. Yang, M., Culhane, J. C., Szewczuk, L. M., Gocke, C. B., Brautigam, C. A., Tomchick, D. R., Machius, M., Cole, P. A., Yu, H. Structural basis of histone demethylation by LSD1 revealed by suicide inactivation. *Nat. Struct. Mol. Biol.* **14**, 535–9 (2007).
31. Hakimi, M.-A., Bochar, D. A., Chenoweth, J., Lane, W.S., Mandel, G., Shiekhattar, R. A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7420–5 (2002).
32. Shi, Y.-J., Matson, K., Lan, F., Iwase, S., Baba, T., Shi, Y. Regulation of LSD1 histone demethylase activity by its associated factors. *Mol. Cell* **19**, 857–64 (2005).
33. Chen, C. D., Welsbie, D. S., Tran, C., Beang, S. H., Chen, R., Vessela, R., Rosenfel, M. G., Sawyers, C. L. Molecular determinants of resistance to antiandrogen therapy. *Nat. Med.* **10**, 33–9 (2004).
34. Kahl, P., Gullotti, L., Heukamp, L. C., Wolf, S., Friedrichs, N., Vorreuther, R., Solleder, G., Bastian, P. J., Ellinger, J., Metzger, E., Schüle, R., Buettner, R. Androgen receptor coactivators lysine-specific histone demethylase 1 and four and a half LIM domain protein 2 predict risk of prostate cancer recurrence. *Cancer Res.* **66**, 11341–7 (2006).
35. Metzger, E., Yin, N. & Wissmann, M. Phosphorylation of histone H3 at threonine 11 establishes a novel chromatin mark for transcriptional regulation. *Nat. cell ...* **10**, 53–60 (2007).
36. Wissmann, M., Yin, N., Müller, J. M., Greschik, H., Fodor, B. D., Vogler, C., Schneider, R., Günther, T., Buettner, R., Metzger, E., Schüle, R. Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nat. Cell Biol.* **9**, 347–53 (2007).
37. Metzger, E. *et al.* Phosphorylation of histone H3T6 by PKC β I controls demethylation at histone H3K4. *Nature* **464**, 792–796 (2010).
38. Huang, J., Sengupta, R., Espejo, A. B., Lee, M. G., Dorsey, J. A., Richter, M., Opravil, S., Shiekhattar, R., Bedford, M. T., Jenuwein, T., Berger, S. L. p53 is regulated by the lysine demethylase LSD1. *Nature* **449**, 105–8 (2007).
39. Wang, J., Hevi, S., Kurash, J. K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G., Gaudet, F., Chen, T. The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat. Genet.* **41**, 125–9 (2009).
40. Karytinis, A., Forneris, F., Profumo, A., Ciossani, G., Battaglioli, E., Binda, C., Mattevi, A. A novel mammalian flavin-dependent histone demethylase. *J. Biol. Chem.* **284**, 17775–82 (2009).
41. Binda, C., Valente, S., Romanenghi, M., Pilotto, S., Cirilli, R., Karytinis, A., Ciossani, G., Botrugno, O. A., Forneris, F., Tardugno, M., Edmondos, D. E., Minucci, S., Mattevi, A., Mai, A. Biochemical, structural, and biological evaluation of tranlylcypramine derivatives as inhibitors of histone demethylases LSD1 and LSD2. *J. Am. Chem. Soc.* **132**, 6827–33 (2010).

Bibliography

42. Hoffmann, I., Roatsch, M., Schmitt, M. L., Carlino, L., Pippel, M., Sippl, W., Jung, M. The role of histone demethylases in cancer therapy. *Mol. Oncol.* **6**, 683–703 (2012).
43. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* **7**, 715–27 (2006).
44. Forneris, F., Battaglioli, E., Mattevi, A. & Binda, C. New roles of flavoproteins in molecular cell biology: histone demethylase LSD1 and chromatin. *FEBS J.* **276**, 4304–12 (2009).
45. Willmann, D., Lim, S., Wtzel, S., Metger, E., Jandausch, A., Wilk, W., Jung, M., Forne, I., Imhof, A., Janzer, A., Kirfel, J., Wildmann, H., Schüle R., Buettner, R. Impairment of prostate cancer cell growth by a selective and reversible lysine-specific demethylase 1 inhibitor. *Int. J. Cancer* **131**, 2704–2709 (2012).
46. Wang, J., Lu, F., Ren, Q., Sun, H., Xu, Z., Lan, R., Liu, Y., Ward, D., Quan, J., Ye, T., Zang, H. Novel histone demethylase LSD1 inhibitors selectively target cancer cells with pluripotent stem cell properties. *Cancer Res.* **71**, 7238–49 (2011).
47. Yamada, M. & Yasuhara, H. Clinical pharmacology of MAO inhibitors: safety and future. *Neurotoxicology* **25**, 215–21 (2004).
48. Carotti, A. & Catto, M. Synthesis and monoamine oxidase inhibitory activity of new Pyridazine-, Pyrimidine- and 1, 2, 4-Triazine-Containing tricyclic derivatives. *J. Med. Chem.* **50**, 5364–5371 (2007).
49. Riederer, P., Danielczyk, W. & Grünblatt, E. Monoamine oxidase-B inhibition in Alzheimer's disease. *Neurotoxicology* **25**, 271–7 (2004).
50. Pålhagen, S., Heinonen, E., Häggglud, J., Kaugesaar, T., Maki-Ikola, O., Palm, R. Selegiline slows the progression of the symptoms of Parkinson disease. *Neurology* **66**, 1200–6 (2006).
51. Schenk, T., Chen, W. C., Göllner, S., Howell, L., Jin, L., Habestreit, K., Klein, H.-U., Popescu, A. C., Burnett, A., Casero Jr., R. A., Marton, L., Wooster, P., Minden, M. D., Dugas, M., Wang, J. C. Y., Dick, E. J., Müller-Tidow, C., Petrie, K., Zelent, A. Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia. *Nat. Med.* **18**, 605–612 (2012).
52. Kauffman, E. C., Robinson, B. D., Downes, M., Powell, L. G., Lee, M. M., Scherr, D. S., Gudas, L. J., Mongan, N. P. Role of androgen receptor and associated lysine-demethylase coregulators, LSD1 and JMJD2A, in localized and advanced human bladder cancer. *Mol. Cell. Oncol.* **50**, 931–944 (2012).
53. Lee, M. G., Wynder, C., Schmidt, D. M., McCafferty, D. G. & Shiekhatar, R. Histone H3 lysine 4 demethylation is a target of nonselective antidepressive medications. *Chem. Biol.* **13**, 563–7 (2006).
54. Schmidt, D. M. Z. & McCafferty, D. G. trans-2-Phenylcyclopropylamine is a mechanism-based inactivator of the histone demethylase LSD1. *Biochemistry* **46**, 4408–16 (2007).

Bibliography

55. Benelkebir, H., Hodgkinson, C., Duriez, P. J., Hyden, A. L., Bulleid, R. A., Crabb, S. J., Packham, G., Ganesan, A. Enantioselective synthesis of tranlycypromine analogues as lysine demethylase (LSD1) inhibitors. *Bioorg. Med. Chem.* **19**, 3709–16 (2011).
56. Guibourt, N., Ortega, M. A. & Castro-Palomino, L. J. Patent application. *WO2010084160 A1 Oryzon Gen*, (2010).
57. Harris, W. J., Huang, X., Lynch, J. T., Spencer, G. J., Hitchin, J. R., Li, Y., Ciceri, F., Blaser, J. C., Greystoke, B. F., Jordan, A. M., Miller, C. J., Ogilvie, D. J., Somervaille, T. C. P. The histone demethylase KDM1A sustains the oncogenic potential of MLL-AF9 leukemia stem cells. *Cancer Cell* **21**, 473–87 (2012).
58. Guibourt, N., Ortega, M. A. & Castro-Palomino, L. J. Patent application. *WO2010043721 Oryzon Gen*, (2010).
59. Guibourt, N., Ortega, M. A., Castro-Palomino, L. J. & Fyfe, M. C. T. Patent application. *WO 2011035941 Oryzon Gen*, (2011).
60. Suzuki, T. & Miyata, N. Lysine demethylases inhibitors. *J. Med. Chem.* **54**, 8236–50 (2011).
61. Ueda, R., Suzuki, T., Mino, K., Tsumoto, H., Nakagawa, H., Hasegawa, M., Sasaki, R., Mizukami, T., Miyara, N. Identification of cell-active lysine specific demethylase 1-selective inhibitors. *J. Am. Chem. Soc.* **131**, 17536–7 (2009).
62. Mimasu, S., Umezawa, N., Sato, S., Higuchi, T., Umehara, T., Yokoyama, S. Structurally designed trans-2-phenylcyclopropylamine derivatives potently inhibit histone demethylase LSD1/KDM1. *Biochemistry* **49**, 6494–503 (2010).
63. Pollock, J. a, Larrea, M. D., Jasper, J. S., McDonnell, D. P. & McCafferty, D. G. Lysine-Specific Histone Demethylase 1 Inhibitors Control Breast Cancer Proliferation in ER α -Dependent and -Independent Manners. *ACS Chem. Biol.* **7**, 1221–1231 (2012).
64. Hu, Q., Kwon, Y.-S., Nunez, E., Cardamone, M. D., Hutt, K. R., Ohgi, K. A., Garcia-Bassets, I., Rose, D. W., Glass, K. C., Rosenfeld, M. G., Fu, X.-D. Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. *Proc Natl Acad Sci U S A.* **105**, 19199–19204 (2008).
65. Manetti, F., Cona, A., Angeli, L., Mugnaini, C., Raffi, F., Capone, C., Dreassi, E., Zizzarri, A. T., Tisi, A., Federico, R., Botta, M. Synthesis and biological evaluation of guanidino compounds endowed with subnanomolar affinity as competitive inhibitors of maize polyamine oxidase. *J. Med. Chem.* **52**, 4774–85 (2009).
66. Ha, H. C., Woster, P. M., Yager, J. D. & Casero, R. a. The role of polyamine catabolism in polyamine analogue-induced programmed cell death. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 11557–62 (1997).
67. Hu, R. H. & Pegg, a E. Rapid induction of apoptosis by deregulated uptake of polyamine analogues. *Biochem. J.* **328** (Pt 1, 307–16 (1997).

Bibliography

68. Binda, C., Coda, A., Angelini, R., Federico, R., Ascenzi, P., Mattevi, A. A 30-angstrom-long U-shaped catalytic tunnel in the crystal structure of polyamine oxidase. *Structure* **7**, 265–76 (1999).
69. Bianchi, M., Polticelli, F., Ascenzi, P., Botta, M., Federico, R., Mariottini, P., Cona, A. Inhibition of polyamine and spermine oxidases by polyamine analogues. *FEBS J.* **273**, 1115–23 (2006).
70. Cona, A., Manetti, F., Leone, R., Corelli, F., Tavladoraki, P., Polticelli, F., Botta, M. Molecular basis for the binding of competitive inhibitors of maize polyamine oxidase. *Biochemistry* **43**, 3426–35 (2004).
71. Stránská, J., Sebelá, M., Tarkowski, P., Rehulka, P., Chmelik, J., Popa, I., Pec, P. Inhibition of plant amine oxidases by a novel series of diamine derivatives. *Biochimie* **89**, 135–44 (2007).
72. Federico, R., Leone, L., Botta, M., Binda, C., Angelini, R., Venturini, G., Ascenzi, P. Inhibition of pig liver and *Zea mays* L. polyamine oxidase: a comparative study. *J. Enzyme Inhib.* **16**, 147–55 (2001).
73. Huang, Y., Greene, E., Stewart, T. M., Goodwin, A. C., Baylin, S. B., Woster, P. M., Casero Jr., R.A. Inhibition of lysine-specific demethylase 1 by polyamine analogues results in reexpression of aberrantly silenced genes. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 8023–8 (2007).
74. Huang, Y., Stewart, T. M., Wu, Y., Baylin, S. B., Marton, L. J., Perkin, B., Jones, R. J., Woster, P. M., Casero Jr., R. A. Novel oligoamine analogues inhibit lysine-specific demethylase 1 and induce reexpression of epigenetically silenced genes. *Clin. Cancer Res.* **15**, 7217–28 (2009).
75. Schmitt, M. L. Synthese und Testung neuer Inhibitoren der Lysinspezifischen Demethylase 1. 214 (2013).
76. Bernstein, F. & Koetzle, T. The Protein Data Bank: a computer based archival file for macromolecular structures. *J Mol Biol.* **185**, 584–591 (1978).
77. Leach, A. R. *Molecular Modelling, Principles and application.* (2001).
78. Warshel, A. & Levitt, M. Theoretical studies of enzymatic reactions. *J. Mol. Biol.* 227–249 (1976).
79. Höltje, H.D., Sippl, W., Rognan, D., and Folkers, G. *Molecular Modelling: Basic principles and applications.* (Wiley-VCH, 2003).
80. McCammon, J.A., Gelin, B.R., Karplus, M. Dynamics of folded protein. *Nature* **267**, 585–590 (1977).
81. Duan, Y. Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution. *Science (80-)*. **282**, 740–744 (1998).

Bibliography

82. Gunsteren, W. van & Mark, A. Validation of molecular dynamics simulation. *J. Chem. Phys.* **108**, 6109–6116 (1998).
83. Case, D.A., Darde, T.A., Cheatham III, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Pearlman, D.A., Crowley, M., Walker, R.C., Zhang, W., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Wong, K.F., Paesani, F., Wu, X., Brozell, S., Ts, M. & D.H., Schafmeister, C., Ross, W.S., Kollman, P. A. Amber 9. *Univ. California, San Fr.* (2006).
84. Lindahl, E., Hess, B. & Spoel, D. Van Der. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Mol. Model. Annu.* **7**, 306–316 (2001).
85. Brooks, B. & Bruccoleri, R. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
86. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161**, 269–88 (1982).
87. Jones, G., Willett, P., Glen, R. C., Leach, a R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–48 (1997).
88. Baxter, C. a, Murray, C. W., Clark, D. E., Westhead, D. R. & Eldridge, M. D. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins* **33**, 367–82 (1998).
89. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2010).
90. Friesner, R., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–49 (2004).
91. Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V & Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided. Mol. Des.* **11**, 425–45 (1997).
92. Meng, E. C., Shoichet, B. K. & Kuntz, I. D. Automated Docking. **13**, 505–524 (1992).
93. DeWitte, R. S. & Shakhnovich, E. I. SMOG: de Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence. *J. Am. Chem. Soc.* **118**, 11733–11744 (1996).
94. Ishchenko, A. V & Shakhnovich, E. I. SMAll Molecule Growth 2001 (SMoG2001): an improved knowledge-based scoring function for protein-ligand interactions. *J. Med. Chem.* **45**, 2770–80 (2002).
95. Gohlke, H., Hendlich, M. & Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **295**, 337–56 (2000).

Bibliography

96. Muegge, I. & Martin, Y. C. Articles A General and Fast Scoring Function for Protein-Ligand Interactions : A Simplified Potential Approach. *J. Med. Chem.* **42**, 791–804 (1999).
97. Böhm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided. Mol. Des.* **8**, 243–56 (1994).
98. Weiner, S. & Kollman, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784 (1984).
99. Shoichet, B. K., Leach, a R. & Kuntz, I. D. Ligand solvation in molecular docking. *Proteins* **34**, 4–16 (1999).
100. Miyamoto, S. & Kollman, P. a. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. *Proteins* **16**, 226–45 (1993).
101. Jorgensen, W. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).
102. Meier, R., Pippel, M., Brandt, F., Sippl, W. & Baldauf, C. ParaDockS: a framework for molecular docking with population-based metaheuristics. *J. Chem. Inf. Model.* **50**, 879–89 (2010).
103. Kennedy, J. & Eberhart, R. Particle swarm optimization. *Proc. IEE Int. Conf. Neural Networks* **4**, 1942–1948 (1995).
104. Walters, W., Stahl, M. & Murcko, M. Virtual Screening – an overview. *Drug Discov. Today* **3**, 160–178 (1998).
105. Stahura, F. & Bajorath, J. Virtual screening methods that complement HTS. *Comb. Chem. Throughtput Screen.* **7**, 259–269 (2004).
106. Guner, O., Clement, O. & Kurogi, Y. Pharmacophore modeling and three dimensional database searching for drug design using catalyst: recent advances. *Curr. Med. Chem.* **11**, 2991–3005 (2004).
107. Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov. Today* **11**, 580–94 (2006).
108. Lipinski, C. a, Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).
109. Muresan, S. & Sadowski, J. “In-house likeness”: comparison of large compound collections using artificial neuronal networks. *J. Chem. Inf. Model.* **45**, 888–893 (2005).
110. Pearce, B. & Sofia, M. An empirical process for the design of high-throughput screening deck filters. *J. Chem. Inf. Model.* **46**, 1060–1068 (2006).

Bibliography

111. Brandsdal, B., Österberg, F. & Almlöf, M. Free energy calculations and ligand binding. *Adv. Protein Chem.* **66**, 123–158 (2003).
112. Jain, A. N. Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. *J. Comput. Aided. Mol. Des.* **10**, 427–440 (1996).
113. Tembe, L. & Mccammon, J. A. LIGAND-RECEPTOR INTERACTIONS. *Comput. Chem.* **8**, 281–283 (1984).
114. Kollman, P., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., Cheatham, T. E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–97 (2000).
115. Hansson, T., Marelus, J. & Aqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aided. Mol. Des.* **12**, 27–35 (1998).
116. Aqvist, J., Medina, C. & Samuelsson, J. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **7**, 385–391 (1994).
117. Honig, B. & Nicholls, a. Classical electrostatics in biology and chemistry. *Science* **268**, 1144–9 (1995).
118. Srinivasan, J., Cheatham III, T. E., Cieplak, P., Kollmann, P. A. & Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *Am. Chem. Soc.* **120**, 9401–9409 (1998).
119. Wang, J., Morin, P., Wang, W. & Kollman, P. a. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.* **123**, 5221–30 (2001).
120. Huo, S., Wang, J., Cieplak, P., Kollman, P. a & Kuntz, I. D. Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design. *J. Med. Chem.* **45**, 1412–9 (2002).
121. Dubey, K. D. & Ojha, R. P. Binding free energy calculation with QM/MM hybrid methods for Abl-Kinase inhibitor. *J. Biol. Phys.* **37**, 69–78 (2011).
122. Lin, H. & Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **117**, 185–199 (2006).
123. Murphy, R. B., Philipp, D. M. & Friesner, R. A. Mechanics (QM / MM) Method for Large-Scale Modeling of Chemistry in Protein Environments. **21**, 1442–1457 (2000).
124. Hayik, S. A., Dunbrack, R. & Merz, K. M. Mixed Quantum Mechanics / Molecular Mechanics Scoring Function To Predict Protein-Ligand Binding Affinity. *J Chem Theory Comput.* **6**, 3079–3091 (2010).

Bibliography

125. Hauser, A.-T., Jung, M. & Jung, M. Assays for Histone Deacetylases. *Curr. Top. Med. Chem.* **9**, 227–234 (2009).
126. Remiszewski, S.-W. The Discovery of NVP-LAQ824: From Concept to Clinic. *Curr. Med. Chem.* **10**, 2393–2402 (2003).
127. Zhou, M., Diwu, Z., Panchuk-voloshina, N. & Haugland, R. P. A Stable Nonfluorescent Derivative of Resorufin for the Fluorometric Determination of Trace Hydrogen Peroxide: Applications in Detecting the Activity of Phagocyte NADPH Oxidase and Other Oxidases. *Anal. Biochem.* **168**, 162–168 (1997).
128. Forneris, F., Binda, C., Dall'aglio, A., Fraajie, M., Battaglioli, E., Mattevi, A. A highly specific mechanism of histone H3-K4 recognition by histone demethylase LSD1. *J. Biol. Chem.* **281**, 35289–95 (2006).
129. Lizcano, J. M., Unzeta, M. & Tipton, K. F. A spectrophotometric method for determining the oxidative deamination of methylamine by the amine oxidases. *Anal. Biochem.* **286**, 75–9 (2000).
130. Gauthier, N., Caron, M., Pedro, L., Arcand, M., Blouin, J., Labonté, A., Normand, C., Paquet, V., Rodenbrok, A., Roy, M., Rouleau, N., Baudet, L., Padrós, J., Rodriguez-Suarez, R. Development of homogeneous nonradioactive methyltransferase and demethylase assays targeting histone H3 lysine 4. *J. Biomol. Screen.* **17**, 49–58 (2012).
131. Chen, Y., Yang, Y., Wang, F., Wan, K., Yamane, K., Zhang, Y., Lei, M. Crystal structure of human histone lysine-specific demethylase 1 (LSD1). *Proc. Natl. Acad. Sci. U. S. A.* **103**, 13956–61 (2006).
132. Mimasu, S., Sengoku, T., Fukuzawa, S., Umehara, T. & Yokoyama, S. Crystal structure of histone demethylase LSD1 and tranlycypromine at 2.25 Å. *Biochem. Biophys. Res. Commun.* **366**, 15–22 (2008).
133. Yang, M., Culhane, J. C., Szewczuk, M. L., Jalili, P., Ball, H. L., Machius, M., Cole, P. A., Yu, H. Structural Basis for the Inhibition of the LSD1 Histone Demethylase by the Antidepressant trans-2-Phenylcyclopropylamine. *Biochemistry* **46**, 8058–8065 (2007).
134. Zibetti, C., Adamo, A., Binda, C., Forneris, F., Toffolo, E., Verpelli, C., Ginelli, E., Mattevi, A., Sala, C., Battaglioli, E. Alternative Splicing of the Histone Demethylase LSD1/KDM1 Contributes to the Modulation of Neurite Morphogenesis in the Mammalian Nervous System. *J. Neurosci.* **30**, 2521–2532 (2010).
135. Baron, R., Binda, C., Tortorici, M., McCammon, J. A. & Mattevi, A. Molecular mimicry and ligand recognition in binding and catalysis by the histone demethylase LSD1-CoREST complex. *Structure* **19**, 212–20 (2011).
136. Tortorici, M., Borrello, M. T., Tardugno, M., Chiarelli, L. R., Pilotto, S., Ciossani, G., Vellere, N. A., Bailey, S. G., Cowan, J., O'Connell M., Crabb, S. J., Packham, G., Mai, A., Baron, R., Ganesan, A., Mattevi, A. Protein Recognition by Short Peptide Reversible Inhibitors of the Chromatin-Modifying LSD1/CoREST Lysine Demethylase. *ACS Chem. Biol.* **8**, 1677–1682 (2013).

Bibliography

137. Klabunde, T. Chemogenomic approaches to drug discovery: similar receptors bind similar ligands. *Br. J. Pharmacol.* **152**, 5–7 (2007).
138. Holm, L., Kääriäinen, S., Rosenström, P. & Schenkel, a. Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24**, 2780–1 (2008).
139. Zhao, G., Bruckner, R. C. & Jorns, M. S. Identification of the oxygen activation site in monomeric sarcosine oxidase: role of Lys265 in catalysis. *Biochemistry* **47**, 9124–35 (2008).
140. Binda, C., Mattevi, A. & Edmondson, D. E. Structure-function relationships in flavoenzyme-dependent amine oxidations: a comparison of polyamine oxidase and monoamine oxidase. *J. Biol. Chem.* **277**, 23973–6 (2002).
141. Li, M., Binda, C., Mattevi, A. & Edmondson, D. E. Functional role of the “aromatic cage” in human monoamine oxidase B: structures and catalytic properties of Tyr435 mutant proteins. *Biochemistry* **45**, 4775–84 (2006).
142. Culhane, J. C. & Cole, P. a. LSD1 and the chemistry of histone demethylation. *Curr. Opin. Chem. Biol.* **11**, 561–8 (2007).
143. Culhane, J. C., Szewczuk, L. M., Liu, X., Da, G., Marmorstein, R., Cole, P. A. A mechanism-based inactivator for histone demethylase LSD1. *J. Am. Chem. Soc.* **128**, 4536–7 (2006).
144. Schmitt, M., Hauser, A.-T., Carlino, L., Pippel, M., Schulz-Fincke, J., Metzger, E., Willmann, D., Yiu, T., Barton, M., Schüle, R., Sippl, W., Jung, M. Non-peptidic propargylamines as inhibitors of lysine specific demethylase 1 (LSD1) with cellular activity. *J. Med. Chem.* **56** (18), 7334-7342 (2013).
145. Molecular Operating Environment (MOE) 2011.10. Chemical Computing Group Inc., Montreal, Quebec, Canada. (2006).
146. Hauser, A.-T., Bissinger, E.-M., Metzger, E., Rapenning, A., Bauer, U.-M., Mai, A., Schüle, R., Jung, M. Screening assays for epigenetic targets using native histones as substrates. *J. Biomol. Screen.* **17**, 18–26 (2012).
147. Manetti, F., Castagnolo, D., Raffi, F., Zizzarri, A. T., Rajamaki, S., D'Arezzo, S., Visca, P., Cona, A., Fracasso, M. E., Doria, D., Posteraro, B., Sanguinetti, M., Fadda, G., Botta, M. Synthesis of new linear guanidines and macrocyclic amidinourea derivatives endowed with high antifungal activity against *Candida* spp. and *Aspergillus* spp. *J. Med. Chem.* **52**, 7376–9 (2009).
148. Binda, C., Angelini, R., Federico, R., Ascenzi, P. & Mattevi, a. Structural bases for inhibitor binding and catalysis in polyamine oxidase. *Biochemistry* **40**, 2766–76 (2001).
149. Irwin, J. J. & Shoichet, B. K. ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. *J Chem Inf Model.* **45**, 177–182 (2005).
150. Wetzel, S., Wilk, W., Chammaa, S., Sperl, B., Roth, A. G., Yektaoglu, A., Renner, S., Berg, T., Arenz, C., Giannis, A., Oprea, T. I., Rauh, D., Kaiser, M., Waldmann, H. A scaffold-tree-

Bibliography

- merging strategy for prospective bioactivity annotation of gamma-pyrones. *Angew. Chem. Int. Ed. Engl.* **49**, 3666–70 (2010).
151. Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. & Stewart, J. J. P. AM1 : A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **107**, 3902–3909 (1985).
 152. Böhm, H. J. Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput. Aided. Mol. Des.* **12**, 309–23 (1998).
 153. Seth A. Hayik, Roland Dunbrack, Jr, and K. M. M. J. A Mixed QM/MM Scoring Function to Predict Protein-Ligand Binding Affinity. *J Chem Theory Comput* **6**, 3079–3091 (2011).
 154. Chimenti, F., Secci, D., Bolasco, A., Chimenti, P., Bizzarri, B., Granese, A., Carradori, S., Yáñez, M., Orallo, F., Ortruso, F., Alcaro, S. Synthesis, molecular modeling, and selective inhibitory activity against human monoamine oxidases of 3-carboxamido-7-substituted coumarins. *J. Med. Chem.* **52**, 1935–42 (2009).
 155. <http://www.chembridge.com/internal/catchAll/>.
 156. Binda, C., Lin, M., Hubálek, F., Restelli, N., Edmondson, D. E., Matteci, A. Insights into the mode of inhibition of human mitochondrial monoamine oxidase B from high-resolution crystal structures. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9750–5 (2003).
 157. Binda, C., Wang, J. & Pisani, L. Structures of human monoamine oxidase B complexes with selective noncovalent inhibitors: safinamide and coumarin analogs. *J. Med. ...* **50**, 5848–5852 (2007).
 158. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
 159. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).
 160. Pastor, R. W., Brooks, B. R. & Szabo, A. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* **65**, 1409–1419 (1988).
 161. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089 (1993).
 162. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **65**, 712–725 (2006).

APPENDIX A

Force Field Terms

Bond Stretching term

Represent the energy to stretch a molecular bond and is giving by equation A.1:

$$E_{Stretching} = \sum \frac{1}{2} k (d - d_0)^2 \quad (\text{A.1})$$

Where **k** is the bond stretching force constant between two atoms present in the molecule (expressed as kcal/mol* Å²). **d** is the length of the atoms (expressed in Å), and **d₀** is the reference length of the described atoms.

Angle Bending term

Bond angle vibrations among a triplet atoms is represented, like the previous case, as a harmonic potential (A.2):

$$E_{Bending} = \sum \frac{1}{2} k (\theta - \theta_0)^2 \quad (\text{A.2})$$

Where **k** is the bending constant of the angle. **θ** is the angle value between two contiguous atoms (expressed in degrees). **θ₀** is the reference angle of the described atoms.

Torsion angle term

Expresses the energy to rotate a torsional angle and is described as cosine function (A.3):

$$E_{Torsion} = \sum \frac{1}{2} k_{\varphi} [1 + \cos n(\varphi - \varphi_0)] \quad (\text{A.3})$$

Where k_{φ} is the constant force of a torsional angle. φ is the torsional angle. φ_0 is the reference torsional angle and n is the multiplicity value.

Out-of-plane bending term

In general when three atoms are on the same plane and a fourth atom links the middle atom from the starting triplet, an out-of-plane term must be included to the energy system calculation. Usually out-of-plane term might be treated as an improper torsional angle using equation A.3 or, in case the fourth atoms lies above the plane, might be described by an harmonic potential as in equation A.2 or A.1.

Van der Waals term

Describes the interaction energy between two atoms not directly bonded. It is expressed as Lennard-Jones potential as shown in equation A.4:

$$E_{vdW} = \sum (A_{ij} r_{ij}^{-12} - C_{ij} r_{ij}^{-6}) \quad (\text{A.4})$$

Where r_{ij} is the distance between atom i and the atom j , A_{ij} and C_{ij} are constants which depend on the van der Waals radii of the corresponding atom (A corresponds to the repulsive forces and C corresponds to the attractive forces).

Electrostatic term

Coulomb's law is used to calculate electrostatic interactions between two molecules or within different atoms of the same molecule, according to equation A.5:

$$E_{ele} = \sum \frac{q_i q_j}{D r_{ij}} \quad (\text{A.5})$$

Where q_i is the partial charge of atom i and q_j is the partial charge of atom j . D is the dielectric constant while r is the distance between i and j

Computational Details***Docking studies***

Several X-ray crystal structures of LSD1 were used for the docking studies conducted during the thesis. The crystal structures were chosen taking into account the resolution of the structures, the co-crystallized structure, and the geometries of putative conserved water molecules and were downloaded from Protein Data Bank. Before docking studies, ions and water molecules (not for 2DW4, where the considered conserved water molecule has been kept) were deleted from the crystal structures. Then the structures were protonated using a specific module implemented in MOE suite called "Protonation 3D".¹⁴⁵ After that the right protonation state of histidine side chains was checked by visual inspection partial charges were added to all the residues using Amber99 force field. While MMFF94x was the force field used to add the partial charges to the FAD cofactor. Then the proteins were minimized with conjugate gradient methods using a derivative converge criterion until a value threshold of $0.01 \text{ kcal}(\text{mol} \cdot \text{\AA})^{-1}$ using Amber99 force field for all the residues and MMFF94x force field for the cofactor.

All the ligands structures were generated using MOE suite¹⁴⁵ then the structures were protonated and the partial charges were added using MMFF94x force field. The structures, were then energy minimized using MMFF94x force field and the conjugate gradient method until the derivative

convergence criterion reached the selected value threshold of $0.0001 \text{ kcal}(\text{mol} \cdot \text{\AA})^{-1}$ using MOE software package.¹⁴⁵

A sphere of 20 Å around the oxygen atom of Tyr 761 sidechain was defined as LSD1 binding site for ligand docking. Gold 4.0,⁸⁷ Glide⁹⁰ and ParadockS¹⁰² were the most used docking software during the thesis. Concerning Gold all the torsional angles of the ligands were allowed to freely rotate for every molecule and at least 10 docking runs were performed. GoldScore and ChemScore were the chosen scoring functions. Water molecules were replaced or used as mediators between ligand and protein through a module implemented in Gold called “toggle mode”. For Glide software default values were selected and for every ligand at least 20 docking solutions were generated. GlideScore “standard precision” (SP mode) was the selected scoring function. Concerning ParadockS, defaults setting were also chosen, while P-score and PMF04 were the scoring functions selected. Also here at least 10 docking solutions were generated for every ligand.

Molecular dynamics

MD of LSD1 with 15 residues of the N-terminal histone tail.

All molecular dynamics were run using AMBER 11.0 software package and the amber ff99SB force field.⁸³ From the proteins starting structures (PDB code 2V1D) all ions and water molecules were removed. From the initial PDB structure histone tail Met 4 residue, was mutated into a Lys 4 residue (carrying none, one, two or three methyl flags according to the molecular dynamic conducted (Chapter 3.1)) using MOE.¹⁴⁵ Atom types and partial charges of the new histone tail were calculated using amber99 force field implemented in MOE¹⁴⁵ and then saved as GAFF mol2 file format. The GAFF force field parameters¹⁵⁸ for the histone tail atoms were then generated using *parmchk* module implemented in AMBER11.0.⁸³ Concerning the FAD cofactor, the atom types and partial charges were created using *antechamber* module implemented in AMBER,⁸³ while the missing GAFF force field parameters for the cofactor were generated using *parmchk* module. Now, using *LEAP* module, implemented in AMBER suite,⁸³ the complexes formed by the protein, the cofactor and the histone tail were generated; neutralized with 7 Cl⁻ counterions and solvated in an octahedral box with TIP3P¹⁵⁹ water molecules leaving at least 15 Å between the solute atoms and the borders of the box (25741 water molecules were added to the system), generating a systems formed by 864954 atoms.

Appendix A

Now all the molecular dynamics of the generated systems were conducted using *sander* program already implemented in AMBER suite.⁸³ In order to remove bad close contacts, the starting systems were initially relaxed in three steps. In the first step the counterions and solvent molecules were relaxed for 800 iterations using steepest descent gradient and for 800 iterations using conjugate gradient while the solute was constrained with a positional restraint by a force constant of 10 kcal/mol*Å. In the second step, solvent atoms and protein atoms were relaxed using first steepest descent gradient (800 iterations) and then conjugate gradient (800 iterations), while FAD cofactor, histone tail, and conserved water molecule atoms were constrained to their initial geometries by a force constant of 10 kcal/mol*Å. In the last relaxation step, all constraints were removed and the whole system was relaxed for 1000 iteration using steepest descent gradient and 1000 iterations using conjugate gradient method.

Following the relaxation step, the systems were gradually heated until 300 K through 300 ps of molecular dynamic with positional restrains (10 kcal/mol*Å) on solute atoms, allowing solvent molecules and counterions to move freely. A constant volume periodic boundary was set to equilibrate the temperature of the system by Langevin dynamics¹⁶⁰ by a collision frequency of 1.0 ps⁻¹. After the temperature was equilibrated, the whole system was subject to a pressure equilibration at 1 bar for 300 ps in a constant pressure periodic boundary by an isotropic pressure scaling method employing a pressure relaxation time of 2.0 ps. The temperature was kept 300 K by the Langevin dynamics. Also here the solute was constrained with a positional restrains of 10 kcal/mol*Å. After an equilibration 2 ns long step, where the positional constraints were gradually removed (under conditions of 300 K temperature and 1 bar pressure), the system was subjected to the molecular dynamics simulations for 42 ns. For the all the equilibration steps, 8 Å cutoff was chosen for the short-range non bonded interactions in combination with the Particle Mesh Ewald option.¹⁶¹ While the lengths of hydrogen atoms covalently bounded to heteroatoms were kept constant through SHAKE method¹⁶² implemented in *sander*. The coordination's and the energy parameters were saved every 100 iterations in the relaxations steps, every 400 iterations in the temperature and pressure equilibration step, and every 4000 iterations during the molecular dynamics.

MD of LSD1 with polyamine inhibitors.

All molecular dynamics were run using AMBER 11.0 software package and the amber ff99SB force field.⁸³ The proteins starting structures (PDB code 2DW4) were obtained from the docking studies as PDB file format. The cofactor and all the ligands parameters were generated using general AMBER force field GAFF.¹⁵⁸ Concerning the ligand structures and the FAD cofactor, the atom types and partial charges were added using *antechamber* module implemented in AMBER,⁸³ while the missing GAFF force field parameters were generated using *parmchk* module also implemented in AMBER suite.⁸³ Then for all the ligands and FAD cofactor GAFF atom coordinates and partial charges were replaced from those obtained during the docking studies. Now, using *LEAP* module, implemented in AMBER suite,⁸³ the complexes formed by protein, cofactor and small ligand were generated, neutralized with 2 Cl⁻ counterions and solvated in octahedral box with TIP3P¹⁵⁹ water molecules leaving at least 15 Å between the solute atoms and the borders of the box (26711 water molecules were added to the system), generating a systems formed by 88974 atoms. For the generated complexes, relaxation, equilibration and molecular dynamic steps were conducted using *sander* module present in AMBER 11.0⁸³ analogous to the previously described settings.

MM-PB(GB)SA and QM/MM-GBSA calculations.

The general binding free energy of molecular system is giving by equation A.6:

$$\Delta G = \Delta H - T\Delta S \quad (\text{A.6})$$

Where T is the temperature of the system (300 K), H is the enthalpy and S in the entropy of the system. From the computational point of view, the binding free energy is evaluated using the equation A.7:

$$\Delta G = G_{Complex} - (G_{Protein} + G_{Ligand}) \quad (\text{A.7})$$

Where $G_{Complex}$, $G_{Protein}$ and G_{Ligand} are the absolute free energies related to the complex, the protein and the ligand respectively. Absolute free energies calculated with MM-PB(GB)SA module are calculated according to equation A.8:

$$G \approx E_{MM} + G_{Solv} + TS_{Solute} \quad (\text{A.8})$$

Where E_{MM} is the potential energy of the solute that is determined by equation A.9 (as the sum of van der Waals, Electrostatic and Internal energy) in the gas phase by using *sander* program implemented in AMBER.⁸³ While G_{Solv} is the solvation free energy for transferring the solute from vacuum into solvent and is the sum of electrostatic and non-electrostatic contributions. In AMBER this term is approximated to the sum of the numerical solution of Poisson-Boltzmann equation (PB) or Generalized-Born surface area (GB)¹¹⁷ and the polar free energy contribution giving by the molecular surface area as showed in equation A.10.

$$E_{MM} = E_{Internal} + E_{ele} + E_{VdW} \quad (\text{A.9})$$

$$G_{Solv} = E_{PB(GB)SA} + E_{SASA} \quad (\text{A.10})$$

For the chromane derivatives dataset, the MM-PB(GB)SA binding free energy has been calculated using MM-PB(GB)SA method implemented in AMBER11.⁸³ The complex was generated using *LEAP* and subjected only to a multi-step relaxation procedure in the gas phase as described in the previous session. The dielectric constant of the solute was set as 1.0 while the dielectric constant of the implicit water was set as 80.0 for all the calculations.

Now, absolute free energies of the complex, the protein and the ligand are calculated with QM/MM-GBSA module according to equation A.11:

$$G \approx E_{QM/MM} + G_{Solv} + TS_{Solute} \quad (\text{A.11})$$

Where $E_{QM/MM}$ is the potential energy of the solute that is determined by equation A.12 (as the sum the potential energies related to: QM region and to the MM region, the potential energy at the interface between the two regions, the influence that one region has on the other, and the potential energy derived from the boundaries conditions) in the gas phase.⁸³

$$E_{QM/MM} = E_{QM} + E_{MM} + E_{QM/MM} + E_{Pol} + E_{Boundaries} \quad (\text{A.12})$$

For the chromane derivatives dataset the complexes were generated using *LEAP* and subjected to a multi-step relaxation procedure in the gas phase as described in the previous session. The QM region was set first as the ligand, then as the ligand plus the residues located 4.5 Å around it, and last as the ligand plus all the residues located 6 Å around it (best results obtained (Chapter 6)). The quantum mechanical theory selected was the pseudo-potential AM1, while the total charge of the complex QM area was set as +1.

For all the calculations the entropy value was approximated to the number of rotatable bonds present in the active inhibitor structure (1 kcal/mol for 1 rotatable bond).¹²⁴

Assays details

In vitro peroxidase assays for propargylamine derivatives

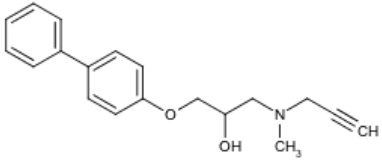
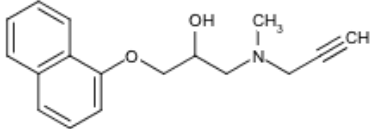
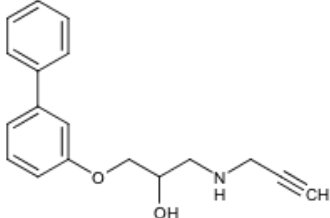
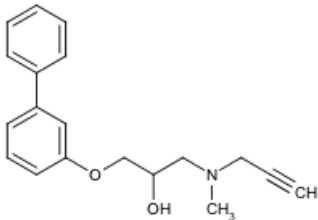
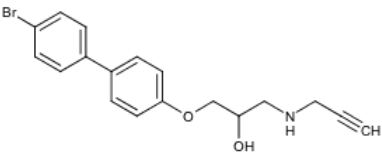
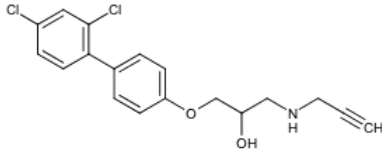
Potency was measured in an established horseradish peroxidase (HRP) assay system based on the Amplex® Red protocol from Invitrogen™ (Life Technologies, Carlsbad, CA). The assay was performed in black 96-well microtiterplates (from Greiner Bio-one GmbH, Germany) with 4 µl LSD1 enzyme, 5 µl inhibitor (dissolved and diluted in DMSO) and 36 µl demethylation buffer (contains 45 mM HEPES, 40 mM NaCl, pH 8.5). This solution was preincubated for 30 minutes, then demethylation was started by adding 5 µl of a solution of 100 µM H3K4(me2) peptide (from Peptide Specialty Laboratories Ltd., Heidelberg, Germany) in buffer. The mixture was incubated for 90 minutes at 30 °C, 50 µl Amplex® Red mixture (containing 100 µM Amplex® Red reagent and 2 U/ml HRP in demethylation buffer) was added and, after 5 minutes, detection performed on a BMG polarstar microplate reader (λ_{ex} : 550 nm, λ_{em} : 615 nm).¹⁴⁴ Details obtained from Martin Leo Schmitt (University of Freiburg).

In vitro DELFIA® assays

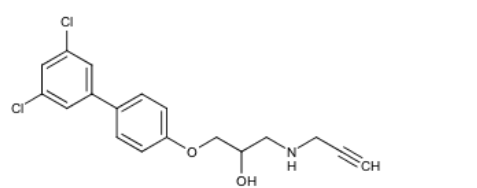
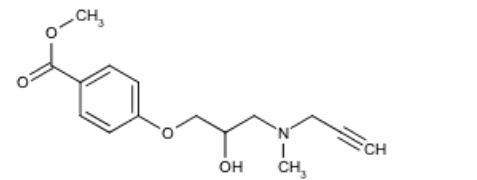
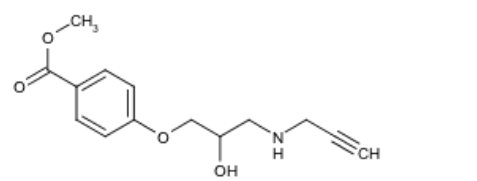
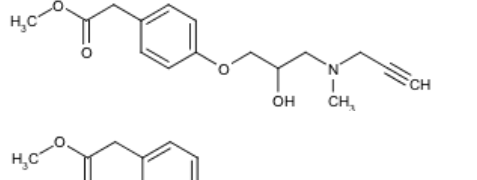
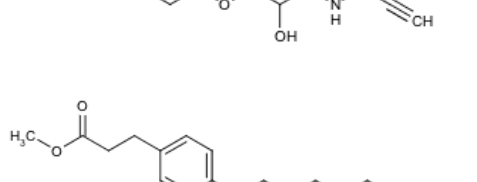
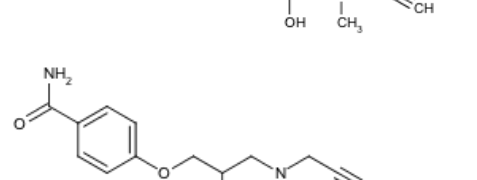
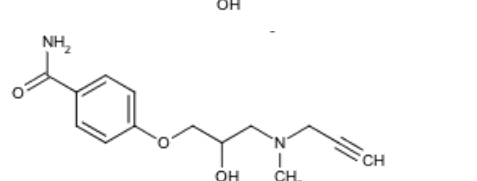
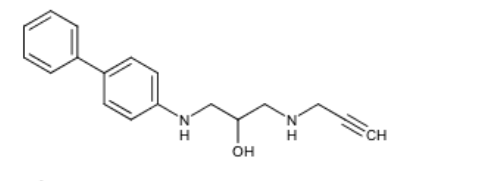
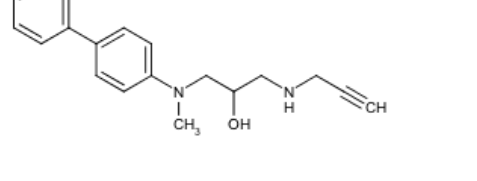

6 µl LSD1, 4 µl demethylase buffer and 2 µl inhibitor- or blank-solution (contains 20% DMSO and 80% buffer v/v, 2% DMSO final) were given to a Eppendorf cup and shaken for 5 minutes at RT. Peptide solution (2.2 µl peptide plus 5.8 µl demethylation buffer) was added to start the reaction. After 15 minutes incubation at 30 °C, the reaction was stopped by adding 380 µl DELFIA wash buffer pH10 and shaken to give a homogeneous solution. 5 µl of this solution and 95 µl DELFIA wash buffer pH10 was given into a 96-well streptavidin coated yellow microplate. After incubation for 1 hour with shaking, the wells were washed 6 times with DELFIA wash buffer. 100 µl of a dilution of anti-H3K4(unmodified) antibody in TRIS incubation buffer (1:10.000) was added and again incubated for 1 hour at 30 °C. After the same washing procedure as before, incubation with 100 µl europium labelled anti-mouse antibody in TRIS incubation buffer (1:250) at 30 °C for 1 hour was started. After a last washing step, 100 µl enhancement solution was added and after 10 minutes shaking at room temperature measured in a Perkin Elmer EnVision with the recommended filters and mirrors from Perkin Elmer. Details obtained from Martin Leo Schmitt (University of Freiburg).

Appendix A

APPENDIX B**Table 1: Propargylamine derivatives dataset**

Structure	Compound Name	Inhibition @ conc Peroxidase Assay
	5b	22.7% @ 100 μ M
	5c	23.4% @ 100 μ M
	5g	36.3% @ 100 μ M
	5h	41.8% @ 100 μ M
	5i	Inactive
	5j	Inactive

Appendix B

	5k	Inactive
	5l	Inactive
	5m	Inactive
	5n	Inactive
	5o	Inactive
	5p	Inactive
	5q	Inactive
	5r	Inactive
	6a	Inactive
	6b	Inactive

Appendix B

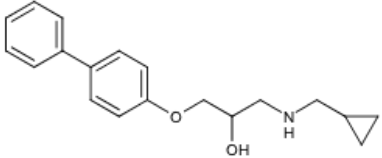
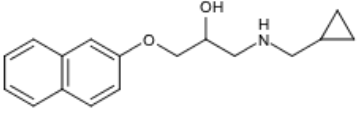
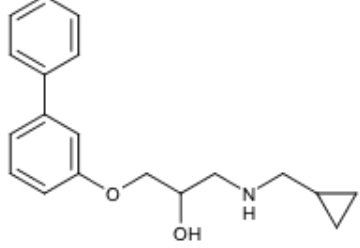
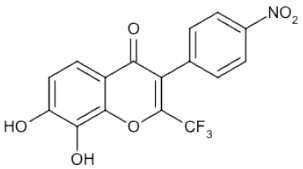
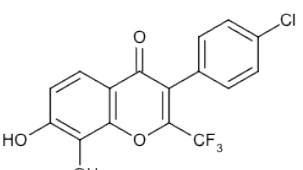
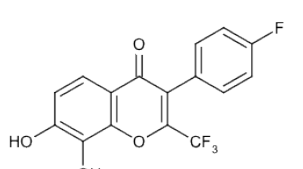
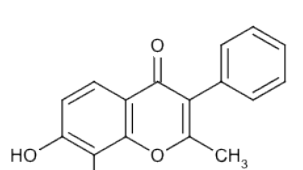
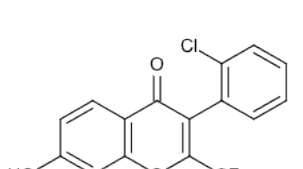
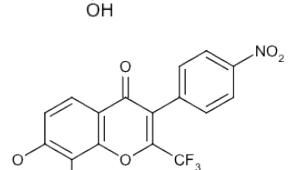
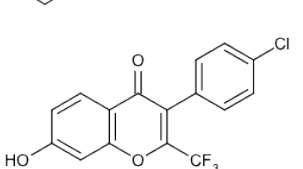
	7a	Inactive
	7b	Inactive
	7d	39.0% @ 100 μ M

Table 2: 4H-chromen-4-one derivatives dataset using Namoline as query structure

ZINC ID	Vendor Name/ Name Code 1	Name Code 2	IC ₅₀ ± SE or Inhibition @ conc Peroxidase Assay	IC ₅₀ ± SE or Inhibition @ conc DELFI A Assay
ZINC00112306	STOCK2S-44578	Namoline, 13	56.03 μM ± 1.62 μM	3.08 μM ± 0.35 μM
ZINC02311497	STOCK1S-16519	Shaolin, 35a	2.47 μM ± 0.38 μM	Inactive @ 20 μM
ZINC00177625	PHAR-000079	32e	24.55 μM ± 2.98 μM	Inactive @ 20 μM
ZINC00112316	STOCK2S-37771	32h	Inactive @ 50 μM	//
ZINC00112328	STOCK2S-49639	36	24.3% @ 100 μM	//
ZINC06668618	MolPort-002746047	33e	26.66 μM ± 3.90 μM	1.98 μM ± 0.18 μM
ZINC00177623	STOCK1S-26451	32f	40.8% @ 50 μM	Inactive @ 20 μM
ZINC00306014	STOCK2S-54080	32g	36.2% @ 100 μM	//
//	MS137	33k	1.43 μM ± 0.41 μM	44% @ 0.2 μM
//	MS142	33l	3.93 μM ± 0.57 μM	0.413 μM
//	MS127	33g	7.06 μM ± 1.07 μM	0.859 μM
//	MS122	32c	Inactive	//
//	MS124	32j	Inactive	//
//	MS112	32i	Inactive	//
//	MS107	32a	Inactive	//
//	MS153	33j	43% @ 100 μM	//
//	MS146	34b	Inactive	Inactive @ 5 μM
//	MS149	34a	Inactive	Inactive @ 1 μM
//	MS148	33c	47% @ 100 μM	Inactive @ 5 μM
//	AW61	33i	Inactive @ 50 μM	//
//	AW60	33h	25% @ 50 μM	//

Table 3: 4*H*-chromen-4-one false positive derivatives dataset using Shaolin as query structure

Structure	Compound Name	IC ₅₀ ± SE or Inhibition @ conc Peroxidase Assay	IC ₅₀ ± SE or Inhibition @ conc DELFIA Assay
	Shaolin, 35a	2.47 μM ± 0.38 μM	Inactive
	35b	2.47 μM ± 0.38 μM	Inactive
	35c	Inactive	//
	35d	102 μM	Inactive
	35e	Inactive	//
	35f	Inactive	//
	35g	Inactive	//

Appendix B

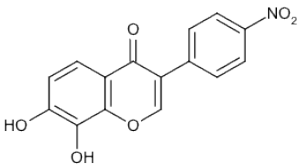
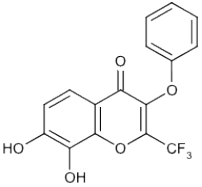
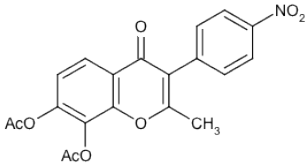
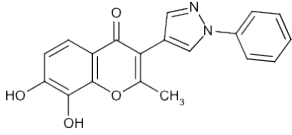
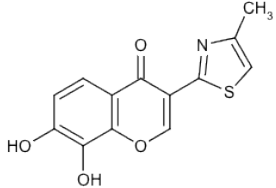
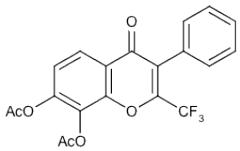
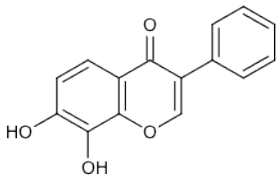
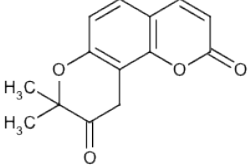
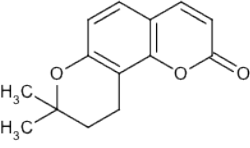
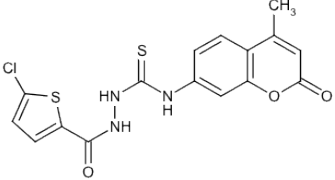
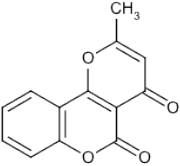
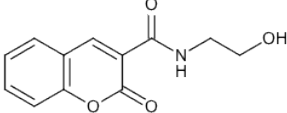
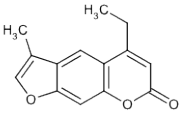
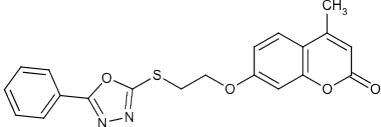
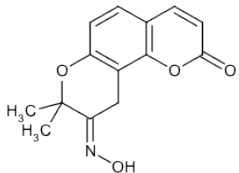
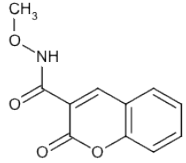
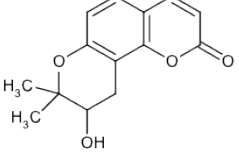
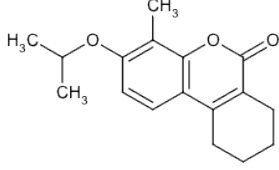
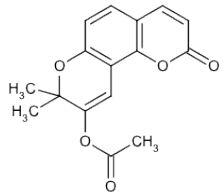
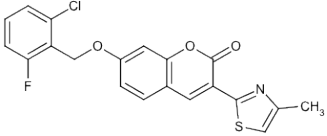
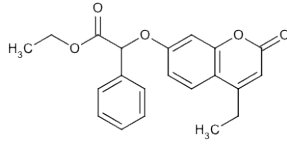
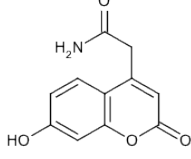
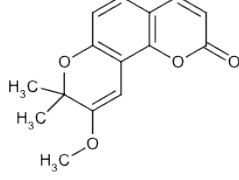
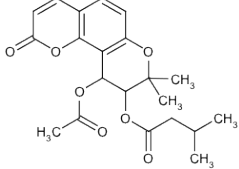
	35h	27.93 μ M \pm 5.94 μ M	Inactive
	35i	//	//
	35k	//	//
	35j	16.36 μ M \pm 2.10 μ M	Inactive
	35l	28.10 μ M \pm 5.50 μ M	Inactive
	35m	21.3% @ 50 μ M	Inactive
	35n	24.54 μ M \pm 4.48 μ M	Inactive

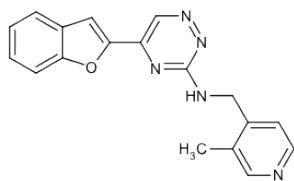
Table 4: Coumarine derivatives from preliminary similarity search using Chembridge database

Structure	Chembridge ID	Compound Name	IC ₅₀ ± SE or Inhibition @ conc Peroxidase Assay
	CB5325408	5325408	5 μM
	CB6670945	6670945	100% @ 50 μM
	CB7785969	7785969	3 μM
	CB7648452	//	Inactive
	CB5102975	//	Inactive
	CB6432612	//	Inactive
	CB7375345	//	Inactive
	CB5325409	//	Inactive

Appendix B

	CB9057353	//	Inactive
	CB5325406	//	Inactive
	CB6238255	//	Inactive
	CB6657683	//	Inactive
	CB6240796	//	Inactive
	CB7353505	//	Inactive
	CB7932577	//	Inactive
	CB6662389	//	Inactive
	CB5325401	//	Inactive

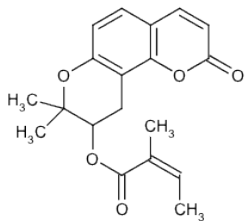
Appendix B



CB41577639

//

Inactive



CB5325407

//

Inactive

Appendix B

List of abbreviations

ADME	Adsorption, Distribution, Metabolism, Excretion
AM1	Austin Model 1
AML	Acute Myeloid Leukemia
AOD	Amio Oxidase Domain
AOL	Amino Oxidase Like
AR	Androgen Receptor
ATRA	All Trans Retinoic Acid
BHC80	BRAF35 (MIM 605535) Histone deacetylase Complex 80
CAMD	Computer Assisted Molecular Design
CoMFA	Comparative Molecular Fields Analysis
CoMSIA	Comparative Molecular Similarity Index Analysis
CoREST	Repressor element-1 Silencing Transcription factor Co-Repressor
DNA	Deoxyribonucleic Acid
DELFI A	Dissociation-Enhanced Lanthanide Fluorescent Immunoassay
DFT	Density Functional Theory
DNMTs	DNA Methyltransferase
Er α	Estrogen Receptor alpha
FAD	Flavin Adenine Dinucleotide
F _c -region	Fragment Crystallizable region
FDH assay	Formaldehyde-dehydrogenase-dependent assay
FEP	Free Energy Perturbation
FMS1	Fenpropimorph-resistance Multicopy Suppressor 1 (Polyamine oxidase)
GA	Genetic Algorithm
H	Histone protein
HDAC	Histone Deacetylase
HDM	Histone Demethylase
HRP	Horseradish Peroxidase (Assay)
HTC116	Colorectal Carcinoma Cells
HTM	Histone Methyltransferase
HTS	High-Throughput screening

List of Abbreviations

JHDM1	JmjC Domain-Containing hstone Demethylase 1
JmjC	Jumonji C (Domain-Containing Histone Demethylases)
LAAO	L-Amino Acid Oxidase
LIE	Linear Integration Energy
LNCap	Androgen-Sensitive Human prostate adenocarcinoma cells
LOO	Leave-One-Out
LSD1	Lysine Specific Demethylase-1 (KDM1A, BHC110, AOF2)
LSD2	Lysine Specific Demethylase-2 (KDM1B, AOF1)
MALDI	Matrix-Assisted Laser Desorption/ionization
MAO	Monoamine Oxidase Proteins
MAO N	Fungal Monoamine Oxidase
MC	Monte Carlo
MD	Molecular Dynamic
MLL	Mixed-Lineage Leukemia
MM	Molecular Mechanics
MM-GBSA	Molecular Mechanics Generalized-Born Surface Area
MM-PBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
MMFF94x	modified Merck Molecular Force Field 94
MRA	Multi Regression Analysis
MS	Mass Spectrometry
mRNAs	micro Ribonucleic Acid filaments
NOD	Nonobese Diabetic
ns	Nanoseconds
p53	Suppressor and transcriptional activator protein 53
PAO	Polyamino Oxidase Proteins
PCPA	Trans-1-Phenylcyclopropylamine
PHD	Plant Homeodomain (finger)
PKC β I	Protein Kinase-C Beta I
PLS	Partial Least Square analysis
PRESS	Predicted Residual Sum of Squares
PRK1	Protein Kinase-C-Related Kinase 1
PRMT1	Arginine N-methyltransferase one protein
PRMT2	Arginine N-methyltransferase two protein

List of Abbreviations

PSO	Particle Swarm Optimizer
q ²	Prediction Coefficient
QM	Quantum Mechanics
QM/MM-GBSA	Quantum Mechanics Molecular Mechanics Generalized-Born Surface Area
QM/MM-PBSA	Quantum Mechanics Molecular Mechanics Poisson-Boltzmann Surface Area
QSAR	Quantitative Structure Activity Relationships
r ²	Description Coefficient
REST	RE-1 Silencing Trascription Factor
RMSE	Root Mean Square Error
ROS	Reactive Oxygen Species
SANT2	Switching-defective protein 3 (Swi3), Adaptor 2 (Ada2), Nuclear receptor co-repressor (N-CoR), Transcription factor (TF) IIIB two
SCID	Severe Combined Immunodeficient
SEP (SDEP)	Standard Error of Prediction (Standard Deviation of the Errors of Predictions
SWIRM	SWI3, RSC80, and MOIRA proteins
TI	Thermodynamic Integration
TR-FRET	Time Resolved-Fluorescence Resonance Energy Trasfer assay
TS	Tabu Search
VS	Virtual screening
ZINC	Zinc Is Not Commercial
ΔG	Gibbs Binding Free energy
ΔH	Enthalpy
ΔS	Entropy

List of Abbreviations

Amino acids letter codes

Amino-acid	Three letters code	Single letter code
Glycine	Gly	G
Alanine	Ala	A
Valine	Val	V
Leucine	Leu	L
Isoleucine	Ile	I
Methionine	Met	M
Phenylalanine	Phe	F
Tyrosine	Tyr	Y
Tryptophan	Trp	W
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Cysteine	Cys	C
Lysine	Lys	L
Arginine	Arg	R
Histidine	His	H
Aspartate	Asp	D
Glutamate	Glu	E
Asparagine	Asn	N
Glutamine	Gln	Q

Amino acids letter codes

List of publications

- I. Ayşe Selcen Alpan, Sülünay Parlar, Luca Carlino, Ayse Hande Tarikogullari, Vildan Alptüzün, Hasan Semih Günes. *Synthesis, biological activity and molecular modeling studies on 1H-benzimidazole derivatives as acetylcholinesterase inhibitors.*
Bioorganic & Medicinal Chemistry. 21(17):4928-4937; 2013.
- II. Schmitt, M.; Hauser, A. T.; Carlino, L.; Pippel, M.; Metzger, E.; Willmann, D.; Yiu, T.; Barton, M.; Schule, R.; Sippl, W.; Jung, M. *Non-peptidic propargylamines as inhibitors of lysine specific demethylase 1 (LSD1) with cellular activity.*
Journal of Medicinal Chemistry. 53(18):7334-7342 2013.
- III. Hoffmann, I., Roatsch, M., Schmitt, M. L., Carlino, L., Pippel, M., Sippl, W., Jung, M. *The role of histone demethylases in cancer therapy.*
Molecular Oncology. 6(6):687-703; 2012.
- IV. Furdas, S. D., Carlino, L., Sippl, W., & Jung, M. *Inhibition of bromodomain mediated protein-protein interactions as a novel therapeutic strategy.*
MedChemComm. 3(2);123-134; 2012.
- V. R. Heinke, L. Carlino, R. Kannan, M. Jung, W. Sippl *Computer- and structure-based approaches to identify novel lead structures for epigenetic targets.*
Bioorg. Med. Chem. 15;19(12):3605-15; 2011

Curriculum Vitae

Personal Information

Name	Luca Carlino
Date of birth	July 02 nd 1984
Place of birth	Caserta, Italy

Education

2009-2013	PhD Student under the supervision of Prof. Dr. Wolfgang Sippl Institute of Pharmaceutical Chemistry, Martin-Luther-University Halle-Wittenberg (Germany)
07/2008-03/2009	Master of Science thesis under the supervision of Prof. Gabriele Costantino Faculty of Pharmacy, department of Pharmaceutical Chemistry, University of Parma (Italy)
2004-2009	Study of Chemistry and Pharmaceutical Technology, Faculty of Pharmacy, University of Parma (Italy)

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertationsschrift selbständig und ohne fremde Hilfe angefertigt, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die aus ihnen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Die Arbeit wurde ausschließlich der Mathematisch-Naturwissenschaftlichen Fakultät der Martin-Luther-Universität Halle-Wittenberg vorgelegt und an keiner anderen Universität oder Hochschule weder im In- und Ausland zur Erlangung des Doktorgrades eingereicht.

Halle/Saale, den 19.11.2013