universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

Identification of HIV Protease Inhibitors by *in Silico* Screening

verfasst von

Birgit Zonsics

angestrebter akademischer Grad

Magistra der Pharmazie (Mag.pharm.)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 449
Studienrichtung lt. Studienblatt:	Pharmazie
Betreut von:	Univ.-Prof. Mag. Dr. Gerhard Ecker

Acknowledgements

This paper would not have been possible without the help of many people that I would like to thank.

First, I want to thank my supervisor Professor Gerhard Ecker, who provided the contact to Professor Gisbert Schneider and gave me the chance to do the practical research for my thesis at ETH Zurich.

I want to thank Professor Gisbert Schneider from ETH for his warm reception in his workgroup and for providing an interesting topic I was allowed to work on. He was a great supervisor who smoothly guided me and encouraged me and gave me the opportunity to learn a lot.

I am grateful to the whole group around Professor Schneider, especially to Tiago Rodriguez, who advised me in the syntheses and had always time and patience for all my questions. He also gave helpful advice for the composition of my thesis. I am also grateful for my colleague Jens Kunze, who introduced me to my topic and was the first person to help me with my problems. I am obliged to Petra Schneider, who had a helping hand during the activity testing of the molecules and supported me a lot in the data evaluation after my return to Vienna. Many thanks to Sarah Haller, who answered all the questions in the lab and explained the handling of the equipment to me.

I would also like to thank the members of the workgroup of Professor Ecker in Vienna, who warmly integrated me into their group after my return and supported me especially in the composition of my thesis and in administrative matters.

I am deeply obliged to Tiago Rodriguez, Daniela Digles, Gwendolin Korinek and my father Erich Zonsics for proofreading my diploma thesis and for the precious remarks and suggestions to improve my text.

My precious thanks to my flatmates Lilian Schaad and Sara Wyss, who warmly received me in their apartments, shared their time with me and became friends during this wonderful time in Zürich.

In addition, I want to thank my parents and grandparents, who encouraged me from the first thought of going abroad for my diploma thesis and without whose financial support this project would never have been possible. They accompanied me through all the ups and downs while I was writing my paper. I am especially grateful for their patience and attendance.

Last but not least I would like to thank all my friends who shared my worries and endured my antics during that difficult process. Each of them contributed in their very special way to help me complete my thesis.

Danksagung

An dieser Stelle möchte ich all jenen danken, die mit ihrer fachlichen und persönlichen Unterstützung zum Gelingen und zum Abschluss dieser Diplomarbeit beigetragen haben.

Ich danke Herrn Professor Gerhard Ecker, dem Betreuer meiner Diplomarbeit, der den Kontakt zu Herrn Professor Gisbert Schneider herstellte und es mir ermöglichte, die Forschungsarbeiten zu meiner Diplomarbeit an der ETH Zürich durchzuführen.

Ich danke Herrn Professor Gisbert Schneider von der ETH für die herzliche Aufnahme in seiner Arbeitsgruppe und für die Bereitstellung dieses interessanten Themas. Ich wurde während der gesamten Forschungsarbeit in Zürich von ihm ausgezeichnet betreut und angeleitet und habe dabei sehr viel gelernt.

Mein Dank gilt der gesamten Arbeitsgruppe um Professor Schneider, besonders Tiago Rodriguez, der mich bei der Synthese meiner Moleküle anleitete und der darüber hinaus immer für Fragen aller Art ein offenes Ohr hatte und mir beim schriftlichen Verfassen der Arbeit viele gute Anregungen gab. Ich danke meinem Kollegen Jens Kunze, der mich in mein Thema einführte und für viele Fragen mein erster Ansprechpartner war. Herzlichen Dank auch an Petra Schneider, die mir bei der Aktivitätstestung der synthetisierten Substanzen helfend zur Seite stand und mir auch nach meiner Rückkehr nach Wien bei der Datenauswertung behilflich war. Vielen Dank an Sarah Haller, die als Laborantin fast alle praktischen Fragen im Labor beantworten konnte.

Ich möchte mich an dieser Stelle auch bei allen anderen Kollegen und Kolleginnen aus Zürich für die angenehme und produktive Zusammenarbeit bedanken!

Ich danke den Mitgliedern der Arbeitsgruppe von Professor Ecker in Wien, die mich nach der Rückkehr aus Zürich bei diversen Fragen zur Gliederung der Arbeit und zum

organisatorischen Ablauf unterstützten und mich auch sehr nett in ihre Gruppe aufgenommen haben.

Ich danke Tiago Rodriguez, Daniela Digles, Gwendolin Korinek und meinem Vater Erich Zonsics für das Korrekturlesen meiner Arbeit und für wertvolle Anregungen, die damit in Verbindung standen.

Ich danke meinen Mitbewohnerinnen Lilian Schaad und Sara Wyss, die mich in Zürich sehr herzlich aufgenommen und beherbergt haben und die in dieser Zeit meine Freundinnen geworden sind.

Ganz besonderer Dank gilt meinen Eltern und Großeltern, die mich von Anfang an in meinem Vorhaben, ins Ausland zu gehen, bestärkt haben und ohne deren finanzielle Hilfe dieser Auslandsaufenthalt nicht möglich gewesen wäre. Sie haben mich auch während der Niederschrift der Arbeit durch alle Hochs und Tiefs begleitet und ich danke ihnen sehr für ihre Geduld und ihren Zuspruch.

Zuletzt möchte ich mich bei allen meinen Freundinnen und Freunden bedanken, die mich in dieser Zeit begleitet und ertragen haben: mit ihren Ermunterungen zur rechten Zeit haben auch sie zu einem positiven Abschluss meiner Diplomarbeit beigetragen.

Die Universität Wien sponsorte diese Diplomarbeit mit einem Stipendium für kurzfristiges wissenschafltiches Arbeiten im Ausland (KWA)

This diploma thesis was sponsored by the University of Vienna

Zusammenfassung

Das humane Immundefizienz Virus (HIV) ist ein Retrovirus das hauptsächlich T-Helferzellen des Immunsystems befällt und nach einer Latenzzeit von ca. 7 Jahren zum Ausbruch von AIDS (erworbenes Immundefizienz Syndrom) führt. Mit der Hemmung des Enzyms HIV Protease wird der letzte Schritt des viralen Entwicklungszyklus, die Reifung von neu gebildeten viralen Partikeln, blockiert und diese verlieren die Fähigkeit andere Zellen zu infizieren. HIV Protease Hemmer sind ein wichtiger Bestandteil der HAART (hoch aktive antiretrovirale Therapie), die die Viruslast im Patienten unter der Nachweisgrenze halten kann und den Infizierten ein nahezu normales Leben ermöglicht.

Ziel der vorliegenden Diplomarbeit war es mit einem *de novo* Design Programm neue potentielle HIV Protease Hemmer zu generieren, diese zu synthetisieren und dann auf ihre Aktivität gegen die Protease zu testen.

Das *de novo* Design Programm DOGS (Design of Genuine Structures) wurde in der Arbeitsgruppe von Professor Gisbert Schneider entwickelt und zählt zu den ligandenbasierten *de novo* Design Programmen. Als Referenzligand, nach dessen Vorbild neue Moleküle generiert werden sollten, wurde Amprenavir ausgewählt.

Die 856 von DOGS generierten Moleküle wurden anhand ihres Molekülgewichtes und ihres SlogP (berechneter Logarithmus des Octanol-/Wasserkoeffizienten) aussortiert und auf eine kleinere Auswahl reduziert. Die 204 kleinsten Moleküle wurden mit Hilfe des Programmes GOLD (Genetic Optimization for Ligand Docking) in die Bindetasche der HIV Protease gedockt. Danach wurden anhand des GOLD-Scores und der Interaktion mit der Bindetasche 6 Moleküle ausgewählt, die synthetisiert werden sollten.

Für 3 Moleküle gelang die vorgeschlagene Synthese und diese wurden nach ihrer Aufreinigung an der HIV Protease getestet.

Dazu wurde ein FRET (Förster Resonanz Energie Transfer) Assay verwendet. Die 3 verschiedenen Testläufe mit unterschiedlichen Bedingungen ergaben widersprüchliche Werte. Es konnte keine klare Aussage über die Aktivität der Moleküle getroffen werden. Eine externe Testung bestätigte, dass die 3 Moleküle nicht gegen HIV Protease aktiv sind.

Weiterführende Untersuchungen wie die Synthese, Derivatisierung und Testung anderer Moleküle aus dem Pool von neu generierten Stukturen könnten aber durchaus zu aktiven Verbindungen gegen die HIV Protease führen.

Abstract

HIV is a retrovirus that primary attacks the CD4+ T-lymphocytes of the immune system. After a latency time of about 7 years it usually causes AIDS (Acquired immunodeficiency syndrome). By inhibiting the enzyme HIV protease the maturation of new viral particles as the last step of the viral life cycle is blocked and the immature viral particles can no longer infect further cells in the organism. Therefore HIV protease inhibitors play an important role in the highly active antiretroviral therapy (HAART). They help to keep the viral load under the detection level and allow a nearly normal life for the infected patients.

The aim of the present diploma thesis was to generate new HIV protease inhibitors using the novel *de novo* design software DOGS (Design of genuine structures) which was developed in the workgroup of professor Gisbert Schneider at ETH Zürich. Further we wanted to synthesize the molecules and test their activity against HIV protease.

The as the reference ligand for the software Amprenavir was chosen. Subsequently 856 unique molecules were generated by DOGS. They were sorted by their molecular weight and the SlogP (logarithm of the calculated octanol/water coefficient). The subset of the 204 smallest molecules were docked into the protein binding pocket with the software GOLD (genetic optimization for ligand docking). Then the molecules were sorted by their GOLD-Scores. The ligand interactions with the protease were also considered to chose 6 molecules to be synthesized.

Three of the these molecules could be synthesized and later tested for their activity against the HIV protease. Therefore a FRET (Förster resonance energy transfer) assay was used. In three different test runs with varied conditions contradicting results were obtained. It was decided to send the substances to an external company to test them and they were all tested inactive against the HIV protease.

It is still possible to find active molecules against HIV protease in the pool of newly generated structures, but therefore more molecules have to be synthesized, modified and tested.

Table of contents

1 Introduction

1.1 General overview

The Human Immunodeficiency Virus (HIV) is the causal agent of Acquired Immunodeficiency Syndrome (AIDS). It is mainly transmitted by sexual contact and by contaminated blood or blood products. HIV/AIDS is especially problematic in poor countries, where the lack of blood screenings, expensive medical treatment and health and sexual education programmes condition the control of the disease. In the developed countries the carefreeness of many young people is raising the numbers of new infections in the last few years. (Aidshilfe Wien 2012)

Since the discovery of the virus many efforts have been made to find a cure but with limited success. (www.hivbook.com 2011) HIV attacks the immune system and it has a high mutation rate. Furthermore, there are several different forms of the virus located in different regions of the world. Due to mutations and multiple infections with different types, subtypes and mutants of HIV the virus can reach a certain state of resistance against the treatment. (Quiñones-Mateu et al. 2008) These problems clearly show the never-ending need for research in the field and the necessity of the development of new drugs and treatment strategies.

1.2 HIV and AIDS

1.2.1 History

At the end of the 1970ies, clinicians in the USA identified a new disease among young gay men and intravenous drug abusers characterized by immune disorder, *Pneumocystis carinii* pneumonia and rare aggressive cancers such as the Kaposi's sarcoma along with opportunistic infections. (CDC 1981; Gottlieb et al. 1981). In 1983, a retrovirus later named HIV (Coffin et al. 1986) was discovered, causing lymphadenopathy (Barré-Sinoussi et al. 1983) and verified in 1984 as the cause of AIDS. The name AIDS was created by the CDC (Centre of Disease Control) in 1982 because of the weak immune defence in the concerned person and the increased rate of infections rather harmless for immunocompetent people usually only occurring in old people and people with suppressed immune system, as well as

the presumption that this disease is acquired by an infection and not inherited. (MMWR Weekly 1983)

It was found that monkeys had transmitted HIV to humans and that it originated from central Africa. The chimpanzee (*Pan troglodytes troglodytes*) was the carrier for HIV-1 (F. Gao et al. 1999) and the sooty mangabay (*Cercocebus atys*) for HIV-2 (Z. Chen et al. 1997). Both viruses lead to AIDS, but HIV-2's pathogenic course is in general longer. Homology between the two forms is about 40%. (Franchetti et al. 1998) Analysing early probes of AIDS-infected people, especially from Africa, the first infection of humans can be dated back to the period of 1884-1920. (Worobey et al. 2008)

The period of intense research following 1983 gave answers to many important questions. The mode of transmission was elucidated (1984-1985) (MMWR Weekly 1983; Clumeck et al. 1984; Piot et al. 1984; Ziegler et al. 1985; Quinn et al. 1986). Shortly after the sequencing of the genome of HIV, the genes and proteins were defined (1984-1985). (Sanchez-Pescador et al. 1985; Ratner et al. 1985; Wain-Hobson et al. 1985) The main target cells were found to be CD4+ T-cells, macrophages and microglial cells in the brain. (Harper et al. 1986; Shaw et al. 1984) With the availability of the blood test in 1984 (Joyce & I. Anderson 1985) the first step to control the disease was made. Closer monitoring of the viral load in patients was possible, and donated blood was screened against the virus to reduce infections during surgeries and among haemophiliacs and other people in need of blood donations. In 1985, Azidothymidine (AZT) originally synthesized for antibacterial purposes was found to be active against HIV. (H Mitsuya et al. 1985) It was approved by the FDA (Food and Drug Administration) as first antiretroviral drug for the treatment of AIDS in 1987. (J. Brown 1987)

1.2.2 Origin, Spread and Epidemiology

At the time of the discovery of HIV and AIDS in the 1980s, it was not appraisable that the virus would spread this rapidly all around the world causing one of the most serious health crises in history. In developed countries, HIV/AIDS is now seen as a chronic disease and the life expectancy of infected people is relatively normal in comparison to healthy people. But especially in Sub-Saharan Africa, HIV/AIDS and the following diseases and cancers have become one of the leading causes of death and shorten the mean lifespan by about 20 years. (www.hivbook.com 2011)

The first documented infection dates back to 1959 and was found in a blood sample from Zaire. (Worobey et al. 2008) After the development of HIV deriving from SIV (simian immunodeficiency virus) in the period between 1900 and 1920, the virus first spread in smaller regions in central Africa. (Sharp & Hahn 2008) From Africa, HIV spread out to the Caribbean (Haiti) and later to the US. From the US, the virus was introduced to Europe and to other regions in the world.

The different subtypes of HIV are depicted in the following graphic (Figure 1). Each subtype can mainly be derived from one concrete zoonosis event with one type of SIV (simian immunodeficiency virus). Multiple infections with more than one subtype of HIV are possible and are likely to cause recombinant forms combining the two subtypes. This happens usually in regions where HIV has a high incidence in the population. (Burke 1997)

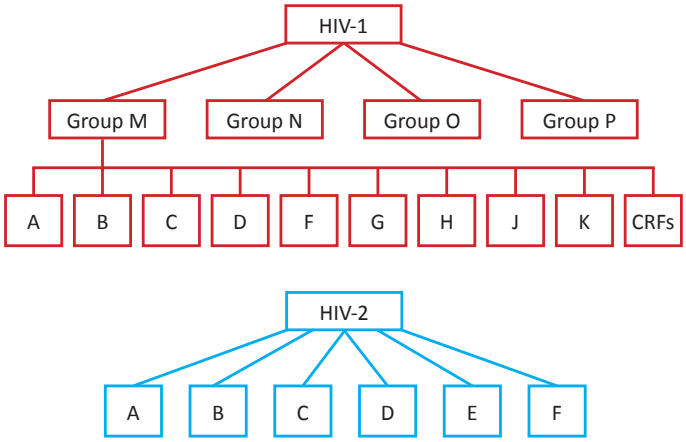


Figure 1: Subtypes of HIV-1 and HIV-2
CRFs (circulating recombinant forms) are combinations of different subtypes

According to the last UNAIDS report (2010) AIDS caused around 25 million deaths worldwide up to now. At the moment, about 34 million people are infected with HIV (see Table 1). (UNAIDS report 2010) AIDS is ranked on the fifth place among the causes of death worldwide. (Infectious Diarrhoea, Pneumonia, Tuberculosis, AIDS, Malaria). (Herold 2011)

Table 1: Epidemiology worldwide data from 2011 (UNAIDS, 2011)

Region	Infected Adults and Children	Prevalence %	New infections	Yearly deaths due to AIDS
Sub Sahara-Africa	22.500.000	5.0	1.800.000	1.300.000
Near and Middle east	460.000	0.2	75.000	24.000
South and southeast Asia	4.100.000	0.3	270.000	260.000
Eastern Asia	770.000	0.1	82.000	36.000
Oceania	57.000	0.3	4.500	1.400
Latin America	1.400.000	0.5	92.000	58.000
Caribbean	240.000	1.0	17.000	12.000
North America	1.500.000	0.5	70.000	26.000
Eastern Europe and Central Asia	1.400.000	0.8	130.000	76.000
Western and Central Europe	820.000	0.2	31.000	8.500
Total	**33.300.000**	**0.8**	**2.600.000**	**1.800.000**

Epidemiology in Austria:

The newest data from Aidshilfe Wien reports about 1-2 new infections daily. In 2011, 525 people were infected with HIV. 7.8% more per year than in 2010 with only 487 new infections. In total, 12.000-15.000 have been infected with HIV from 1983 on. Half of them are living in Vienna, the Capital city of Austria. The distribution between men and women is about two to one third. 3.659 people were affected with AIDS in the period between 1983-2011. 1.945 of them died. 1.714 patients are now living with AIDS in Austria. (Aidshilfe Wien 2012)

1.2.3 Human Immunodeficiency Virus (HIV)

HIV is an RNA virus that belongs to the group of retroviruses. Retroviruses are characterized by a step called reverse transcription where their genetic information stored in an RNA

molecule is transformed to DNA by the enzyme reverse transcriptase for later integration into the host cell DNA. (Campbell 2009)

Retroviruses are divided into several groups. The lentiviruses, the group that also contains HIV, have a very long latency time in the human body before the disease breaks through. HIV can be assigned to that group because it can take about 7 years from the infection until the first symptoms of AIDS appear. (Weiss 1996)

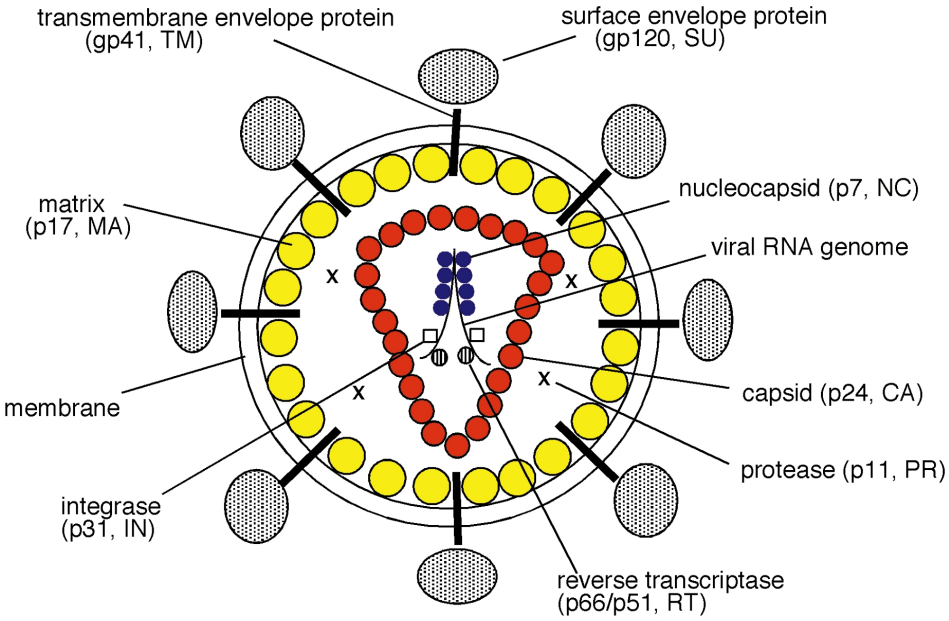


Figure 2: schematic depiction of the HIV virion in the mature state (E.O. Freed, 1998)

The virion has a diameter of about 100 nm. It is a nearly spherical particle. Several layers surround the genetic information of the virus. The outer envelope is formed by a lipid bilayer deriving from the host cell. (Greene 1993) In this membrane, glycoproteins gp120 (surface envelope protein) and gp41 (trans-membrane envelope protein) are embedded. These proteins play a key role in the identification of host cells and promote the binding to specific receptors on the surface of T-helper cells such as CD4 and its co-receptors (CCR5 and CXCR4). (E O Freed 2001b a) On the inner side of this lipid bilayer, the matrix protein p17 creates a shell around a conical capsid core built by the capsid protein p24. (Turner & Summers 1999) This core contains two single strands of viral RNA stabilized by the protein p7 also called nucleocapsid as well as the three essential viral enzymes reverse transcriptase, integrase and protease. Several proteins essential for the communication and regulation of

the host cell are directly packed in the viral particle while others are transcribed from the viral genome in the host cell. (Greene 1993)

1.2.4 Replication

The replicative cycle of HIV can be classified into two phases, the early phase (entry, reverse transcription and integration) and the late phase (transcription and translation, budding and maturation). (E O Freed & M. A. Martin 2001)

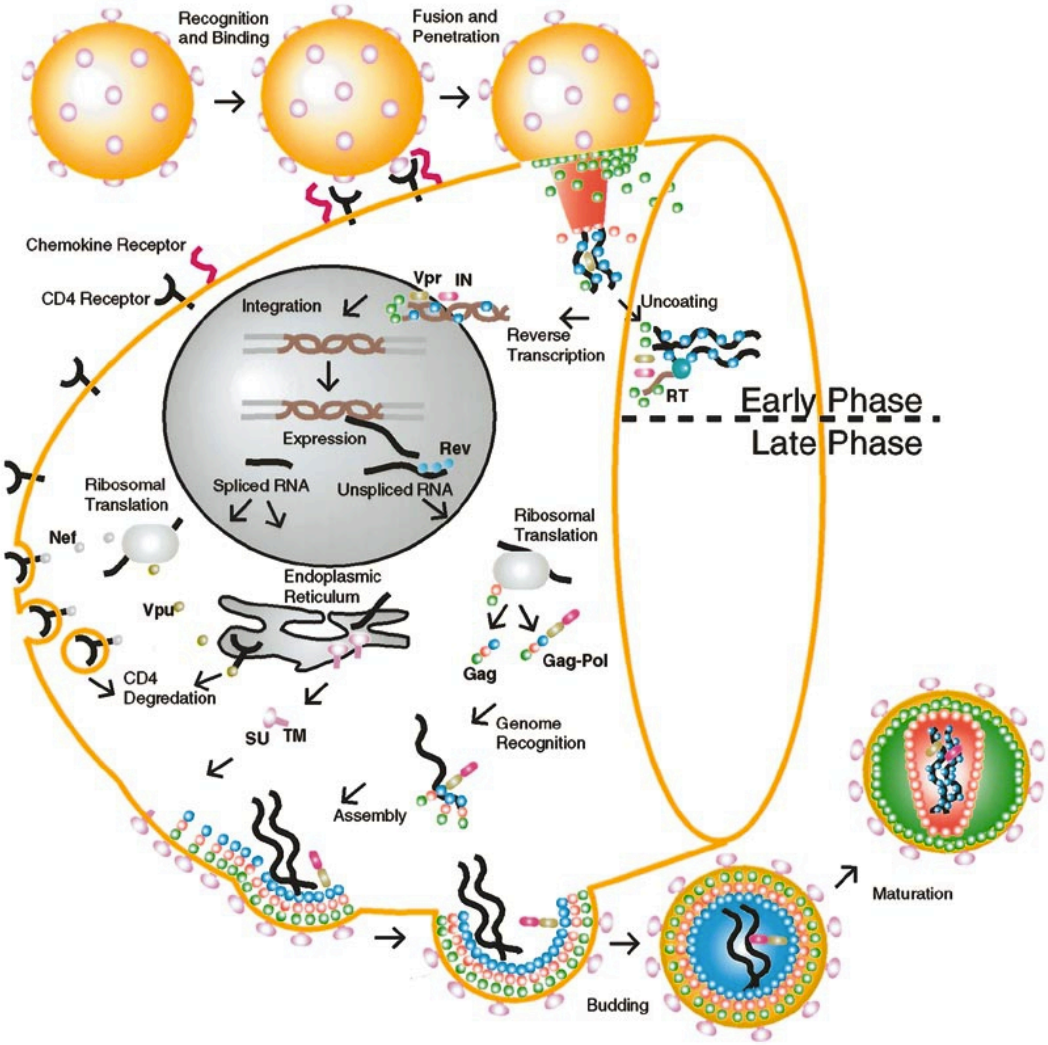


Figure 3: Replicative cycle of HIV (Turner & Summers, 1999)

Virus entry: The infection and entry starts as gp120 recognizes the CD4 receptor on the surface of the target cell and attaches to the host cell. Gp41 starches into the membrane of the host cell and, due to a conformational change induced by co-receptor binding to either CCR5 or CXCR4 on the surface of the host cell, membrane fusion of the virus and the cell is

initiated. (Clapham & R. A. Weiss 1997) The viral membrane with its envelope proteins becomes part of the host cell membrane.

Core entry: By a process, which is still not well understood, the p24 capsid core breaks open and several viral enzymes and two single strands of RNA are released into the cytoplasm of the cell. It seems that this process is facilitated or even guided by a cellular protein called cyclophilin A (hCyp-18), which can be found in the core layer of mature virions. (Luban 1996)

Reverse transcription: Reverse transcription is the characterizing step in the life cycle of retroviruses. The viral enzyme reverse transcriptase (RT) first uses the viral RNA as a template to form a hybrid double helix (viral RNA human DNA). Then it breaks down the RNA and in a second step forms a DNA double helix, which can be integrated into the host cell genome. During the reverse transcription many other proteins may also play a crucial role but are left out here for the sake simplicity. (Whitcomb & Hughes 1992)

Transfer to the nucleus: After the reverse transcription, the DNA together with accessory proteins form the pre-integration complex that is actively transported through a nucleopore into the nucleus of the host cell. (Whitcomb & Hughes 1992)

Integration: The pre-integration complex is guided towards a convenient region of the genome for the integration and is then processed and integrated into the host cell genome by the viral enzyme integrase (IN). The integrated viral DNA that becomes part of the cellular genome is referred to as provirus. It is the cause for lifelong infection. Only if the cell has been activated, the replication of the virus is promoted. (Ratner 1993)

Transcription and translation: The late phase of the viral life cycle starts with the activation of the host cell that starts to transcribe the integrated viral DNA and produces three kinds of mRNA transcripts: Highly spliced mRNAs are encoding the viral auxiliary proteins Tat, Rev and Nef. Unspliced mRNAs contain the information for Gag and Gag-Pol proteins and act as genome in the new viral particle. Single spliced mRNAs encode for the Env, Vpu, Vif, and Vpr proteins. (Turner & Summers 1999)

Early transcribed genes (highly spliced mRNAs) code for auxiliary proteins such as Tat, Nef, Rev and some others. (Wei et al. 1998) Tat acts as activator of the viral transcription. (K. A. Jones & Peterlin 1994) Nef seems to be a down regulator for CD4 receptors and also a weak inhibitor for the viral transcription. (Niederman et al. 1989) Rev acts as a switch between the

transcription of accessory proteins and the expression of structural proteins and enzymes. (Emerman et al. 1989)

In the second phase of the transcription, unspliced mRNAs are translated to form the Gag and Gag-Pol polyproteins. Rev mediates the export out of the nucleus and the polyproteins assemble under the cell membrane. (Ohno et al. 1998) The Env proteins are produced as normal membrane proteins involving maturation in the endoplasmic reticulum and transport to the membrane in Golgi vesicles. They are either incorporated into the membrane of the cell or into budding viral particles. (Freed 2001)

Budding: The process of budding starts with the assembly of about 2000 Gag proteins and 200 Gag-Pol proteins, some envelope proteins (gp120 and gp41) and two copies of single strand viral RNA. Viral tRNA and cellular proteins e.g. cyclophilin A are incorporated in the bud. Gag seems to mediate the formation of a buckle (Freed 1998) and Vpu (also a viral protein) probably releases the immature bud from the cell surface. (Schubert et al. 1996; Garnier et al. 1998; Wilk et al. 2001)

Maturation: The released particle has to become mature in order to infect new target cells. The viral protease is crucially involved in this process. The protease is encoded in the Gag-Pol precursor protein. Two molecules of Gag-Pol have to dimerize and by autocatalytic processing (John M Louis et al. 2007) protease is cleaved. Then the other proteins can be cleaved by the protease and after that assemble to form the matrix shell and the core and become functional integrase, reverse transcriptase and protease.

The maturation can be followed using electron microscopy and is completed if a conical core can be distinguished. (Ratner 1993; Adamson & Freed 2007)

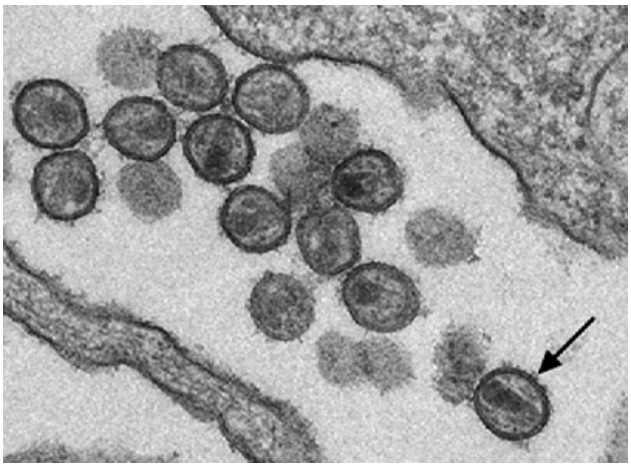


Figure 4: Virions after Budding, Mature particles show the conical core in the centre of the virion. (Adamson & Eric O Freed 2007)

1.2.5 Drug Targets

To find targets for anti HIV chemotherapy one must consider that HIV is an intracellular parasite. The most effective way to kill the virus would be to kill its host cell, but due to a very generalized infection and the rapid replication rates of the virus, such a chemotherapy would not be supported by the patients. Adequate drug targets are unique to the virus and, therefore, are specific and selective to kill only the virus and not the host cell. (Aktories et al. 2009; Schubert-Zsilavecz & Steinhilber 2010)

The first antiretroviral drugs were blockers of reverse transcriptase, an enzyme entering the host cell together with the viral RNA. This enzyme does not exist in healthy human cells and is therefore a perfect target for antiretroviral chemotherapy. Even though AZT is still on the market and still used, many patients have developed resistance against the drugs due to mutations in the virus. Therefore it is necessary to develop new agents to overcome the resistance and to be able to treat also patients with high resistant stems of the virus. A brief overview of the relevant mechanisms and targets for current anti-HIV chemotherapy is given below:

Entry: The most effective way to protect cells from a viral infection is to block the entry of the virus to the host cell. Even though it is very difficult to neutralize or block the receptors of CD4+ cells with antibodies and to target the viral surface proteins directly, some entry inhibitors have already reached the market. The entry itself can be further divided into two different stages: A) the receptor and co-receptor binding and B) later the fusion of the viral

membrane with the host cell's membrane. Each step is targeted by a specific class of inhibitors: A) Co-receptor antagonists inhibit the interaction between the co-receptor of the host cell and the viral envelope glycoproteins by blocking or manipulating the binding sites. B) Fusion inhibitors are the second class of entry inhibitors. They also act on the cell surface and prevent HIV from entering the CD4+ T-lymphocytes by blocking the fusion of the viral and host cell membranes. (Prabakaran et al. 2007)

A) Maraviroc is a CCR5 receptor antagonist blocking the co-receptor of CD4 and therefore hinders the entry of the virus into the host cell.

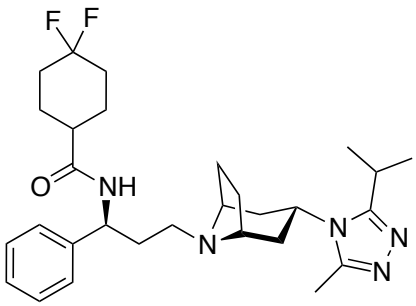


Figure 5: Maraviroc

B) Enfuvirtid is a fusion inhibitor that prohibits an essential conformational change in gp120, so that gp41 cannot starch into the membrane of the host cell to initiate the fusion. Enfuvirtid is a peptide with 36 amino acids and must be applied parenteral twice daily.

Reverse Transcription: There are currently two classes of inhibitors blocking reverse transcriptase:

a) *Nucleoside/Nucleotide reverse transcriptase inhibitors* mimic the nuclein bases of the host cell. During the transcription of the viral genome from RNA to DNA, the artificial nucleoside/nucleotide analogues are inserted to the nascent DNA strand. With the lack of a 3' hydroxyl group no other bases can be attached to the strand. The transcription stops at that point, releasing incomplete DNA fragments, which are degraded by host cell enzymes. Some of the nucleoside analogues also inhibit the mitochondrial DNA polymerase-γ which could lead to a mitochondrial dysfunction; (Schubert-Zsilavecz & Steinhilber 2010)

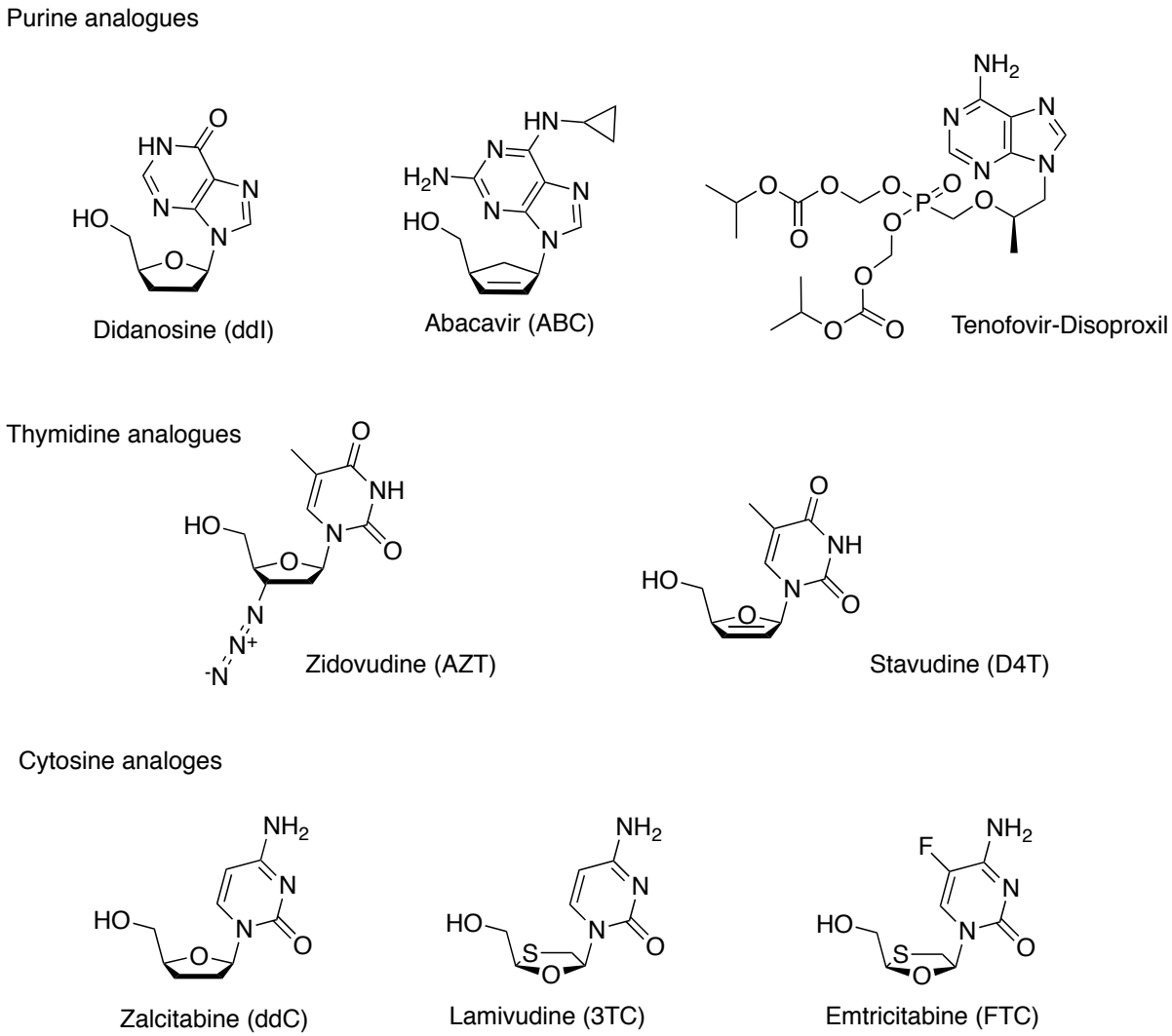


Figure 6: Nucleoside reverse transcriptase inhibitors

b) *Non-nucleoside reverse transcriptase inhibitors* bind to the enzyme in an allosteric binding site and block its function directly by arresting it in an inactive state. An important tool for the "survival" of the virus is missing and host cell enzymes will degrade the viral RNA before its transcription. These drugs are subsumed under the term non-nucleoside/nucleotide reverse transcriptase inhibitors because they structurally do not resemble the nucleosides/nucleotides. (Aktories et al. 2009)

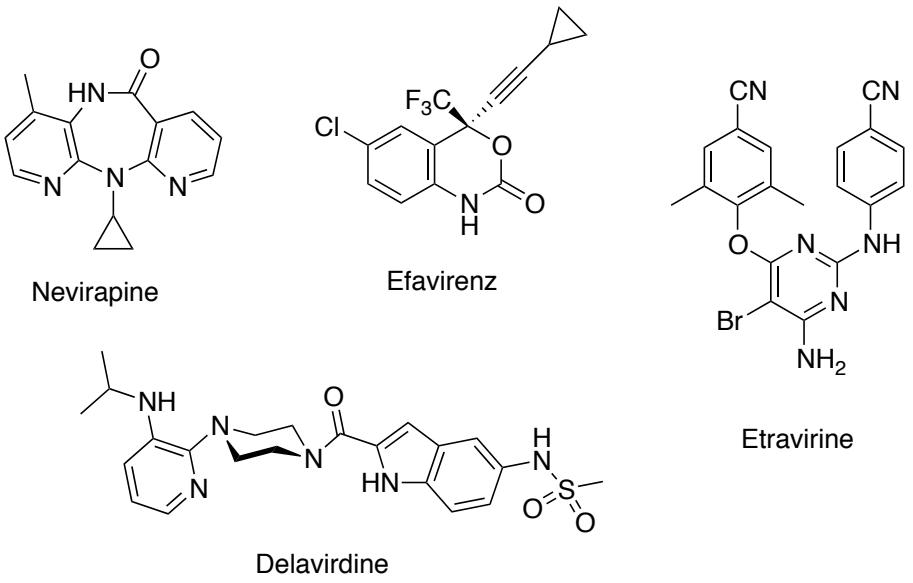


Figure 7: Non-nucleoside reverse transcriptase inhibitors

Integration: Only if the RNA was successfully transcribed to DNA and is then integrated into the host cell's genome, the virus stays in that cell for good and cannot be eliminated by the immune system anymore. The viral enzyme integrase (IN) catalyses this integration process and is therefore also a promising target for antiretroviral chemotherapy. Only recently the first integrase inhibitor (Raltegravir, 2007) entered the market. (FDA, 2007)

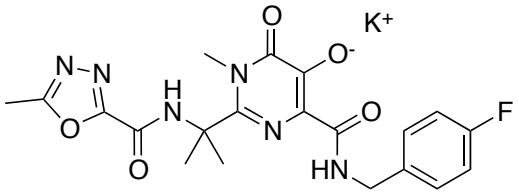


Figure 8: Raltegravir

After the integration the viral genome is treated as the host cell genome and the viral proteins are transcribed by the cellular protein biosynthesis apparatus. There is no possibility to target the translated proteins directly because they are behaving like cellular proteins.

Budding: By now the process of budding is not very well understood. The bud is formed by the Gag proteins that are aggregating under the HIV specific surface proteins and due to their interactions form a round particle that is later released from the cell surface. There is

by now no literature about inhibitors that trigger the budding process. (Adamson & Freed 2007)

Viral maturation: After the transcription and translation of the pro-viral genome into proteins by the host cell translation machinery, in the last step of the viral life cycle, the viral polyprotein precursor has to be processed by the protease that cleaves it into the structural proteins and functional enzymes. These are needed in the next cycle of the infection. This process happens during or shortly after budding of the new virions. If the viral protease is blocked, only uninfective particles can bud off the cell. (John M Louis et al. 2007)(See chapter 1.3.4 HIV protease inhibitors)

1.2.6 Current Clinical Treatments

1.2.6.1 Regular therapy

The first antiretroviral drug AZT (nucleoside reverse transcriptase inhibitor) entered the market in 1987 (about ten years after the discovery of HIV) and was the first great hope for people suffering from HIV/AIDS. But due to the high mutation rate in the virus and adverse drug reactions in patients, there was a need for better drugs against HIV. Other nucleoside analogues such as ddC (Zalcitabine), ddI (Didanosine) and d4T (Stavudine) approved between 1991-1994 were very limited in their use as monotherapy due to side effects. Only in 1995 it was found that a combination therapy of two different types of nucleoside analogues from the beginning of the therapy on was more effective than each of them in monotherapy. (Studies: Delta 1995 and ACTG 175 in 1996 –Hammer) (www.hivbook.com 2011) In late 1995, the first protease inhibitors entered the market. They were the first drugs where the knowledge of the HIV protease's crystal structure was used to design new compounds. (Lapatto et al. 1989; Navia et al. 1989; Wlodawer et al. 1989) The clinical trials of the first three Protease inhibitors ended in marketing three potent drugs between December 1995 and March 1996. Hoffmann - La Roche won the race with Saquinavir followed by Ritonavir (1996 Abbott) and Indinavir (1996 MSD).

From that point on, the antiretroviral treatment entered a new phase. The diction Highly Active Antiretroviral Therapy (HAART) was created which held a concept of combining two different nucleoside reverse transcriptase inhibitors with one non-nucleoside reverse

transcriptase inhibitor or one protease inhibitor. David Ho spread a totally new opinion about the treatment strategy: “Hit hard, Hit early” (D. D. Ho 1995). The development of PIs (protease inhibitors) was seen as the first breakthrough in antiretroviral therapy, but it did not stop there. (www.hivbook.com 2011) Only one year later, another new class of drugs entered the market. With Nevirapine (Boehringer Ingelheim, 21. June 1996), the first non-nucleoside reverse transcriptase inhibitor passed the clinical trials and was introduced to antiretroviral chemotherapy. But just after a few years of success with HAART, patients complained about some serious side effects as well as the high pill burden. Some patients were still resistant to therapy to some point or their viral load did not decrease under the clinically detectable threshold. There was a clear need for approvals of new drugs with less adverse effects and to overcome resistance of HIV against the used drugs. Later, different types of drugs like entry inhibitors, fusion inhibitors and integrase inhibitors were approved by the FDA giving a better perspective to patients in long-term treatment with a high rate of resistances. The typical HAART contains two different nucleoside reverse transcriptase inhibitors and one non-nucleoside reverse transcriptase inhibitor or protease inhibitor. Many of them are already provided in a predefined combination. Up to the end of 2011, 24 different mono- and 6 combination preparations have been approved (see Table 2). (www.hivbook.com 2011) Atripla® (Gilead, 12th July 2006) is the newest combination and has the great advantage that the patient has to take only one pill once a day. (De Clercq 2009)

Table 2: Approved antiretroviral drugs in the USA and Europe (De Clercq, 2009)

Generic name	**Manufacturer**	**Date of FDA approval**
Zidovudine	GlaxoSmithKline	19 March 1987
Didanosine	Bristol-Myers Squibb	9 October 1991
Zalcitabine	Hoffmann-La Roche	19 June 1992
Stavudine	Bristol-Myers Squibb	24 June 1994
Lamivudine	GlaxoSmithKline	17 November 1995
Saquinavir	Hoffmann-La Roche	6 December 1995
Ritonavir	Abbott Laboratories	1 March 1996
Indinavir	Merck	13 March 1996
Nevirapine	Boehringer Ingelheim	21 June 1996
Nelfinavir	Agouron Pharmaceuticals	14 March 1997

Delavirdine	Pfizer	4 April 1997
Efavirenz	Bristol-Myers Squibb	17 September 1998
Abacavir	GlaxoSmithKline	17 December 1998
Amprenavir	GlaxoSmithKline	15 April 1999
Lopinavir + ritonavir	Abbott Laboratories	15 September 2000
Tenofovir disoproxil fumarate (TDF)	Gilead Sciences	26 October 2001
Enfuviride	Hoffmann-La Roche & Trimeris	13 March 2003
Atazanavir	Bristol-Myers Squibb	20 June 2003
Emtricitabine	Gilead Sciences	2 July 2003
Fosamprenavir	GlaxoSmithKline	20 October 2003
Tipranavir	Boehringer Ingelheim	22 June 2005
Darunavir	Tibotec, Inc.	23 June 2006
Maraviroc	Pfizer	18 September 2007
Raltegravir	Merck & Co., Inc.	12 October 2007
Etravirine	Tibotec Therapeutics	18 January 2008
Fixed dose drug combinations		
Name	**Brand name**	**Manufacturer**
Lamivudine + Zidovudine	Combivir	GlaxoSmithKline
Abacavir, Zidovudine, Lamivudine	Trizivir	GlaxoSmithKline
Abacavir + Lamivudine	Epzicom (USA) Kivexa (Europe)	GlaxoSmithKline
TDF + Emtricitabine	Truvada	Gilead Sciences
Efavirenz, Emtricitabine + TDF	Atripla	Bristol-Myers Squibb & Gilead Sciences

1.2.6.2 Post exposure prophylaxis (PEP) (www.hivbook.com 2011)

The post exposure prophylaxis is a drug combination that can be used after the contact with HIV containing material that could lead to a persistent HIV infection for the exposed person. In general the PEP should be applied as soon as possible after the contact with the infected body fluids. Already 72h after exposure, the chance for success is rather low. Furthermore, it must be considered that HIV is not very easily transmitted compared to Hepatitis B or C. Therefore, the success and urgency of PEP depends very much on the transmission pathway, the infectivity of the viral subtype, the constitution and the immunostatus of the recipient and the viral load in the infective material.

By now there is no drug combination marketed as PEP, but there are suggestions and guidelines to combine NRTIs and PIs as well as Efavirenz to form the PEP.

Table 3: Combination possibilities for Post exposure Prophylaxis (PEP)

TDF + FTC (Truvada)		Lopinavir+Ritonavir (Kaletra)
or	+	or alternative Protease Inhibitor
ATZ + 3TC (Combivir)		or Efavirenz

In the following cases the application of PEP is highly recommended:

- deep cuts an stiches with visible blood on the instrument
- intravenous injection or transfusion of contaminated material like infected blood or blood products
- needle sharing among intravenous drug users
- stitches with needles that can hold a higher amount of fluid
- unprotected receptive sex with a HIV positive person

In hospitals and HIV clinics, there are guidelines to assess the situation and to recommend PEP. (Brodt 1998)

1.2.7 Transmission

The main transmission routes for HIV are

- unsafe sex with an HIV infected partner
- sharing needles for intravenous injection (drug abuse)
- vertical transmission from mother to child during birth and by breastfeeding
- transmission due to infected blood donations in countries where donated blood is not regularly screened for HIV

The risk for infection always exists if blood to blood contact or contact with the diverse body fluids takes place. Not all the body fluids contain the same amount of viral particles. For example, tears or saliva have nearly no potential to transmit HIV. The risk of infection also depends on the viral load of the patient.

In case of doubt, there is the possibility to take a post exposure prophylaxis. It is able to prohibit the integration of the viral genome into the genome of the host cell. (see 1.2.6.2) (AIDSHILFE wien, (www.hivbook.com 2011)

1.2.8 Course of the infection with HIV

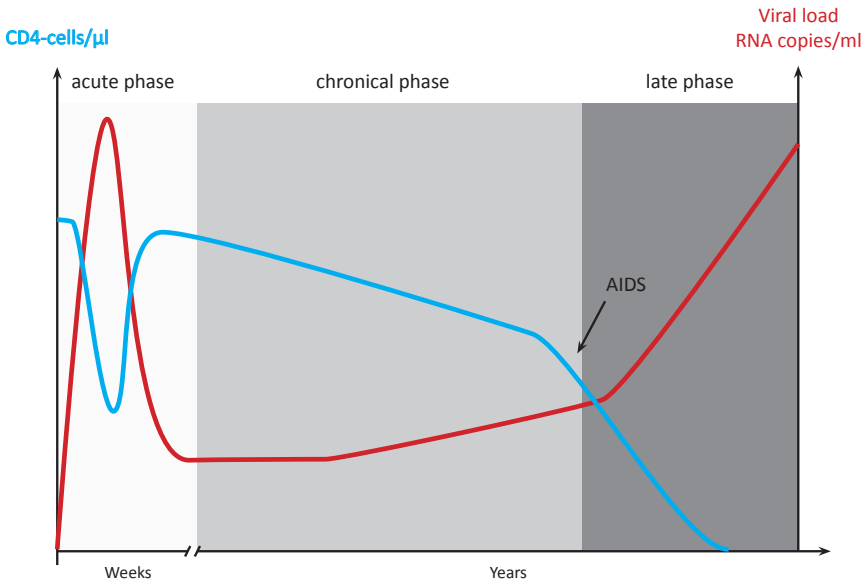


Figure 9: Course of the untreated HIV infection

The course of the HIV infection can be followed by regular blood testing. Two parameters, the CD4+ cells and the plasma viral load, are used to characterize the state of the infection and can also give a prognosis for the patient or indicate the efficacy of the anti retroviral treatment. (Herold 2011) The three episodes of the HIV infection corresponding to the tables 2 and 3 are shaded in grey (see Figure 9).

The virus infects CD4+ cells and while reproducing in the cells it hinders them in assisting the immune system by destroying those cells. The viral load in the patient's blood indicates the reproduction rate of the virus. The T-cell count is a parameter for the progress of the disease and the ability of the immune system to handle the virus.

Table 4: Classification of HIV-disease according to 2008 classification

Stage	AIDS defining illness	CD4+ T-cell count
1	None	>500/µl ≥29%
2	None	200-499/µl 14-28%
3	Documented AIDS defining illness	<200/µl <14%

Acute phase of the HIV infection or acute retroviral syndrome: About 30% of the newly HIV infected patients suffer from mononucleosis-like symptoms with fever, swollen lymphatic nodes, angina, myalgia etc.

Asymptomatic infection or latency time: The major group of patients undergo an asymptomatic infection where after about one to three or at least six months antibodies are present in the blood. Those patients are carrying the virus often without noticing for up to ten years. Even without any obvious symptoms these people can infect their sexual partner or children with HIV.

Persistent generalized lymphadenopathy or lymphadenopathy syndrome (LAS): Around 40% of all AIDS patients develop the lymphadenopathy syndrome. In that case they are positive for HIV antibodies and suffer from a generalized lymphadenopathy for over three months. At least two different lymph nodes apart from the inguinal region are enlarged. Normally no general symptoms as in the acute state can be observed.

Later, during the infection all the diseases for which HIV can be the initiating factor especially because of cellular immune disorder can occur. These diseases are not classified as AIDS-defining diseases. If one or more of these illnesses occur, an HIV infection should always be taken into consideration. Examples are: Subfrebile temperature (lower than 38°C) or chronic diarrhoea (longer than one month), Listeriosis Bacillary angiomatosis, Herpes zoster with more than one dermatome affected or recurrence in the same dermatome, Oral hairy leucoplakia (OHL), Oropharyngeal candidiasis (oral thrush), Cervical dysplasia or carcinoma in situ

AIDS-defining diseases: If the immune system is no longer able to cope with the virus the classical AIDS-defining diseases occur:

- Wasting syndrome

- HIV-associated Enzephalopathy (HIVE)

- Opportunistic infections defining AIDS:

    - Protozoan infections: cerebral toxoplasmosis, kryptosporidiosis or isosporasis

    - Fungal infections: Pneumocystis carinii pneumonia, candidiasis of the lung, trachea and bronchia, infections of the oesophagus, histoplasmosis

    - Bacterial infections: Bacterial pneumonia, atypical mycobacteriosis, tuberculosis, and salmonella sepsis

    - Viral infections: Cytomegalovirus infection, herpes simplex infection, progressive multifocal leukoencephalopathy

    - AIDS defining cancers: Karposi sarcoma, Non-Hodgkin-Lymphoma, invasive cervix carcinoma

(Herold 2011) CDC, 1993; www.hiv-leitfaden.de;

1.3 HIV-Protease

Below, the importance of HIV protease in the viral life cycle, its structure and function as well as the marketed inhibitors is presented.

1.3.1 Structure

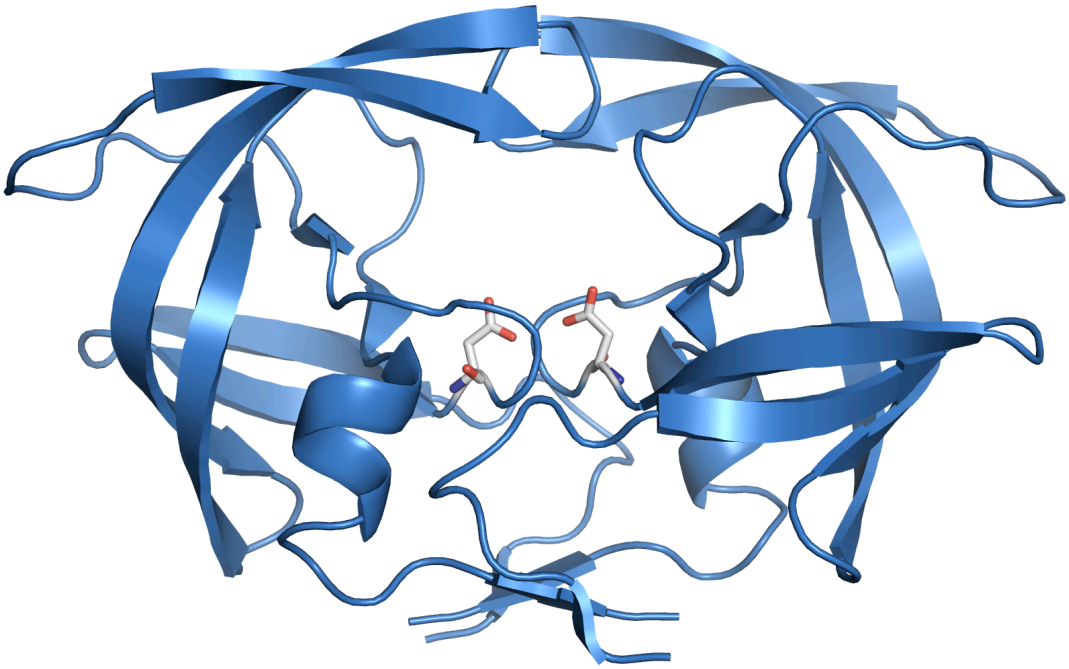


Figure 10: HIV protease (PDB: 3EKV) in the closed state with aspartate groups (Asp 25 and Asp 25') in the active site.

Together with reverse transcriptase and integrase, HIV protease is one of the essential enzymes in the human immunodeficiency virus. HIV protease is part of the aspartic proteinases family and its subunit contains 99 amino acids. In contrast to human aspartic proteinases, that structurally are bilobal polypeptide chains with functional domains, the retroviral aspartic proteinase has to form a C2-symmetric homodimer to be active. The structure (here PDB: 3EKV (King et al. 2012)) is characterized by several ß-sheets and only one short helix. Each subunit contributes one aspartic acid (Asp 25) to the active site and forms one half of the binding cavity. The binding cavity is covered by one flexible flap for each monomer that is rather open if the protease is empty and closed if a substrate or an inhibitor is bound. (See Figure 10) In the so-called half-open state, the cleaved substrate can be released. (Hornak & Simmerling 2007) The active site with the two aspartic acids is on the

bottom of a lipophilic tunnel that can accommodate peptides with eight amino acids in length. The nomenclature of the pockets after Schechter and Berger (1967) in the binding cavity is shown in Figure 11 (Schechter & Berger 1967). The dimer is mainly formed and stabilized by in total four antiparallel ß-sheets formed by the last residues of the C-terminus and the N-terminus of each subunit. (J M Louis et al. 1999; John M Louis et al. 2007; Gulnik et al. 2010) HIV protease was the first HIV protein that was resolved by X-ray diffraction. (Lapatto et al. 1989; Navia et al. 1989; Wlodawer et al. 1989)

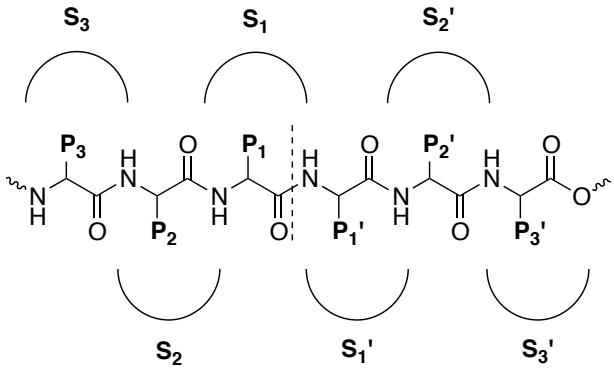


Figure 11: Pocket nomenclature after Schechter and Berger. The dashed line indicates the cleavage site of the peptide or protein cleaved by HIV protease. (Schechter & Berger 1967)

1.3.2 Role of the HIV protease

HIV protease is the crucial enzyme for the maturation of the nascent virions. It is able to cut the large poly-proteins into all their functional fragments – structural proteins and enzymes.

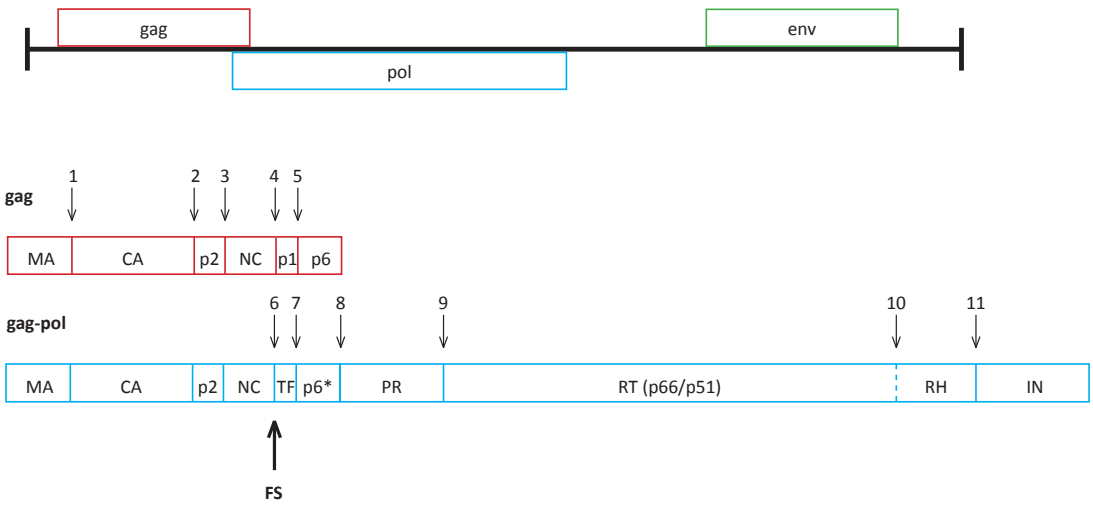


Figure 12: Cleavage sites of gag-pol protein p160
Abbreviations: MA matrix, CA capside, NC nucleocapside, TF transframe peptide, p6* p6-pol, PR protease, RT reverse transcriptase, RH RNase H, IN integrase.

The protease itself is synthesized in combination with other viral enzymes as a large polyprotein precursor called p160 Gag-Pol. This large polyprotein is only generated if a frame shift (FS marked with the bold arrow in Figure 12) during the translation at the ribosome takes place. That happens only at a rate of 5% among the single spliced mRNA transcripts. Normally in all the other cases, the translation results in Gag (p55). Only the Gag-Pol p160 polyprotein contains the information for the viral enzymes reverse transcriptase, integrase and protease. (Xie et al. 1999; Gulnik et al. 2010)

In the first step, HIV protease autocatalytically cuts itself out of the precursor protein. (Agniswamy et al. 2012) Then it can further process the remaining Gag-Pol protein and the Gag protein. The assembly of the structural proteins can be followed with the electron microscope. HIV protease is able to cleave the polyproteins at all the indicated cleavage sites (marked with enumerated arrows from 1-11 in Figure 12) exhibiting several different peptide sequences (view Figure 13). The secondary structure as well as the lipophilicity of the protein sequence seems to be the relevant signal for the recognition of the cleavage site on the part of the protease. The concrete amino acids only play an inferior role for the cleavage site recognition and the cleavage itself. Most frequently, HIV protease cleaves protein sequences between phenylalanine and proline in contrast to human proteases, which possess more specificity for other amino acids. (Aktories et al. 2009)

The characterization of HIV protease with X-ray diffraction in 1989 by Lapatto, Navia and Wlodawer and their co-workers lead to a rapid progress in the drug design against HIV. (Lapatto et al. 1989; Navia et al. 1989; Wlodawer et al. 1989) The first structure-based HIV protease inhibitors entered the market in 1995 (Saquinavir, Hoffmann-La Roche). Up to now (January, 2012) around 400 different HIV protease structures can be found in the PDB. (Berman 2000) They are composed of apo-proteins, protease-inhibitor complexes, and also several drug resistant mutants of the protease are published. By selective mutation analysis, it was shown that malfunction in the protease leads to uninfective viral particles. (Kohl et al. 1988) This shows the crucial role of the protease in the viral life cycle and points to a very successful strategy to reduce the viral load in the human body.

Cleavage site	P5	P4	P3	P2	P1	P1’	P2’	P3’	P4’	P5’
1	V	S	Q	N	Y	P	I	V	Q	N
2	K	A	R	V	L	A	E	A	M	S
3	S	A	T	A	M	M	Q	R	G	N
4	E	R	Q	A	N	F	L	G	K	I
5	R	P	G	N	F	L	Q	S	R	P
6	E	R	Q	A	N	F	L	R	E	D
7	E	D	L	A	F	L	Q	G	K	A
8	V	S	N	N	F	P	Q	V	T	L
9	C	T	L	N	F	P	I	S	P	I
10	G	A	E	T	F	Y	V	D	G	A
11	I	R	K	V	L	F	L	D	G	I

Figure 13: Sequences cleaved by HIV protease

1.3.3 Mechanism of action

As mentioned before the active site is situated at the bottom of a hydrophobic tunnel formed by the two subunits. The centre is built by the amino acids Asp 25, Thr 26 and Gly 27 of each of the subunits. The cleavage is catalysed by the two aspartic acids Asp 25 and Asp 25’ and one water molecule. This attribute classifies the HIV protease as a member of the aspartic proteinases family. It is an endopeptidase that cleaves peptides and proteins at specific sites with the help of the conserved water molecule adjacent to the aspartic acids in the centre of the binding site under hydrolysis. (Schubert-Zsilavecz & Steinhilber 2010; Davies 1990; Dunn 2002; Hyland et al. 1991; T. D. Meek et al. 1994; Northrop 2001; Polgar et al. 1994)

The more detailed mechanism is postulated as following (Figure 14): one of the aspartic acids acts as general acid because it is originally protonated. The other aspartic acid is

negatively charged and serves as general base. The adjacent water molecule acts as the nucleophile and attacks the carbonyl carbon atom of the substrate at the cleavage site. A tetrahedral transition state is postulated forming a geminal diol. The diol breaks down resulting in the cleavage of the peptide bond. (Brik & Wong 2003)

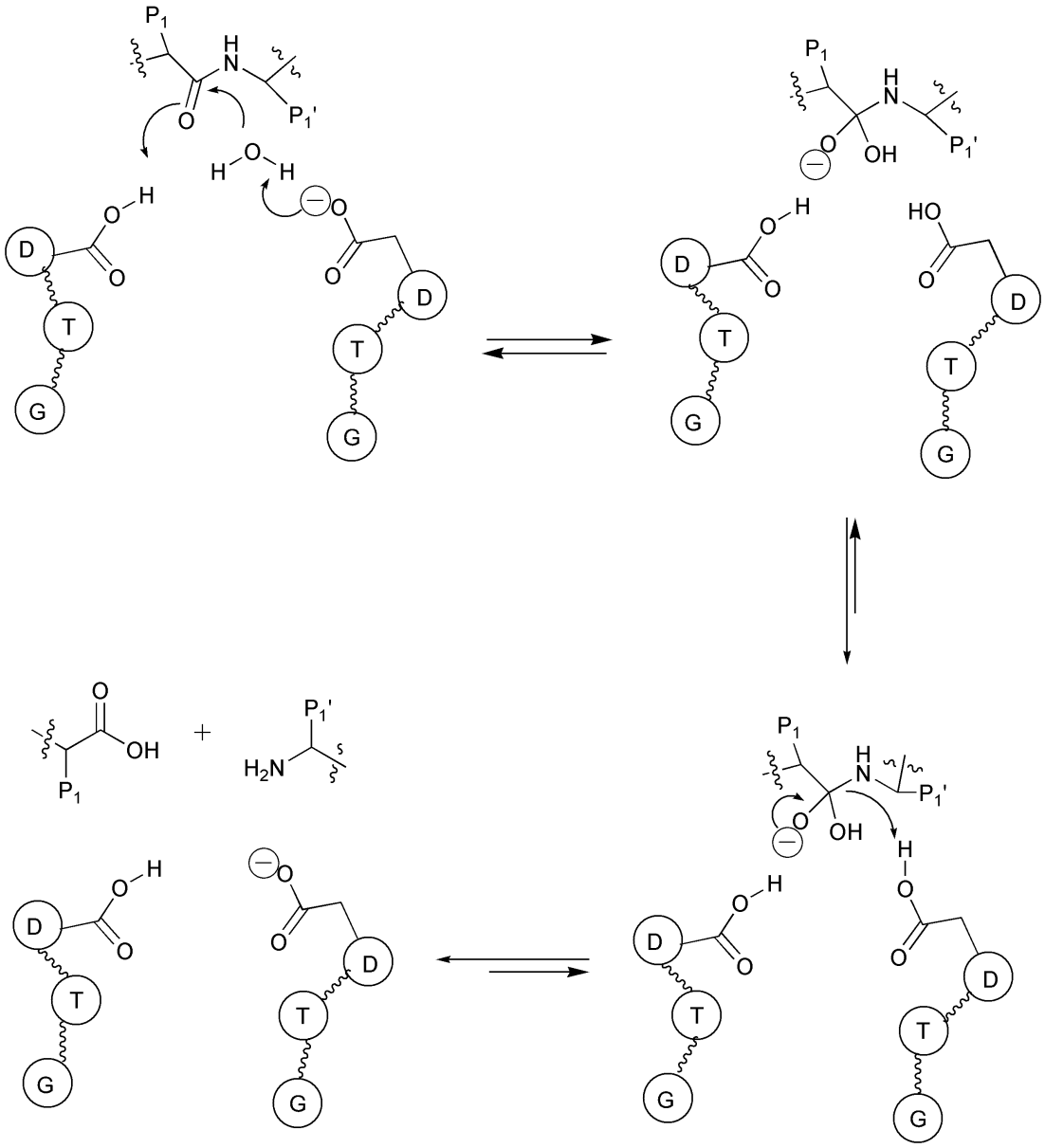


Figure 14: Catalytic mechanism of HIV protease (Brik & Wong, 2003)

The postulated mechanism can be found in experimental biochemical data as well as in structure analysis and theoretical calculations. (Brik & Wong 2003)

Based on that mechanism of action, different groups of peptidomimetics were designed as inhibitors. The atoms and bonds mimicking the peptide bond in the first cage are color-coded in red. (Figure 15)

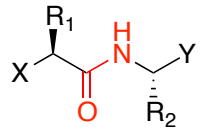
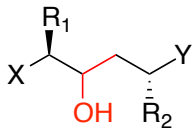
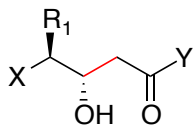
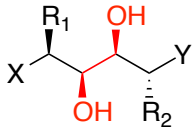
			
Peptide bond	S-Hydroxyethylene	Statin	Dihydroxyethylene

Figure 15: Peptide bond versus peptide bond mimicking substructures

1.3.4 HIV protease inhibitors

These nine protease inhibitors are now on the market.

Table 5: Currently marketed protease inhibitors

Name	Abreviation	Marketed in	Company
Saquinavir	SQV	1995	Hoffman-La Roche
Ritonavir	RTV	1996	Abbott
Indinavir	IDV	1996	Merck & Co
Nelfinavir	NFV	1997	Agouron Pharmaceuticals
Amprenavir	APV	1999	Vertex Pharmaceuticals
Fosamperenavir	FPV	2003	Glaxo Smith Kline
Lopinavir	LPV	2000	Abbott
Atazanavir	ATV	2003	Ciba-Geigy
Tipranavir	TPV	2005	Pharmacia Upjohn/Boehringer Ingelheim
Darunavir	DRV	2006	Tibotec-Virgo NV

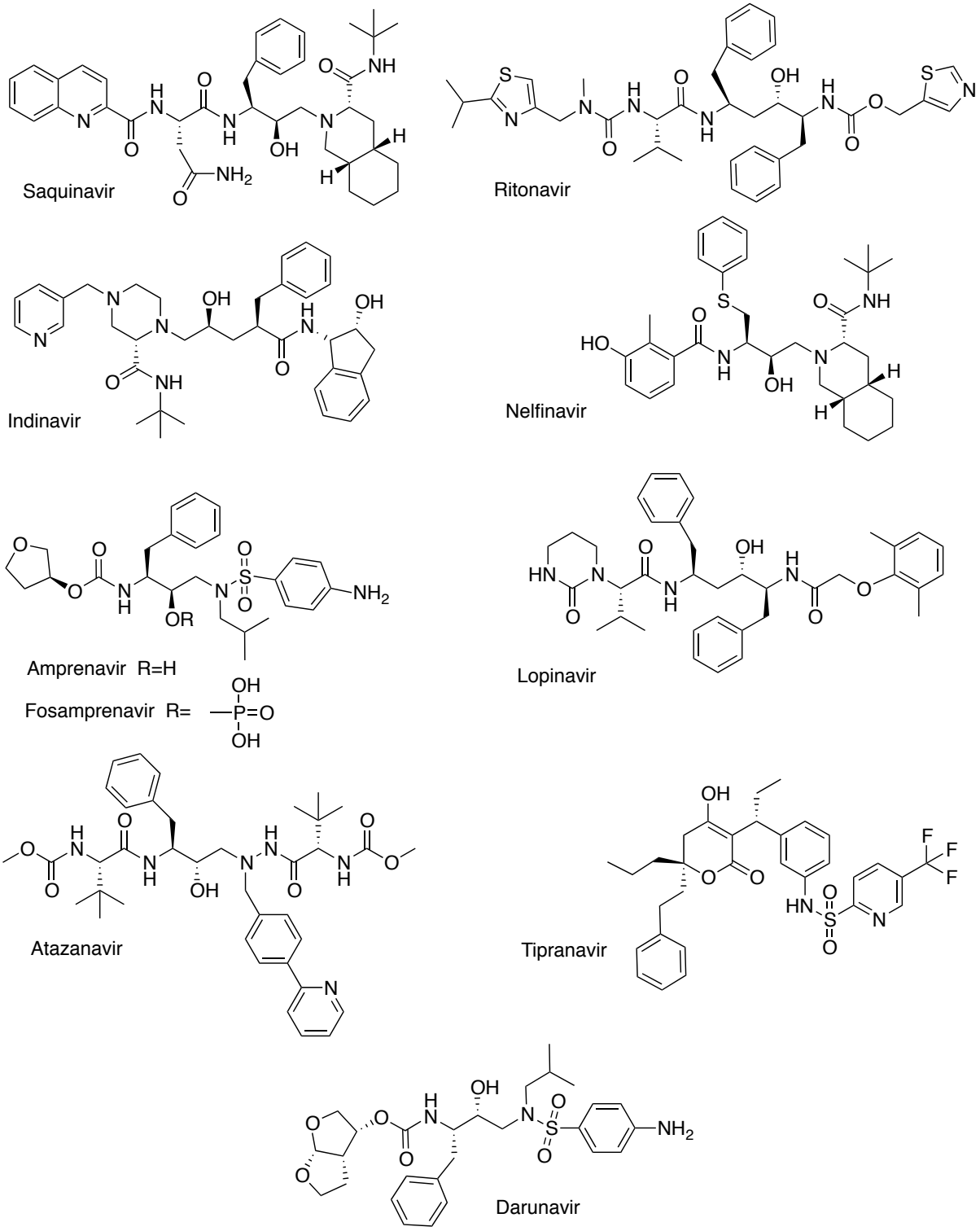


Figure 16: Marketed HIV protease inhibitors

1.3.5 Boosting with Ritonavir

Ritonavir is biologically not the most active protease inhibitor on the market, but it became one of the most important in therapy. Because of the multiple side effects and the low bioavailability, protease inhibitors had to be taken several times a day in former times. Their elimination in the liver is catalysed by the cytochrome P450 (CYP450) system. Ritonavir was found to have an inhibitory effect on cytochrome P450 3A4. In studies where a protease inhibitor was combined with ritonavir, the plasma levels of the other protease inhibitor were much higher. (Kempf et al. 1995) Only then it was possible to reduce the doses and especially the high pill burden for the patients. Another positive effect is the reduction of resistance of the virus, because it then has to overcome two different protease inhibitors at the same time in higher concentrations. The problem in inhibiting the CYP450 enzymes is that many other drugs, but also substances in our nutrition and endogenous products are metabolized by CYP450. This results in side effects and toxicity. Even the doses of the boosted protease inhibitors can reach dangerous levels if the monitoring is not carried out precisely and carefully enough. (Zeldin & Petruschke 2004; Moyle & Back 2001)

1.4 Computational Introduction

1.4.1 Drug discovery and design – a brief retrospective

In the very beginning drugs were empirically discovered when some plants or other agents showed effects on the human body. The next step was to isolate and purify drugs out of plants to have a better reproducibility in the effect and dosage of the drug. In that time the experiments with animals have been established and the search for bioactive agents also became interesting for synthetic chemists. More and more products came out of the colour industry. Molecular test systems and models were used to diminish the need for experiments with animals. With the developments in chemistry, physics and biology, theoretical methods like protein crystallography, molecular modelling, QSAR and many more could be used to computationally design molecules targeting a concrete protein especially with the help of new software packages. The discovery of genetic mechanisms and the ability to sequence the genome of many organisms as well as the possibility to manipulate the genome willingly, made it possible to evoke diseases in animals on a genetic level. That gave insights to new therapy strategies. (Gerhard Klebe 2009)

Today targets are often identified as proteins out of sequenced genomes. Then computational drug design is used to create or find small molecules, which later are optimized and synthesized and then tested on a molecular level in in vitro testing runs. Then to find the correct formulation and dose for the application as drug the substance is tested in animal models. This thesis copes with the level of computational and rational drug design against a known target protein. An overview of the drug design process is given in Figure 17.

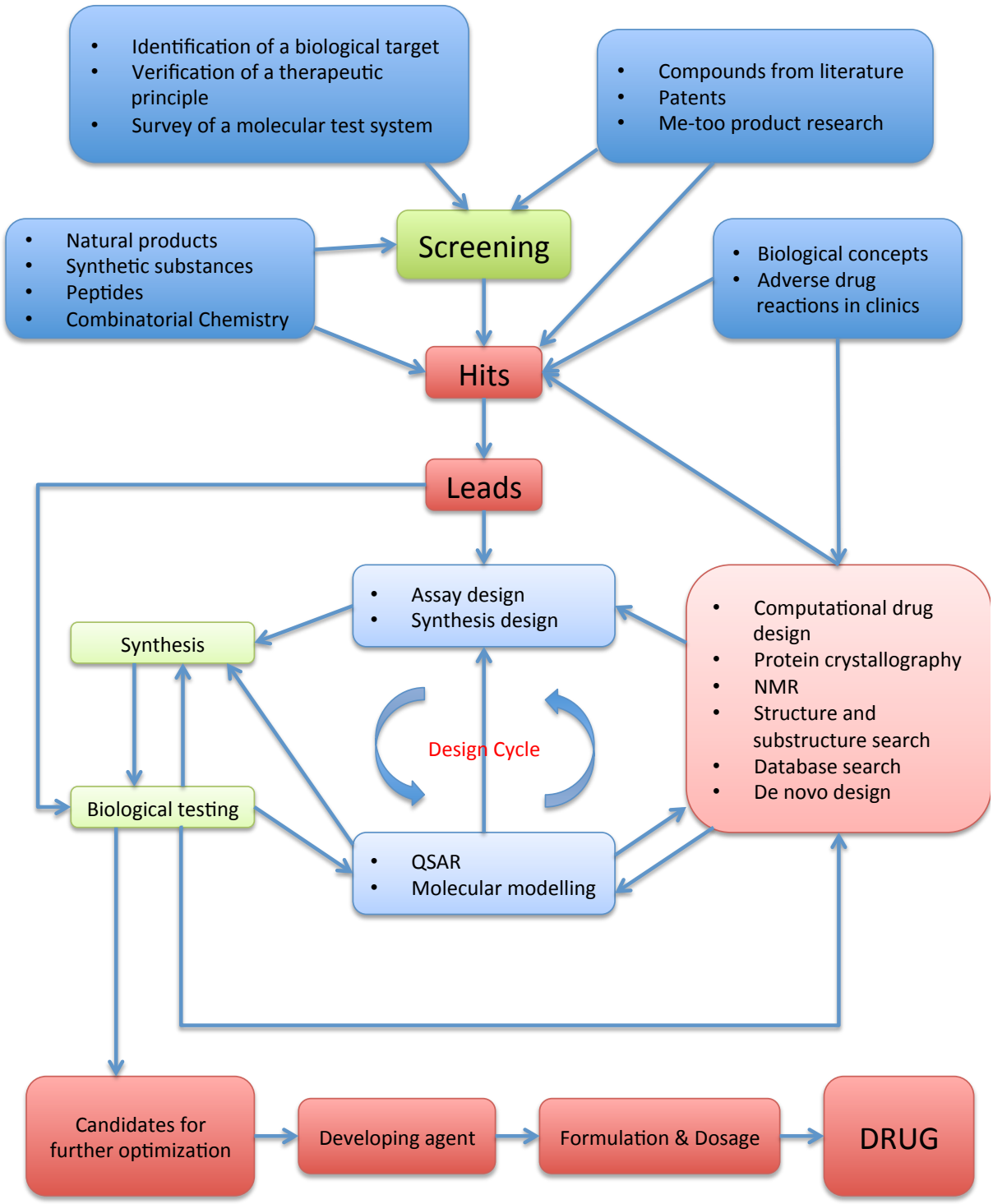


Figure 17: Drug design workflow (Gerhard Klebe 2009)

Today, rational drug design is significantly supported by the computer. Many tasks are perfectly made for its use. A future goal of computational drug design is to calculate physicochemical properties and interaction profiles for newly and virtually invented molecules with a high activity and a drug-like profile and send the results directly to a robot to synthesize them and further in that pipeline also test them with *in vitro* systems. (Markus Hartenfeller & Gisbert Schneider 2011b) Nowadays, that is still an illusion because the underlying physicochemical processes are not yet elucidated in an appropriate way, so it is still necessary to work with approximations and empirical rules. Even if it is possible to solve some of the basic equations for very small molecules, computational resources and time expenses might not be worth the probably small gain in exact knowledge compared to the output of simplified and knowledge based equations. It is therefore important to use the knowledge of medicinal chemists and the experiences in classical drug development to check the accuracy of the virtual models with synthesis and *in vitro* testing and then implement the gained knowledge into the development process for new software.

In the classical drug design process, synthesis and *in vitro* high throughput screening are the main focus of companies to detect new hits that can be further optimized. (Gerhard Klebe 2006)

The initial ideas for new structures to be synthesized can be derived from many different sources. They range from tested natural products and their derivatives to the rational and specific design of compounds against a known target. Nevertheless, many drugs are discovered only by chance (in the field often termed as serendipity).

In drug design, screenings are used to find new potential drug candidates (hits) or new promising scaffolds for further development. It would be impossible to analyse thousands of chemical compounds against hundreds of targets of interest in different assay conditions manually. High throughput screening (HTS) is an automatized technology where a large amount of chemical compounds is screened for activity against a set of targets. It is widely used in the pharmaceutical industry because of its time efficacy and reproducible results. However, detecting totally new chemical entities and scaffolds with HTS is very time-consuming and expensive. Often only the already synthesized compounds in the storage of the companies are tested several times. Pharmaceutical companies typically use HTS to regularly test their compound libraries containing thousands of synthesised molecules and intermediates in standardized assays against new targets to identify new hit candidates.

Therefore, the compounds must be available physically, which narrows the chemical search space drastically. To randomly synthesize molecules even with combinatorial chemistry would not really make sense because it is rather time-consuming and expensive to synthesize compounds without scientific backing. (Gerhard Klebe 2009; Markus Hartenfeller & Gisbert Schneider 2011b)

1.4.2 Computer aided drug design

The techniques of computational drug design can support the classical drug design process in every step. Especially in the rational design of compounds against targets with a known crystal structure the techniques are already very sophisticated. But also the *de novo* design techniques experience a revival at the moment and serve as idea generator for new scaffolds and for distinct molecules.

The two main techniques, synthesis and *in vitro* testing, are mirrored by the two *in silico* (computational) concepts *de novo* design and virtual screening. Nowadays, many algorithms, which were originally designed for a specific task and could be assigned to a certain computational concept have proved to be applicable for a broader field of computational puposes. Often they are implemented in multifunctional software packages. Scoring functions and docking may serve as examples. (Gisbert Schneider & Fechner 2005)

In this thesis, only a subset of virtual screening techniques is used especially to check the possible fitting of the new compounds into the binding cavity with structure-based docking and then rank the molecules on the basis of the quality of the interactions with the protein.

The main focus of the computational part was set on *de novo* design in the present work.

a) *De novo* Design: (Markus Hartenfeller & Gisbert Schneider 2011a; Markus Hartenfeller & Gisbert Schneider 2011b; Gisbert Schneider & Fechner 2005; Gisbert Schneider & Baringhaus 2008)

A *de novo* design program "invents" new structures that should fit into the binding pocket or show a certain similarity to a template compound. Historically, the concept was structure-based, but several newer programs implementing a ligand-based approach were developed. 15 to 20 years ago, when the first *de novo* design programs entered the field of drug design, they were limited in designing synthesizable and meaningful structures. However,

developers optimized their programs in several directions so now they can be seen as useful idea generators for synthetic chemists. The first approaches were atom-based. The assembly of molecules atom by atom was a time-consuming process. Moreover, atom-based approaches often led to chemically intractable structures and molecules that were not drug-like. Nowadays, most of the software packages use a fragment-based approach, automatically narrowing the chemical space to a more drug-like subspace because the fragments are (i) taken from already existing drugs, or (ii) from building block libraries. *De novo* design software has to deal with the following tasks: (Markus Hartenfeller & Gisbert Schneider 2011b)

1. How to generate the molecules?
2. How to assess the quality of the designed molecules?
3. How to optimize the compounds in the next design round and navigate to promising areas in chemical space?

The aim of all these programs is to imitate a chemist who rationally designs molecules for a certain purpose.

1. How to generate the molecules? – Assembly

The methods to choose or generate the molecules are diverse, but one possibility would be to use retrosynthetic rules *i.e.* to cut down known drug-like molecules into small pieces. The fragments are then re-assembled. If the software works ligand-based, the design process usually starts with one fragment and grows the molecule with synthesis-based rules at specified attachment sites up to a size limit. In the structure-based approach, fragments are placed in the binding pocket and can either be grown until they produce steric clashes with the protein, or two or more fragments are placed into the protein pocket complementary to the main interaction sites and are then linked with a convenient linker.

2. How to assess the quality of the designed molecules? – Scoring

In the ligand-based approach, the designed molecules are evaluated according to the similarity to the reference molecule. In the structure-based design, molecules are typically assessed by docking the new compounds into the receptor binding-pocket, and then subsequently scoring them according to their placement.

3. How to optimize the compounds in the next design round and navigate to promising areas in chemical space? – Optimization

- Compound optimization and navigation towards the optimum in chemical space

The goal is to optimize the designed compounds towards a global optimum in the chemical space. Therefore, after construction of the molecules and also after each round of the design cycle, the compounds must be assessed. In structure-based design, the evaluation is usually done by docking, in fragment-based design usually with the help of other scoring functions. Since only fragments are used in molecule construction, the basically endless chemical space is already narrowed. However, it is still difficult to lead a design process towards a global optimum and to find that optimum.

- Some stochastic optimization techniques used in diverse software packages:

The chemical space could be seen as a landscape on which one can “walk along” by changing the molecule (starting point) in a smooth manner. (Gisbert Schneider 2008) Tiny changes in the molecule are expected to evoke small steps along the wavy surface. (Johnson et al. 1990; Rechenberg 1973)The changes are assessed by the scoring function that leads the changes towards the optimum. The optimum can be seen as the top of a hill on that surface where any change in the molecule cannot bring any further improvement. This principle is based on local neighbourhood exploration. An important factor is the size of the steps along the search and optimization path. Small changes in the molecule should bring only small progress in optimization. If the steps are too large or the surface is very rough, the step size can no longer be extrapolated to the improvement of the molecular structure. (Gisbert Schneider 2008)

Other ways to optimize the compounds are evolutionary algorithms (EA). This principle comes from Darwin’s evolution theory and uses mutation and cross-over to create compound diversity. These techniques can sometimes overcome local optima and find the global optimum. (Gisbert Schneider & Fechner 2005; Hiss et al. 2010)

Other implementations use particle swarm optimization (PSO) (Kennedy & Eberhart 1995). A particle swarm (one can imagine animals in a swarm) is searching for the best solution (e.g. food) and therefore is communicating with the other members of the swarm. Promising points are visited by more particles and less promising regions are neglected. All the

information is stored in a common memory and the particles can learn from each other. (Markus Hartenfeller & Gisbert Schneider 2011b)

In addition to stochastic optimization procedures, deterministic approaches are used for the compound optimization. For example, all the possible fragments are docked into a receptor pocket and obtain a score. For compound assembly, only the fragments with a good score are first considered and the scores are added. This saves computing time, but it might result in erroneous scores because binding-free energies are usually not only additive.(Proschak, Sander, et al. 2009; Nikitin et al. 2005; G Klebe & H. J. Böhm 1997; M Hartenfeller 2010)

The previous listing only gives a brief and incomplete overview of the most commonly used techniques in the field of computational drug design, especially virtual screening and *de novo* design.

Table 6: Fragment-based *de novo* design software (table taken from Schneider & Fechner, 2005 and Hartenfeller & Schneider, 2011)

Method/Name	Scoring Strategy		Year	Reference
	Ligand Based	Structure Based		
HSITE/2D Skeletons		X	1989	(Danziger & P M Dean 1989; Lewis & P M Dean 1989b; Lewis & P M Dean 1989a)
3D Skeletons		X	1990	(V.J. Gillet et al. 1990)
Builder v1		X	1992	(Lewis et al. 1992)
LUDI		X	1992	(H. J. Böhm 1992a; H. J. Böhm 1992b; H. J. Böhm 1993; H. J. Böhm 1994)
NEWLEAD		X	1993	(Tschinke & N. C. Cohen 1993)
SPLICE		X	1993	(C. M. W. Ho & Marshall 1993)
GroupBuild		X	1993	(Rotstein & Mark A. Murcko 1993)
CONCEPTS		X	1993	(Pearlman & Mark A. Murcko 1993)
SPROUT		X	1993	(Gillet et al. 1993; Mata et al. 1995)

MCSS & HOOK		X	1994	(Eisen et al. 1994; A Miranker & M Karplus 1991; Caflisch et al. 1993)
GrowMol		X	1994	(Bohacek & McMartin 1994)
Chemical Genesis	X	X	1995	(Glen & Payne 1995)
PRO LIGAND	X	X	1995	(Clark et al. 1995; Waszkowycz et al. 1994; Westhead et al. 1995; Frenkel et al. 1995; Clark & Murray 1995; Murray et al. 1995)
SMoG		X	1996	(DeWitte & Shakhnovich 1996; Ishchenko & Shakhnovich 2002; Grzybowski et al. 2002)
CONCERTS		X	1996	(Pearlman & Mark A. Murcko 1993)
RASSE		X	1996	(Z. Luo et al. 1996)
PRO SELECT		X	1997	(Murray et al. 1997; Eldridge et al. 1997)
SkelGen	X	X	1997	(Todorov & Dean 1998; Stahl et al. 2002)
Nachbar	X		1998	(Nachbar 1998; Nachbar 2000)
Globus	X		1999	(Globus et al. 1999)
DycoBlock		X	1999	(Liu et al. 1999; Zhu, H. Yu, et al. 2001)
LEA	X		2000	(D Douguet et al. 2000)
LigBuilder		X	2000	(Wang et al. 2000)
TOPAS	X		2000	(G Schneider, M. L. Lee, et al. 2000; G Schneider, Clément-Chomienne, et al. 2000)
F-DycoBlock		X	2001	(Zhu, H. Fan, et al. 2001)
ADAPT		X	2001	(Pegg et al. 2001)

Pellegrini & Field	X	X	2003	(Pellegrini & Field 2003)
SYNOPSIS		X	2003	(Vinkers et al. 2003)
CoG	X		2004	(N. Brown et al. 2004)
BREED	X		2004	(Pierce et al. 2004)
Nikitin		X	2005	(Nikitin et al. 2005)
LEA3D		X	2005	(Douguet et al. 2005)
Flux	X		2006	(Fechner & G. Schneider 2006; Fechner & G. Schneider 2007)
FlexNovo		X	2006	(Degen & Rarey 2006)
Feher	X		2008	(Feher et al. 2008)
GANDI	X	X	2008	(Dey & Caflisch 2008)
COLIBREE	X		2008	(Markus Hartenfeller et al. 2008)
SQUIRRELnovo	X		2009	(Proschak, Zettl, et al. 2009; Proschak, Sander, et al. 2009)
Hecht & Fogel	X	X	2009	(Hecht & Fogel 2009)
FOG	X		2009	(Kutchukian et al. 2009)
MED-Hybridise		X	2009	(Moriaud et al. 2009)
MEGA	X	X	2009	(Nicolaou et al. 2009)
Fragment-Shuffling	X	X	2009	(Nisius & Rester 2009)
AutoGrow		X	2009	(Durrant et al. 2009)
NovoFLAP	X		2010	(Damewood et al. 2010)
PhDD	X		2010	(Q. Huang et al. 2010)
GARLig		X	2010	(Pfeffer et al. 2010)
DOGS	X		2010	(Markus Hartenfeller & Gisbert Schneider 2011a; Markus Hartenfeller et al. 2012)

b) Virtual screening:

A large number of theoretical molecules, of which the physico-chemical properties can be calculated by the computer, is ranked by scoring functions. Usually, the molecules are classified as active or inactive according to the scoring function. To score the molecules, two basic approaches can be pursued:

1) Structure-based approach: The structure of the target protein or at least a similar protein with the same interaction profile is known. The tested compounds are assessed according to their fitting into the binding pocket of the protein. Docking and Scoring are two examples commonly used in this field.

2) Ligand-based approach: In ligand-based design, the structure of the target protein is not known. However, at least one but usually several active compounds against a target are known but not the exact binding pocket or the interaction mode. In that case, the molecules are ranked by their similarity to the reference ligand or to a consensus model of several reference ligands and their activity can be compared to each other. Then, a model to predict other active compounds can be derived. For the ligand-based design, the most popular ways for screenings are structure and substructure searches, QSAR (quantitative structure-activity relationships) models or pharmacophore modelling and search. In ligand-based design, huge databases can be screened to identify compounds for deeper investigation. (Gerhard Klebe 2006), Schneider, 2010)

1.4.3 DOGS – Design of Genuine Structures (Markus Hartenfeller et al. 2012)

The software DOGS - Design Of Genuine Structures is a reaction-driven de novo design software to generate new drug candidates or lead molecules on a ligand-based principle which should mimic a template structure in its biological, chemical and pharmacological properties. No structural information about the receptor is needed to generate new structures. Only one single reference molecule is required for the generation of molecules that are similar to the reference compounds, but have a different molecular scaffold.

The outstanding advantage of DOGS is to generate new molecules on the fly, while providing the structures of the required educts and a complete synthetic route for each designed compound. The fragments used to build the new compounds are manually selected and pre-

processed out of a vendor catalogue, when the software was designed, and can therefore directly be ordered for the synthesis of the in silico generated molecules.

The building blocks are defined by their attachment points stored in a database. 58 different reactions can be applied to the molecules. The reactions where recommended by medicinal chemists because of their high yields, synthetic feasibility, laboratory equipment needed to perform them, chemicals used, applicability to many building blocks among others. Only the smallest part of the structure that is needed to participate in the reaction is considered.

Choice of fragments for the software generation:

The start fragments undergo numerous pre-processing routines: For example, only fragments with a molecular weight between 30 and 300 Da are selected, structures with too many atypical atoms for drugs are not considered, and reactive compounds are eliminated. By elimination of unwanted fragments, the chemical space is drastically narrowed and the software automatically yields more drug-like molecules in its outcome. Fragments with more than one attachment point are not considered in the primary database either.

All building blocks are downloaded as part of the Sigma Aldrich catalogue accessed at the ZINC database (Irwin & Shoichet 2005), pre-processed with the MOE (Chemical Computing Group) washing routine (for details see 2.1.1) and subsequently collected in the building block database MySQL. Therefore, all of these substances are purchasable and can directly be ordered for synthesis if needed.

Molecule construction with DOGS:

Every molecule (start fragments and intermediate products) is processed by "dummy" reactions. First, an attachment point (functional group that can interact in the reactions of the reaction library) has to be chosen. Most of the molecules exhibit more than one group that could interact in the reactions from the library. A reaction is performed with a dummy fragment only containing the features, atoms and bonds necessary to interact with the reaction partner. In case of a one step reaction, the modification to the building block can be applied directly without creating a dummy fragment. In case of a two-step reaction, a dummy fragment has to be used for every reaction partner.

In the first cycle, all the building blocks from the library are processed to find the *n* most promising ones. The parameter *n* is set by the user and indicates the number of start fragments and consequently also approximately the number of constructed molecules resulting from that design run.

The dummy products resulting from these steps are scored by the scoring function that implicitly scores the most promising reactions. For the *n* best dummy products, the reactions are applied to yield the first generation of offspring – the first intermediate products. To each of the starting fragments, the best reaction is then applied with all possible building blocks suitable to participate in the chosen reaction. These products are ranked by the scoring function. The molecules with the best score are further considered for the next design cycle, whereas all the others are discarded. If two molecules are considered best, both are kept for the next round so the underlying start fragment will yield in at least two different final products.

With the intermediated products the procedure is the following:

First the appropriate reaction is chosen by dummy fragments added to the intermediate products and the rank of these dummy products according to the scoring function. The reaction that leads to the best dummy product will then be carried out with all the building blocks that can participate in the given reaction. The resulting products will be evaluated with the scoring function and only the best molecule is kept as intermediate product.

This process is repeated until at least one of the stop criteria is fulfilled.

1) The first stop criterion is defined by a mass threshold defined as 70% to 130% of the mass of the reference molecule. The molecule grows until it is passing the lower mass boundary.

After that, an addition of a new building block is only accepted if it improves the score of the molecule and does not exceed the upper mass boundary. Otherwise, the last intermediate product is added to the list of final products without the changes that lead to the unfavourable score or mass.

2) The second stop criterion regulates the number of fragments added to the starting fragment even if the mass threshold is not exceeded. This contributes to the aim of the program to create synthesizable molecules. The more virtual reaction cycles take place the

more real reactions are necessary to synthesize the molecule. If the number of growing steps exceeds the stop criterion, the last intermediate is added to the list of final products.

The parameter *n* defines the number of start fragments for the run. For every start fragment considered within the *n* most promising fragments DOGS usually creates at least one molecule.

A start fragment does not form any product if no building block can be attached to the start fragment before exceeding the lower mass boundary. DOGS removes that fragment from the list of considered start fragments and adds the next building block to the list to make sure that at least *n* molecules are constructed.

Usually one run creates more molecules than the considered start fragments because if two intermediate dummy products with their underlying reactions are equally ranked on top, both reactions are applied to the intermediate product and the synthesis pathway splits into two branches.

The design routine of DOGS is strictly deterministic and two design runs with exactly the same parameters will therefore always yield exactly the same results.

The following scheme is taken from Hartenfeller et al. 2012: (Markus Hartenfeller et al. 2012)

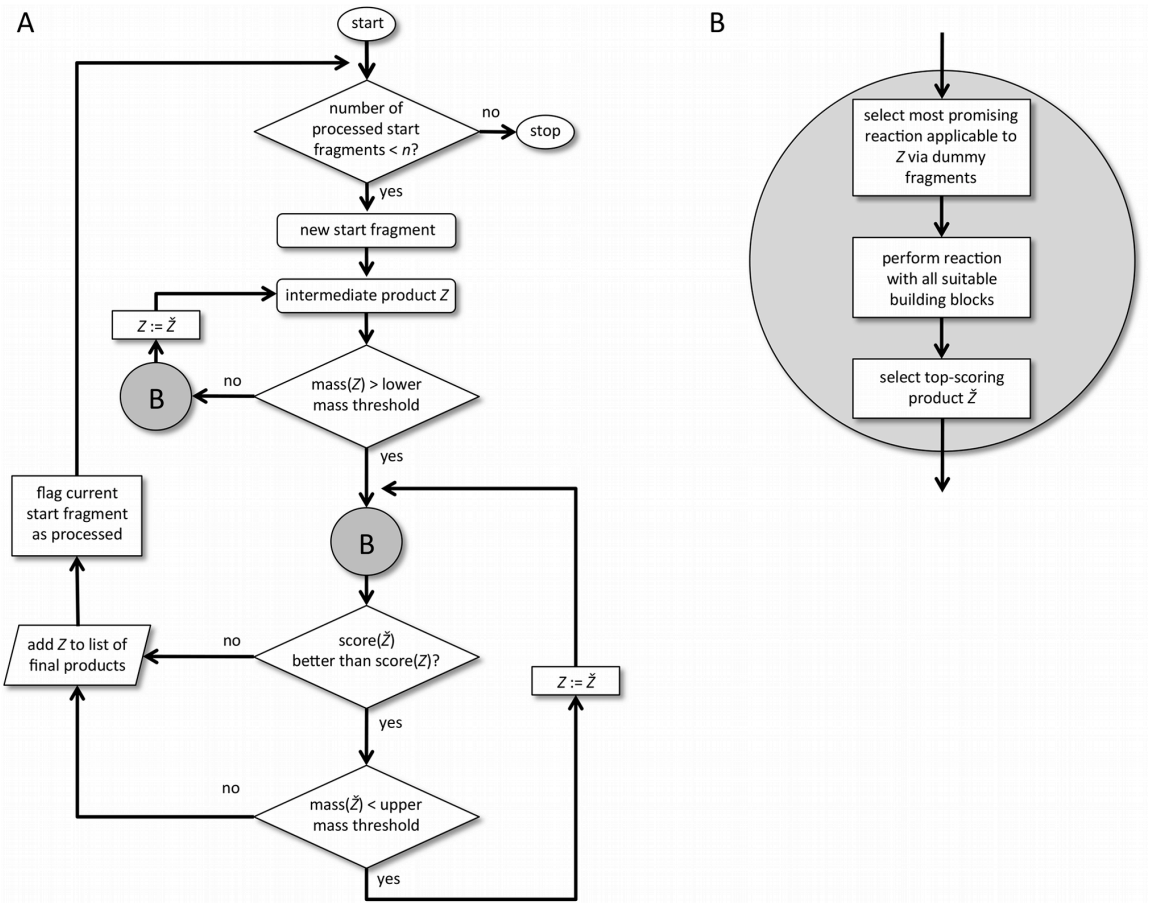


Figure 18: Flowchart of molecule generation with DOGS
A: Flowchart of the DOGS design algorithm. The stop criterion for maximum number of reaction steps is not included. B: Detailed description of flowchart element B (grey circle). It comprises the key steps to modify intermediate product Z in order to yield Ž by applying in silico reactions.

Scoring:

DOGS has implemented two types of molecular representations, the molecular graph and the reduced graph representation. Both are topological 2D representations of the molecules. For the molecular graph representation, all hydrogen atoms are removed from the molecule and the atoms referred to as vertices are labelled with pharmacophoric features (discrete variables). The bonds are referred to as edges between the vertices. The reduced graph representation of the molecules subsumes ring systems with one or more rings as one vertex, resulting in a higher level of abstraction. For the generation of the structures in the present thesis, only the molecular graph representation was used.

The ISOAK scoring function was originally developed for virtual screening purposes and is a method to compare two molecules and to compute their similarity. (Rupp et al. 2007) It is a 2D graph kernel method. All products (dummies, intermediate, final) are evaluated with the ISOAK scoring function. The function calculates the similarity between the newly synthesized compound and the template molecule and thereby evaluates the constructed compound.

The comparison works as following: The graphs are first compared according to the labels assigned to every vertex. The similarities are stored in a matrix. Then the environment of the vertices is taken into account and the similarity according to the neighbourhood of every vertex is calculated. (Two vertices are similar if their neighbourhood is similar.) Both factors influence the overall similarity that is calculated as a weighted sum according to the ISOAK parameter α that can be chosen by the user and has to be a number between 0 and 1. Higher values of α take the neighbourhood of the vertices more into account than lower values.

In the next step the calculated similarities are used to make sure that the molecules are compared to each other in a meaningful alignment and orientation. The scoring function assigns one vertex of the smaller graph to exactly one corresponding vertex of the larger graph. None of the vertices can be assigned to two vertices of the other graph (and vice versa).

The overall similarity is then computed by summing up the similarities between the corresponding vertices of the two graphs.

1.5 Aim

In the current thesis the goal was to develop inhibitors for HIV protease, which is an essential enzyme of the virus. This specific protein was chosen because of its well-known binding pocket and several marketed inhibitors that could serve as templates for the *de novo* design software DOGS (Markus Hartenfeller et al. 2012) that we used to design new potential compounds.

DOGS, a ligand-based de novo design software, it needs only one known inhibitor as a template molecule to create a set of diverse compounds that exhibit a certain similarity to the template, yet with different molecular scaffolds. The most promising compounds were manually selected and synthesised according to the synthesis pathways suggested by DOGS. After purification and analysis, these compounds were tested in a fluorescence-based assay for inhibitory activity against HIV protease.

The following list gives an overview of the workflow of the project. The main tasks were to

- collect activity and structural data for HIV protease inhibitors on the market,
- find a convenient reference molecule as a template for the *de novo* design software,
- prepare the template molecule(s) for the *de novo* design software,
- design the structures,
- evaluate the generated structures with different *in silico* screening methods,
- select promising structures for the subsequent synthesis,
- synthesise the compounds with the suggested pathways given by DOGS,
- adapt the synthesis if necessary and improve the reactions,
- purify and characterize the products,
- test the compounds for potential activity in a fluorescence assay kit.

2 Practical Part

2.1 Computational Part

HIV protease inhibitors on the market and in the drug development pipeline, as well as their molecular scaffolds and their crystal structures were compiled and examined. One of these structures was selected to be the reference molecule for the structures to be designed with the *de novo* design software DOGS. The inhibition constants (K_i) and the dissociation constants (K_D) for the inhibitor molecules, if available, were collected. The molecular weight and the SlogP (is the log of the octanol/water partition coefficient (including implicit hydrogens). This property is an atomic contribution model (Wildman & Crippen 1999) that calculates logP from the given structure; i.e., the correct protonation state (washed structures). Results may vary from the logP(o/w) descriptor. The training set for SlogP was ~7000 structures.) were calculated with the software package MOE. (Molecular Operating Environment (MOE), 2012.10; Chemical Computing Group Inc., Montreal, QC, Canada) The search for those structures was started in the PDB, which is helpful to automatically select meaningful crystal structures with its inhibitor for the docking experiments carried out later.

Table 7: Molecules active against HIV-protease

Name	MW	SlogP	K_i	K_D 25°C
Amprenavir	519.66	2.4028	0.23 nM[a]	0.725 nM[c] 1.10 nM[b]
Atazanavir	704.87	4.4780	0.24 nM[a]	0.403 nM[b,c]
Brecanavir	705.85	3.8887	15 fM[e]	
Darunavir	547.67	2.3753	16 pM[f]	
DMP-450 Mozenavir	536.68	4.7659	0.28 nM[h]	
Indinavir	613.80	3.2287	0.31 nM[a]	1.68 nM[c] 1.07 nM[b]

Lopinavir	628.81	4.3281	1.3 pM 0.08 nM[a]	0.101 nM[b,c]
Nelfinavir	567.80	4.7476	1.71 nM 0.54 nM[a]	3.67 nM[c] 1.64 nM[b]
Palinavir	708.90	4.7035	6 pM[g]	
PL-100	624.80	4.1196	0.036 nM [i]	
Ritonavir	720.96	6.4379	0.59 nM[a]	0.831 nM[c] 0.608 nM[b]
Saquinavir	670.86	3.0924	0.23 nM[a]	0.297 nM[c] 0.315 nM[b]
Tipranavir	602.67	7.6370	0.008 nM	

[a] (Shuman, Vrang, et al. 2004); [b] (P. Markgren et al. 2002); [c] (Shuman, Hämäläinen, et al. 2004); [d] (Alterman et al. 1998); [e] (Lalezari et al. 2007); [f] (Ghosh et al. 2008); [g] (Lamarre et al. 1997); [h] (Hodge et al. 1996), [i] (Dandache et al. 2007)

The currently marketed HIV protease inhibitors are mostly peptidomimetics that imitate the cleavage site of the natural substrates of HIV protease. Therefore, those molecules can block the HIV protease in its transition state. The inhibitors were also explored in the corresponding crystal structure to learn about the protein ligand interactions. More about the crucial interactions between the HIV protease and its inhibitors can be found in Section 2.1.8 Cherry picking.

2.1.1 De novo design with DOGS - Parameters

The selected template molecules were prepared for the DOGS experiments and some test runs were performed with most of the marketed molecules. Due to the specific properties of Amprenavir such as low molecular weight, size and the SlogP and the limited time period for

the experiments it was finally decided to only use Amprenavir as a template compound (Figure 18)

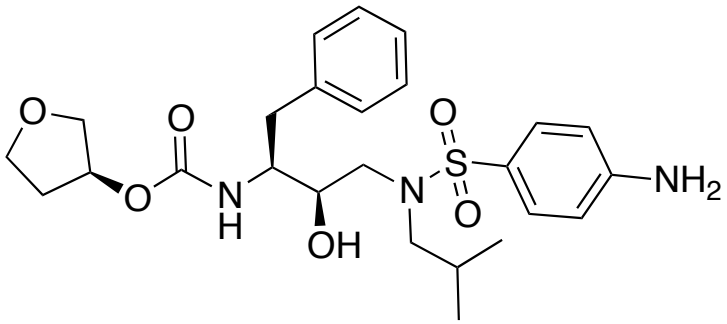


Figure 19: Amprenavir

Beside the runs with Amprenavir test runs with Darunavir, Nelfinavir and Brecanavir were performed, but due to time limitations it was decided to further pursue only one of the reference compounds (Amprenavir).

The chemical structure of Amprenavir (extracted from PDB: 3EKV) was drawn in ChemDraw and later prepared with the MOE washing routine: The three dimensional X-ray conformation was protonated, explicit hydrogens were added and the protonation state was set to “deprotonate strong acids and protonate strong bases”. The reference molecule was saved in the SDF (Structure Data Format) file format (Dalby et al. 1992).

Then a folder for the results and the files generated during the run had to be created. DOGS creates two files for each designed molecule. The first is an SDF-file where the molecular structrue is stored. The second contains the suggested synthesis pathway and the building blocks for the synthesis indicated as the product number of the vendor catalogue where it is extracted from and the name of that vendor catalogue. One additional file is created for each run, where all the created molecules are subsumed. This file is called “all_created_mols.sdf”.

In the created folder, the dogs_config.txt file with the input parameters necessary for each run was stored as well. In that file, the absolute paths to the input files have to be specified. The locations of the files were separated by a tab. The most important setting is the output directory. If that path is not changed, old results may be overwritten. The config file contains the following parameters set by the user:

- Path to the reference ligand (absolute path)
- Label for the generated molecules (it is helpful to encode some of the parameters in that label)
- Number of start fragments considered
- Graph representation method
- ISOAK α (parameter for similarity comparison of the scoring function)
- Path to output directory
- Path to the atom type definitions

The program is written in Java and can be executed from the command line in the following way: A new shell (mac terminal window) has to be opened. In the jar file, the provided memory for the run is specified, followed by the path to the program file and then the path to the config file.

Java –Xmx “memory_in_MB” –jar “path_to_jar” “path_to_config-file”

For all runs the graph representation 0, molecular graph, was used as well as 200 start fragments for each run. The ISOAK α was varied between 0.1 and 0.9 in 0.1-steps.

2.1.2 DOGS results

Table 8 gives the statement of the program when it has finished the molecule construction: The number of constructed molecules for each run, the overall fitness and the standard deviation of the fitness. Fitness here is a synonym for similarity to the reference molecule. The higher the fitness the more similar are the generated molecules to the given reference with respect to their molecular graph. The direct neighbourhood of each atom has more

influence during the assignment of the two graphs than the correct atom type. If the ISOAK α is smaller the atom types play a major role for the score of the new molecule.

Table 8: Statement of DOGS after one run

ISOAK α	Number of Molecules	Overall fitness	STD
0.1	230	0.067	+-0.032
0.2	238	0.088	+-0.031
0.3	237	0.111	+-0.030
0.4	227	0.137	+-0.032
0.5	265	0.160	+-0.03
0.6	254	0.198	+-0.033
0.7	272	0.240	+-0.033
0.8	316	0.298	+-0.039
0.9	299	0.397	+-0.043

In total, 2338 molecules were constructed. After the fusion of the "all_constructed_mols.sdf" files of each α value into one SDF, the multiples were counted and removed. 856 unique molecules resulted from the nine DOGS runs. The ten most frequently constructed molecules are shown in Figure 20. In the legend the molecule code and their occurrence frequency *n* are indicated.

One molecule can be constructed more than once in one single run and can also be constructed in several different runs. In Figure 20 only the molecule tags for the smallest α are listed. The evaluation of DOGS which delivers a distance or fitness value assesses the same molecules differently depending on the ISOAK α value for the run. Therefore a molecule appearing in the run α 0.1 has a different distance/fitness than the same molecule appearing in the run α 0.9.

By optical inspection of the molecules generated by DOGS the first eye catching substructure is the quinazolin-4(3H)-one structure. It appears in 9 of the 10 most frequently created structures and is also frequently seen in the other structures. Compared to Amprenavir it

might replace the sulphanilamide group. Its ketone group may form similar interactions to the protein and the size might be similar to the aniline ring.

The carboxylic acids often appear in the intermediates and give attachment points for the software. It also makes the molecules more soluble in water.

In the centre of the molecules frequently an amide group is situated which is evident if one consider the desired products as peptidomimetics. In the listed molecules the carbamide group is appearing 3 times. It is also part of Amprenavir in combination with the tetrahydrofuranol. Four of the six molecules chosen to be synthesized feature this tetrahydrofuranol substructure. One of them contains the quinazolin-4(3H)-one.

The hydrazine group at the end of longer aliphatic sidechains may mimic the amino group at of the sulphanilamide substructure in Amprenavir.

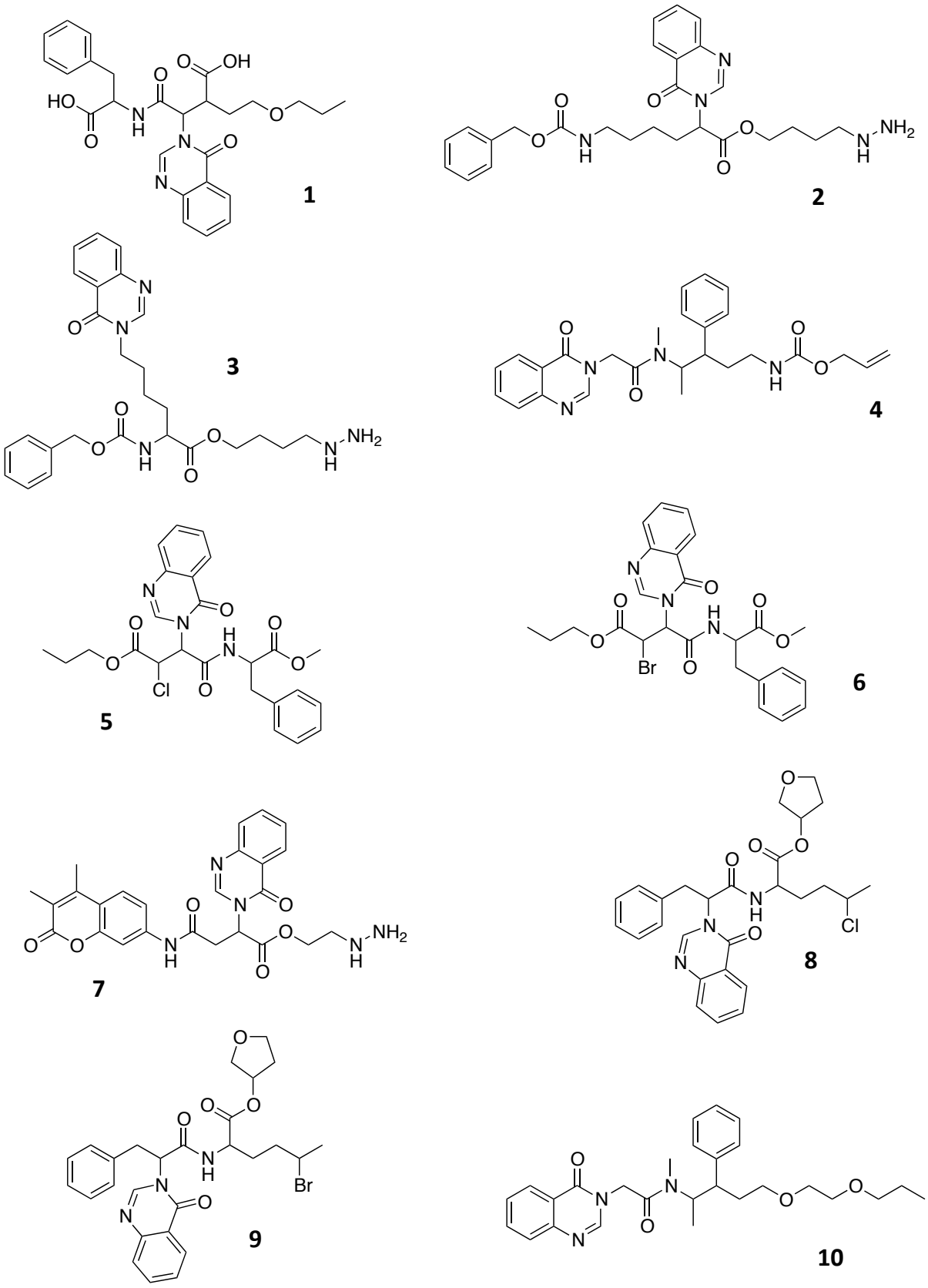


Figure 20: Most frequently constructed molecules
1 mol_107_a03 n=20; **2** mol_5_a01 n=18; **3** mol_1_a01 n=18; **4** mol_185_a01 n=16; **5** mol_a23_a02 n=15; **6** mol_18_a02 n=15; **7** mol_93_a01 n=15; **8** mol_85_a07 n=12; **9** mol_84_a07 n=12; **10** mol_145_a05 n=12

2.1.3 Scaffold analysis

With all the structures generated by DOGS, a scaffold analysis based on the scaffold definition of Bemis and Murcko (G W Bemis & M A Murcko 1996) was carried out. The scaffolds were extracted from the molecules with the program "Stripper" (version 1.0.4.). (Silicos NV, 2005-2011) Copyright (C) 2005-2011 by Silicos NV.

A central goal of *de novo* design software is the so-called scaffold hopping. Usually the scaffolds of marketed drugs are protected intellectual property. But by bioisosteric replacement it is possible to find molecules and scaffolds or interaction patterns that exhibit the same biological effect by different chemical groups. (Gisbert Schneider & Baringhaus 2008)

In the scaffold analysis that we carried out the focus was set on the basic pattern of the molecule without functional groups, to see which similarities in the shape of the designed molecules are obvious.

The ten most frequently designed scaffolds can be found in Figure 21 including their frequency. In total 278 different scaffolds were created. About 40% of the generated molecules can be described by one of the 10 most frequent scaffolds. Most of the scaffolds have a linear structure including a ring system with two six membered rings. By visual inspection of the molecules this is already obvious because many designed molecules contain a quinazolin-4(*3H*)-one structure.

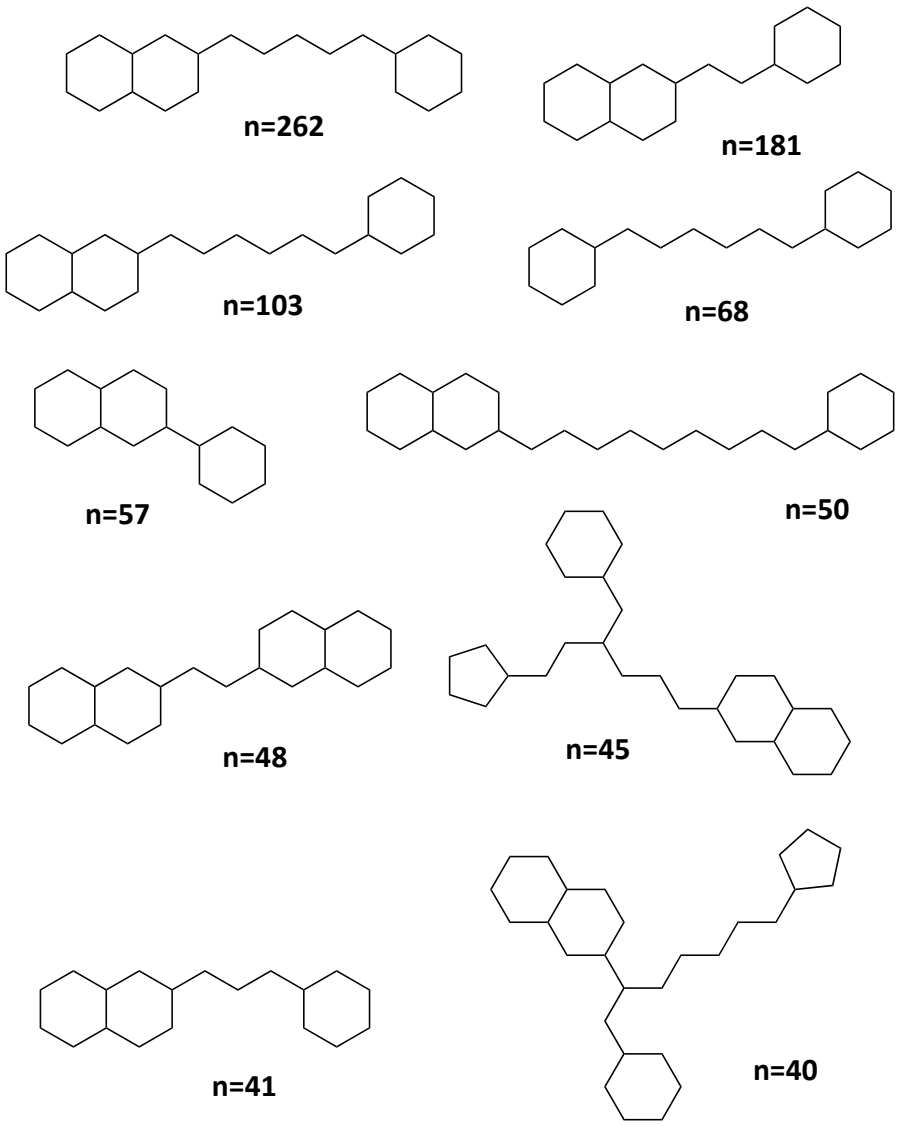


Figure 21: Most frequently constructed scaffolds

2.1.4 Docking

After the transformation of the 2D molecules designed by DOGS into 3D molecules with the help of the software CORINA (J. Gasteiger et al. 1990), it was attempted to gain some information about the possible interaction mode of the molecules with the HIV protease and thereafter select the most promising structures. For that purpose docking was used.

Docking places every molecule into the binding pocket and tries to calculate the binding energies for different docked conformations and to assign as many interaction sites on the protein as possible with complementary functions of the given ligand. Thus, docking faces two problems: Firstly, the ligand has to be placed into the binding pocket. Secondly, the docking pose has to be evaluated with a scoring function.

2.1.5 GOLD – Genetic Optimization for Ligand Docking

For docking, the software package GOLD (version 5.0.1.) (G. Jones et al. 1997) was used. GOLD (Genetic Optimization for Ligand Docking) is a software for automated docking of ligands into a receptor binding pocket. It has an implemented genetic algorithm (GA) for the conformational search that results in the stochastic generation of the docking poses.

The program is able to give the ligand full flexibility and even allows some flexibility in the protein if required. The energy function used in the program GOLD is rather knowledge based on the observations of contacts between the ligand and the receptor in crystal structures. GOLD does not use the conformations to be then scored directly, but in a bit-string representation.

Although a GA does not always find the optimal solutions, it performs well in more complex queries. Therefore, it is a suitable technique to handle the flexibility of large ligands and interacting side chains in the predictions.

2.1.6 Input parameters for GOLD

The structures suggested by DOGS were selected due to their molecular weight. The cut off was set at 470 Da. These small structures from all nine DOGS runs were then docked into the HIV protease substrate pocket of the 3EKV structure from the Protein Database (PDB: 3EKV) (King et al. 2012) after being transformed to 3D structures with CORINA.

The settings were chosen with the help of the GOLD wizard. First the binding pocket was defined by the co-crystallized ligand Amprenavir in the 3EKV structure. By default all protein residues within the radius of 5Å were selected to form the interaction site for docking. Even if there is only one single atom of the residue in the defined region the whole residue was selected. The option to define the binding pocket by a reference ligand has the advantage that the docked compounds will bind in the same region and can be directly compared with the template that was also used for generating the new structures with DOGS.

The 3ekv structure was first prepared for the docking by my colleague Felix Reisen. Sidechains appearing twice in different orientations were removed and adapted. Structure was cleaned up.

For the calculation of the GOLD-Score, the default parameters of the GOLD scoring function were used. Before the docking, the missing hydrogen atoms were added to the protein structure. The water molecules and the co-crystallized ligands (Amprenavir and some solvent molecules) were removed. The protein was kept rigid while the docked molecules obtained full flexibility during the docking procedure. Up to ten poses per molecule were calculated.

It can be observed that larger molecules usually obtain better scores because they can form more potential interactions with the protein. However, the aim was to find small and drug-like molecules as new potential HIV protease inhibitors. The marketed structures are already very large for drug molecules and possess some bioavailability problems, so a cut off of 470Da was set. After the generation of the 3D conformations of the 204 smallest molecules the resulting 227 molecules were docked into the HIV protease active site.

2.1.7 Docking results

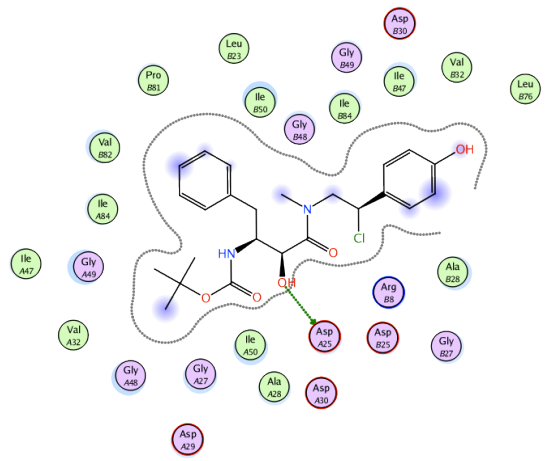
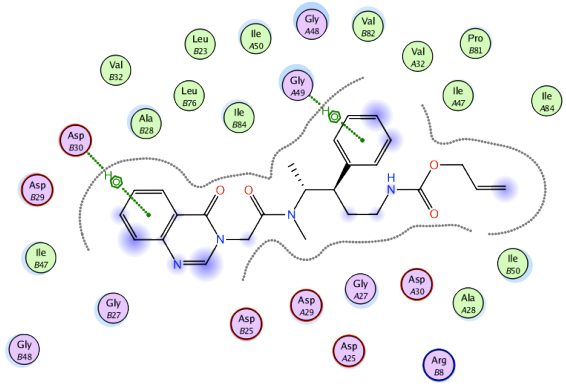
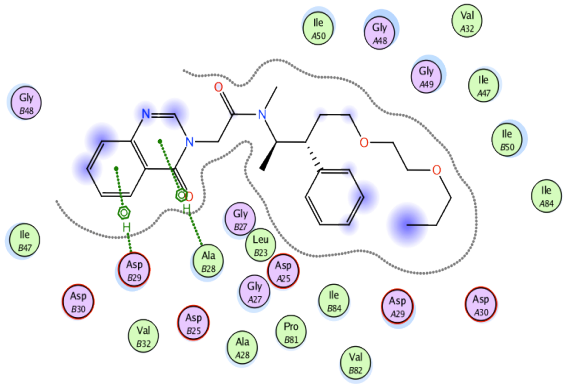
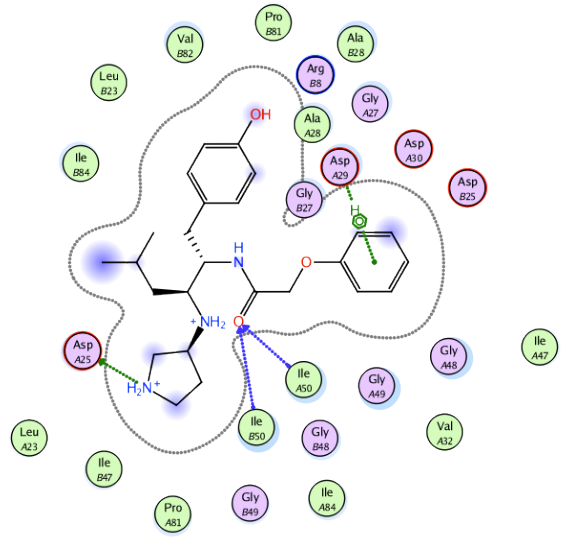
Up to ten poses were generated for each of the 227 molecules resulting in totally 2232 docking poses and scores. The best pose for every molecule according to its GOLD-Score was kept while the others were not considered. After the filtering of the poses, the ligand interactions and the proposed binding mode for the molecule were investigated more closely. Not only the molecules with the best GOLD-Score were selected because then again only the largest molecules out of the already narrowed subset of small molecules would have been obtained. The molecules were especially sorted according to their molecular weight.

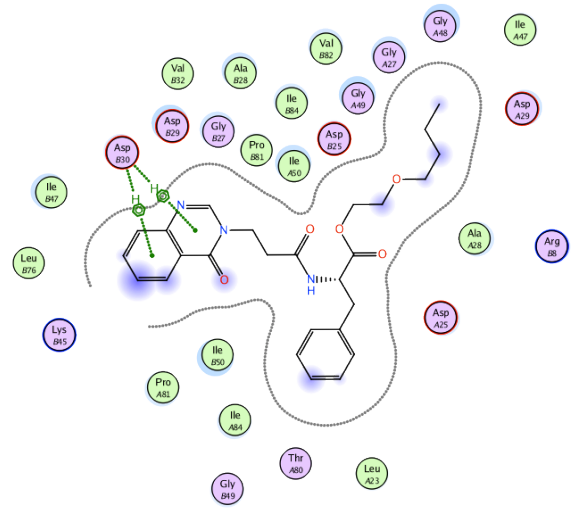
Table 9 and Table 10 show the ten best ranked molecules according to the GOLD-Score. As one can see the molecular weight is in the always over 400 Da. Although the highest score obtained is 81, the ligand interactions do not show as many interactions points as the molecules we selected for synthesis.

Table 9: GOLD-Score, Molecular weight and SlogP for the best ranked molecules

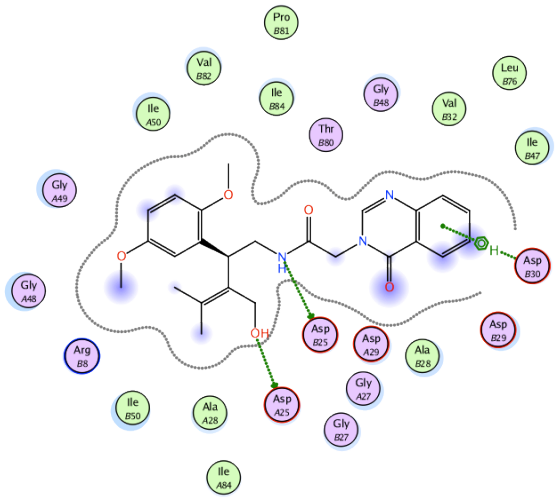
Name	GOLD-Score	MW	SlogP
mol_63_a09	81	462	3.7
mol_185_a01	80	462	3.7
mol_145_a05	80	465	4.3
mol_2_a09	80	427	0.8
mol_197_a02	79	465	3.3
mol_58_a01	78	451	3.0
mol_49_a01	78	467	2.4
mol_222_a04	77	455	4.3
mol_210_a06	76	468	1.2
mol_219_a04	76	457	5.2

Table 10: Ligand interaction charts for the best ranked molecules

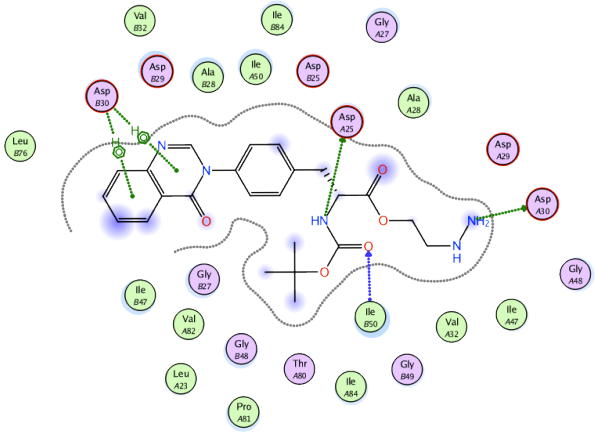
 mol_63_a09	 mol_185_a01
 mol_145_a05	 mol_2_a09



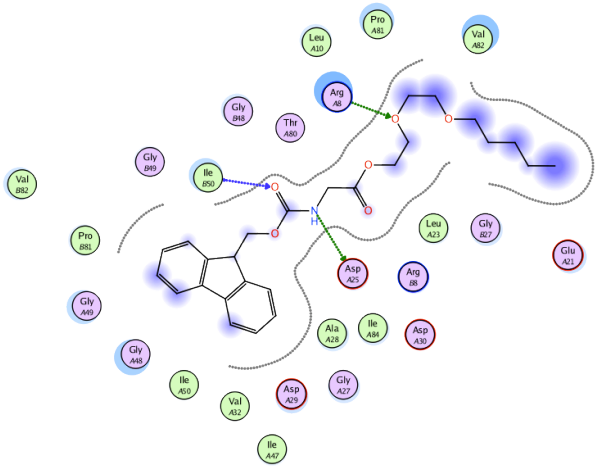
mol_197_a02



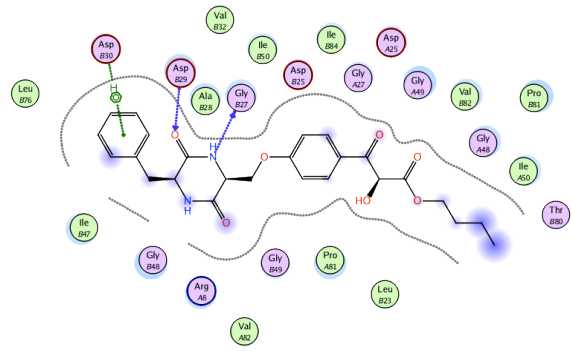
mol_58_a01



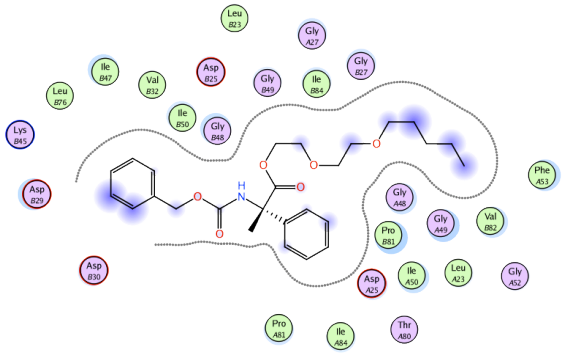
mol_49_a01



mol_222_a04



mol_210_a06



mol_219_a04

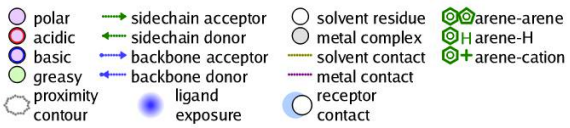


Table 11: GOLD-Scores for the selected structures

Name	GOLD-Score
Alpha 0.8 molecule 217	58.9752
Alpha 0.8 molecule 14	61.8023
Alpha 0.9 molecule 35	62.0555
Alpha 0.9 molecule 5	68.5400
Alpha 0.1 molecule 136	64.6696
Alpha 0.1 molecule 39	60.4809

The ligand interactions for these molecules can be found under the section cherry picking 2.1.8.

2.1.8 Cherry picking

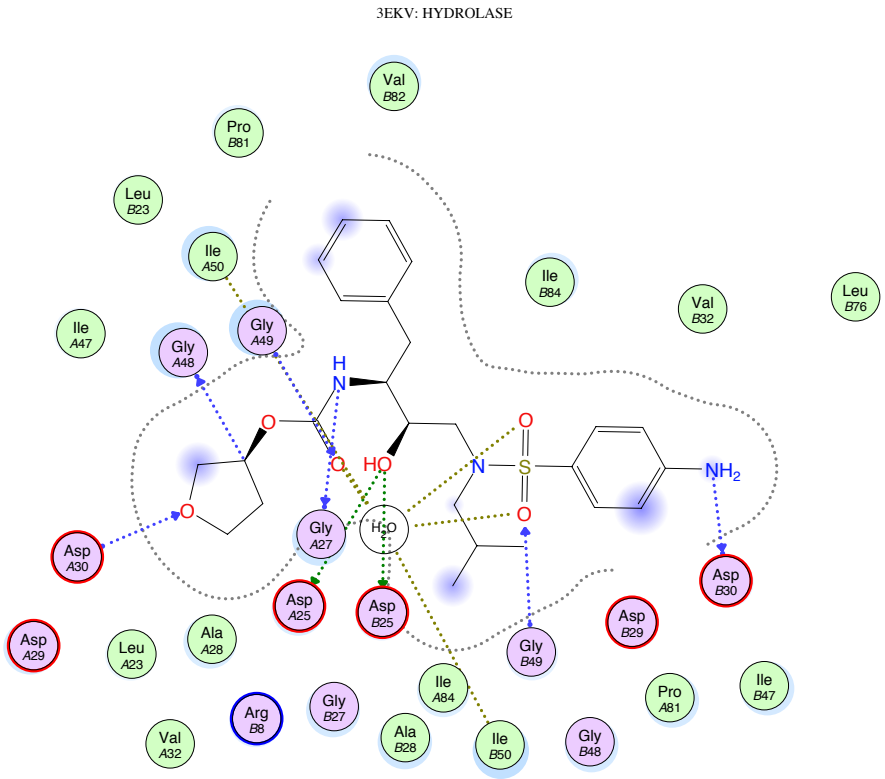


Figure 22: Ligand Interactions of Amprenavir with HIV-protease in the 3EKV structure of the PDB

The main interacting residues are Asp 25/25', which are the catalytic aspartates, Gly 48/48', Asp 30/30' as well as Ile 50/50' over a conserved water molecule. Further interactions can be observed in the pictures of the interactions with Amprenavir and depend on the ligand orientation in the binding cavity. The ligand interactions predicted with MOE depend also on the orientation of the residues in the crystal structure and the protonation state.

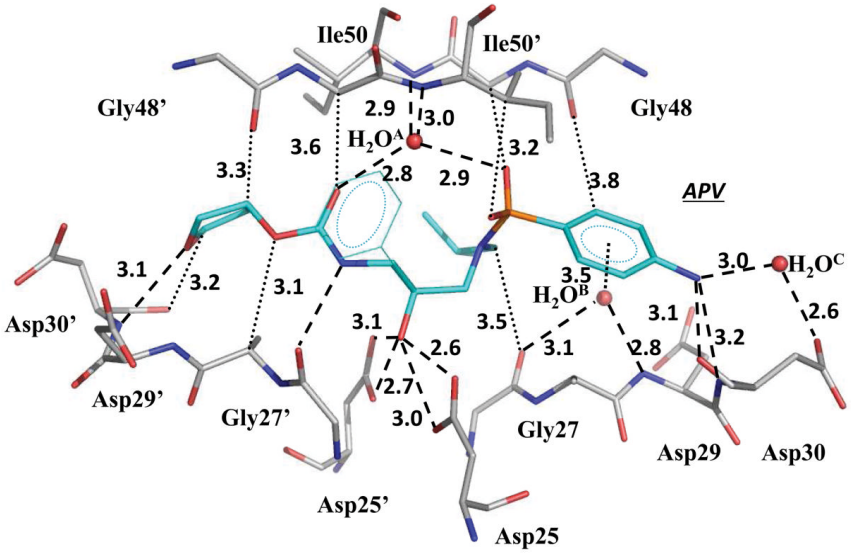


Figure 23: Interactions of APV with HIV-protease PDB: 3NU3 (Shen et al. 2010)
Hydrogen bond, C-H···O and H2O···π interactions between PR (gray) and APV (cyan). Hydrogen bond interactions are indicated by dashed lines. C-H···O and H2O···π interactions are indicated by dotted lines.

The structures were selected after sorting the docking results by their molecular weight and their GOLD score. For the selection of the structures only the highest scored docking pose for each molecule was considered. The docking poses for the molecules were inspected manually in MOE and the pictures for the ligand interactions were generated. Molecules showing some of the same main interactions with the protein as Amprenavir were most promising to be active against the protease in *in vitro* testing. The crucial interactions can be found in Figure 22 and Figure 23.

A second criterion for four of the structures to be selected was that they only showed minor differences in their molecular scaffold. Two of them only differ in one carbon atom in the aromatic sidechain and had the same synthesis suggested by DOGS (mol_217_a08 and mol_14_a08). The other two molecules (mol_5_a09 and mol_35_a09) shared the same molecular scaffold as the larger molecule of the previously described pair (mol_14_a08). Mol_5_a09 has a hydroxyl group on the phenyl ring as the amino acid tyrosine. Mol_35_a09

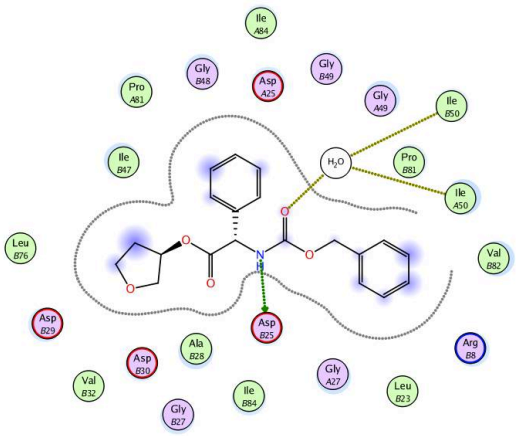
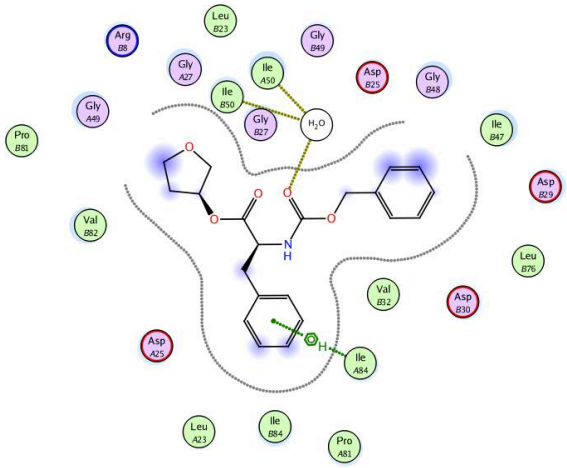
is substituted on the other ring in *para* position to the oxygen forming the phenyl ether. The other ring is unsubstituted corresponding to the amino acid phenylglycine.

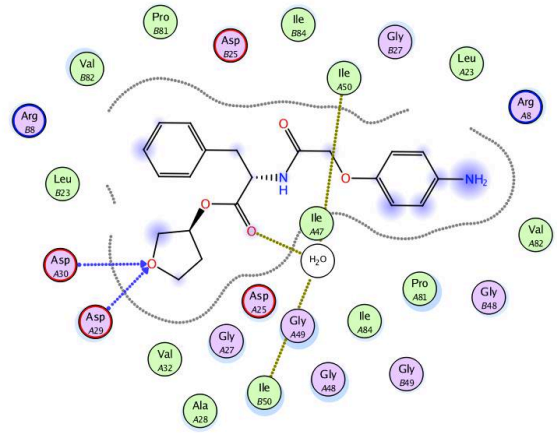
The remaining two structures were selected because of their completely different molecular scaffold compared to classical HIV protease inhibitors. Molecule BZ1 (mol_39_a01) shared a scaffold also appearing in a different project in the group and was therefore interesting for later comparison if it would show some activity. The sixth molecule had a sugar substructure, very unlikely to appear in promising molecules, and therefore caught our interest.

Gisbert Schneider checked the already narrowed subset of compounds to select compounds, which are reasonable as HIV protease inhibitors. Tiago Rodriguez inspected the structures and synthetic suggestions of DOGS with respect to synthetic feasibility in the approximate time period of one and a half months. He helped to find an appropriate synthesis method for each computationally designed compound with the suggested reaction strategy.

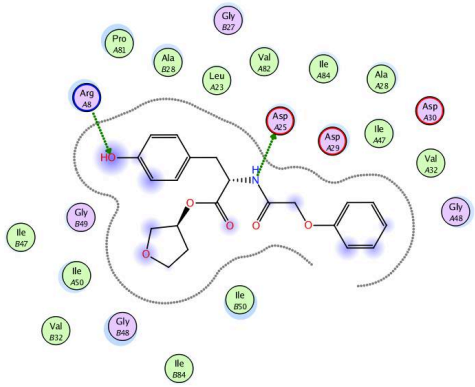
The selected structures and the ligand interactions of their best-ranked docking pose can be found in Table 12. The interaction with ASP 25/25' in the active site is an important criterion for the inhibition. Also a very important interaction is the one to ILE 50/50' in the flaps which coordinates over a conserved water molecule. To get a better prediction of the interaction mode of the structures a more exhaustive docking experiment would be necessary, where the residues in the protein are also optimized according to the hydrogen bond network.

Table 12: Ligand interactions of the molecules selected for synthesis

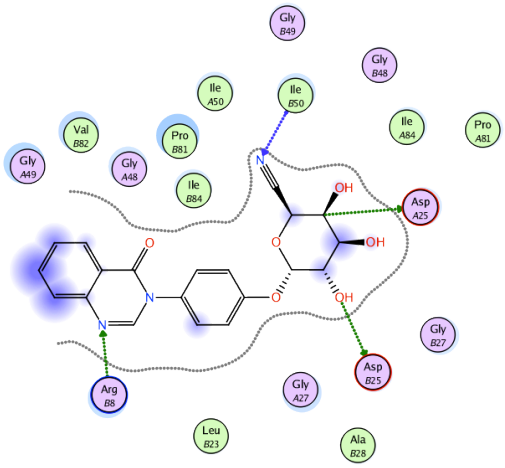
 mol_217_a08 = BZ-2	 mol_14_a08 = BZ-3



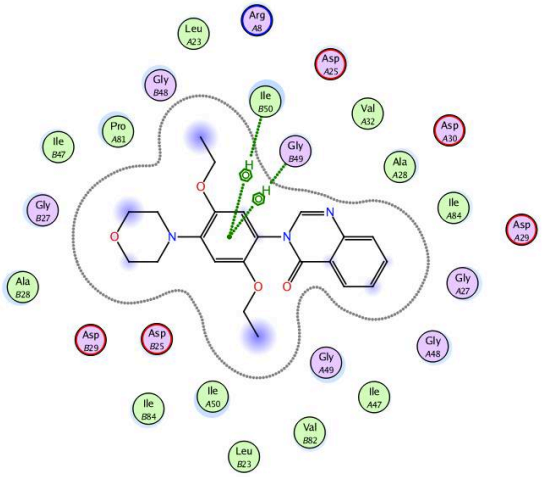
mol_35_a09 = BZ-5



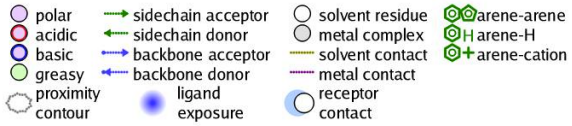
mol_5_a09 = BZ-4



mol_136_a01



mol_39_a01 = BZ-1



2.2 Synthesis

2.2.1 Suggestion by DOGS

In this section, the molecules and the corresponding synthesis pathways suggested by DOGS are listed. For the molecules of interest the exact output of DOGS is shown. The suggestion is described in a text document containing the code of the molecule in its title and the start fragment with the DOGS distance as well as the added building blocks, the coupling reaction, and the fitness value of the final product. For better understanding, the start fragments and the added building blocks are depicted in section 2.2.2.

Alpha 0.8 molecule 217

DB-entry_12705: 73113_FLUKA (Fitness: 0.4312523603439331)

DB-entry_12705: 73113_FLUKA + DB-entry_11399: 296686_ALDRICH --> Mitsunobu Carbonsäure (Ladung 2) --> construct_1 (Fitness: 0.3126789331436157)

Alpha 0.8 molcule 14

DB-entry_12636: 359807_ALDRICH (Fitness: 0.3735860586166382)

DB-entry_12636: 359807_ALDRICH + DB-entry_11399: 296686_ALDRICH --> Mitsunobu Carbonsäure (Ladung 2) --> construct_1 (Fitness: 0.265169620513916)

Alpha 0.9 molecule 35

DB-entry_21790: C1378_SIGMA (Fitness: 0.5497753620147705)

DB-entry_21790: C1378_SIGMA + DB-entry_3395: 09160_FLUKA --> Williamson Ether --> construct_1 (Fitness: 0.4038090705871582)

construct_1 + DB-entry_11399: 296686_ALDRICH --> Mitsunobu Carbonsäure (Ladung 2) --> construct_2 (Fitness: 0.30763572454452515)

Alpha 0.9 molecule 5

DB-entry_21792: C1878_SIGMA (Fitness: 0.5727962851524353)

DB-entry_21792: C1878_SIGMA + DB-entry_3071: 16016_SIGMA-ALDRICH --> Williamson Ether --> construct_1 (Fitness: 0.4421009421348572)

construct_1 + DB-entry_11399: 296686_ALDRICH --> Mitsunobu Carbonsäure (Ladung 2) --> construct_2 (Fitness: 0.34216201305389404)

Alpha 0.1 molecule 136

DB-entry_11025: H5405_SIGMA (Fitness: 0.2588313817977905)

DB-entry_11025: H5405_SIGMA + DB-entry_15732: 00483_ALDRICH --> Michael-Addition --> construct_1 (Fitness: 0.07521235942840576)

construct_1 + DB-entry_5215: 121290_SAJ --> Mitsunobu Carbonsäure (Ladung 2) --> construct_2 (Fitness: 0.07135742902755737)

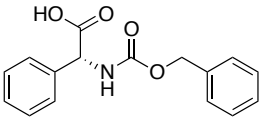
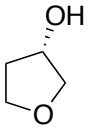
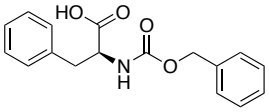
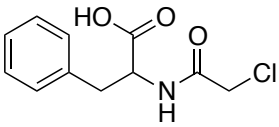
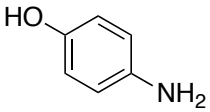
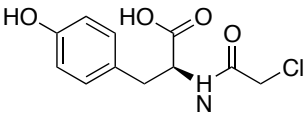
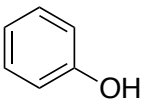
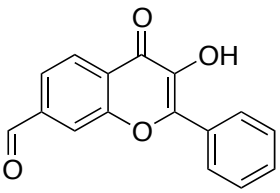
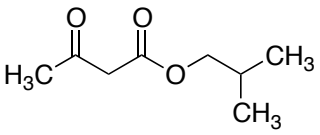
Alpha 0.1 molecule 39

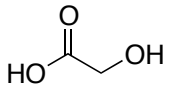
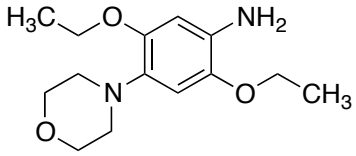
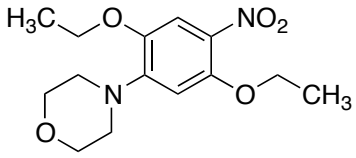
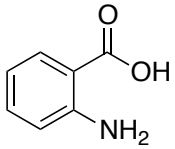
DB-entry_11296: 439363_ALDRICH (Fitness: 0.2861173748970032)

DB-entry_11296: 439363_ALDRICH + DB-entry_8422: 10678_FLUKA --> Quinazolin (Ladung 2) --> construct_1 (Fitness: 0.11678570508956909)

2.2.2 Building blocks

Table 13: Building blocks ordered for synthesis

Structure	Product number	Amount	Vendor	Note
	73113_FLUKA	1g	Fluka	
	296686_ALDRICH	5g	Aldrich	Necessary for 4 reactions Racemate ordered
	359807_ALDRICH	5g	Acros	
	C1378_SIGMA	1g	TCI	
	09160_FLUKA	5g	Acros	Nitrophenol ordered instead of amino
	C1878_SIGMA	1g	TCI	
	16016_SIGMA-ALDRICH	500mg	Fluka	
	H5405_SIGMA			Not ordered
	00483_ALDRICH			Not ordered

	121290_SAJ			Not ordered
				Amino- not available
	439363_ALDRICH	10g	Aldrich	Nitro- ordered
	10678_FLUKA	50g	Sigma	
Additional chemicals				
Sn granula		50g	Sigma	
$Yb(OTf)_3$		1g	Aldrich	
Diethyl 1,2-hydrazinedicarboxylate		25g	Aldrich	
Iodobenzenediacetate		5g	Aldrich	
Triphenylphosphine		100g	ABCR	
Ethanol for synthesis				

The reagents and solvents were ordered from ABCR Chemicals, Acros, Aldrich, Fluka and Sigma and directly used without further purification.

2.2.3 Introduction to the synthetic strategies

DOGS only provides the information shown in 2.2.1. In order to synthesise the compounds, information about reaction conditions has to be obtained from the literature.

All the synthesis methods were discussed with Tiago Rodriguez and Gisbert Schneider whether they were chemically feasible and in the range of the budget for the present project.

Only three of the six selected molecules were finally synthesized (BZ-1 to BZ-3). For five molecules (BZ-1 to BZ-5) the building blocks were ordered. For the sixth molecule, the suggested reaction was not reasonable because it required several different protection groups and the reactivity of the compounds under the given conditions was questionable.

2.2.4 Characterization and Analysis of the synthesized compounds

Melting point:

All melting points were measured with Melting Point M-560 (Büchi, Switzerland). At the starting point the temperature was set to 70°C and was then raised with a gradient of 10°C/min to the maximum temperature of 250°C.

HPLC:

The products and intermediates were analysed and purified on an RP-HPLC (reverse phase high pressure/performance liquid chromatography) (Shimadzu, Columbia, USA). For that purpose two analytical columns and one preparative column were used (MACHEREY-NAGEL GMBH & CO. KG, Düren, Germany). The properties of the different columns are subsumed in Table 14.

Analytical HPLC-MS was carried out in a Shimadzu LC-MS2020 system, equipped with a Nucleodur C18 HTec column, under an appropriate gradient of ACN : H2O (+ 0.1% TFA in each phase), and a total flow rate of 0.5 mL/min. The mass spectrometer was operated in positive-ion mode with ESI. Preparative HPLC was carried out on a Shimadzu LC-8A system, coupled to a Nucleodur 100-5 C18 HTec column, and a SPD-20A UV/Vis detector at 210 nm detection wavelength.

The masses from 80 to 700 Da were recorded with a standard gradient from 30 to 95% ACN/H_2O for the S2 column and different gradients for the S1 column used to check the

appropriate conditions for the preparative runs. For the S1 column three different linear gradients were used: 50-75%, 50-95% and 30-95% ACN/H_2O (always 0.1% TFA).

Table 14: Properties of HPLC columns

	Column S1	Column S2	Column prep
Use	Analytical	Analytical	Preparative
Length	150mm	100mm	150mm
Internal diameter	3.0mm	3.0mm	21.0mm
Pore size	110 Å	110 Å	110 Å
Particle size	5µm	1.8µm	5µm
Packing	C18 Htec	C18 Htec	C18 Htec
Run time	16min	12min	16min
Flow rate	0.5mL/min	0.5mL/min	24.5mL/min

NMR:

The final products were characterized by ^{1}H-NMR (^{1}Proton Nuclear Magnetic Resonance), ^{13}C-NMR (^{13}Carbon Nuclear Magnetic Resonance). ^{1}Proton and carbon nuclear magnetic resonance spectra (^{1}H and ^{13}C NMR, respectively) were recorded on Bruker Avance 400 spectrometer. Chemical shifts (δ) are reported in units of parts per million (ppm) downfield from SiMe4 (δ 0.0) and relative to the respective solvent's peak. Multiplicities are given as: s (singlet), d (doublet), t (triplet), q (quartet), dd (double of doublet) td (triplet of doublet) or m (multiplet). 1H-1H Coupling constants (*J*) are reported in Hertz (Hz).

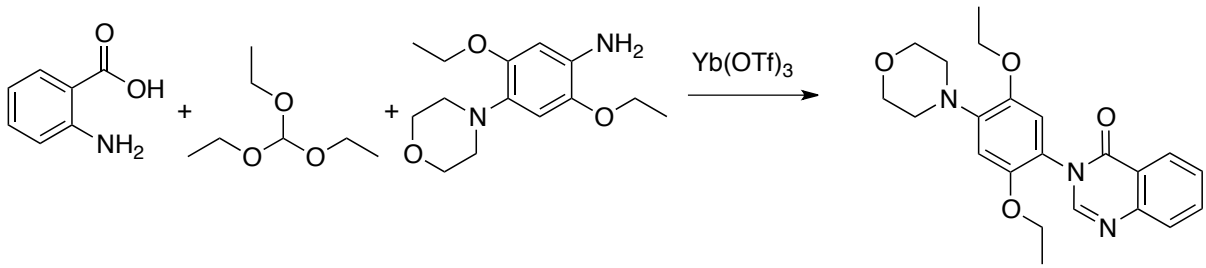
MS:

The mass spectra for the three synthesized compounds were recorded with Burker Daltonics maXis ESI-QTOF in the group of Dr. Xiangyang Zhang.

2.2.4.1 Preparation of 3-(2,5-Diethoxy-4-morpholinophenyl)quinazolin-4(3H)-one (BZ-1)

The molecule with the DOGS code **alpha 0.1 molecule 39** later referred to as **BZ-1** was prepared according to Limin Wang et al. (L. Wang et al. 2003)

Reaction scheme:



Chemicals:

Anthranilic acid	54.856 mg	0.4 mmol	1 eq
Triethoxymethane	86µl, 77.064 mg	0.52 mmol	1.3 eq
2,5-diethoxy-4-morpholinoaniline	127.5 mg	0.479 mmol	1.2 eq
Ytterbium triflate	3.22 mg	0.0052 mmol	0.013 eq
Ethanol as solvent	2 mL		

Preparation of BZ-1:

Anthranilic acid (0.4 mmol, 54.8 mg), triethyl orthoformate (0.52 mmol, 77.1 mg; 86 µL) 2,5-diethoxy-4-morpholinoaniline (0.48 mmol, 127.5 mg) and $Yb(OTf)_3$ (0.0052 mmol, 3.22 mg) in 2 mL ethanol were irradiated in a microwave vial under stirring at 100°C for 30 min at 200W.

Work-up:

After the termination of the reaction H_2O (2 mL) was added to the product. Then, the product was extracted with EtOAc (3x 10 mL). The organic layer was dried over $MgSO_4$ and the solvent was removed on the rotavap. The dark violet residue was dissolved in H_2O/ACN 1:1 (5 mL) and purified by preparative HPLC. The compound was separated over a linear gradient of 35-95% ACN/H_2O (0.1% TFA).

After removing H_2O and ACN the light violet crystals were dried in the vacuum exsiccator.

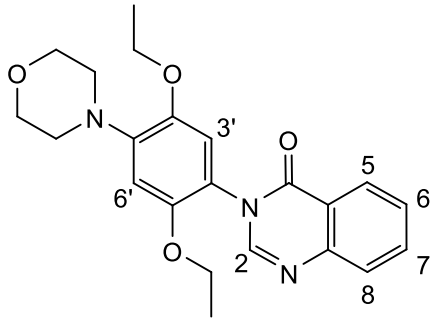
Outcome:

A white to very clear pink coloured solid was obtained. In solution the violet colour was still present. The absolute yield after purification was 37.3 mg, which is 10% of the theoretical yield.

Analysis:

The melting interval was found to be between 157.8-160.1°C. The expected product could be proven with HR-MS ESI $[M+H]^+$ ($C_{22}H_{26}N_3O_4^+$) (expected mass: 396.1918 determined mass: 396.1918).

3-(2,5-Diethoxy-4-morpholinophenyl)quinazolin-4(3H)-one (BZ-1):



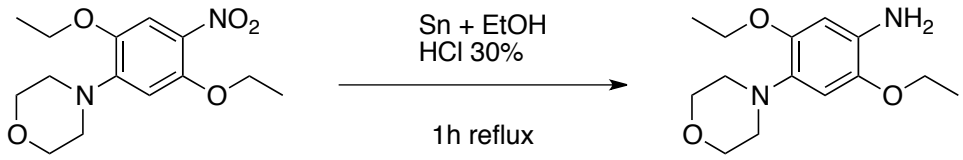
^{1}H NMR (400 MHz, $CDCl_3$): δ = 1.28 (t, *J* = 7.0 Hz, 3 H), 1.50 (t, *J* = 7.0 Hz, 3 H), 3.55 (t, *J* = 4.0 Hz, 4 H), 4.05 (t, *J* = 4.6 Hz, 4 H), 4.13 (q, *J* = 7.1 Hz, 4 H), 7.01 (s, 1 H), 7.13 (s, 1 H), 7.66 (t, *J* = 7.2 Hz, 1 H), 7.88–7.93 (m, 2 H), 8.40 (d, *J* = 7.6 Hz, 1 H), 8.45 ppm (s, 1 H)

^{13}C NMR (100 MHz, $CDCl_3$): δ = 14.48 and 14.74, 51.44, 65.22 and 65.40, 107.10, 114.38, 121.69 and 122.13, 124.90, 127.73, 128.66, 135.61, 137.17, 143.90, 145.44, 148.09, 148.84, 159.61 ppm

2.2.4.2 Preparation of 2,5-diethoxy-4-morpholino-aniline for preparation of BZ-1

Due to an ordering mistake 4-(2,5-diethoxy-4-nitrophenyl)-morpholine was ordered which had to be reduced to the 2,5-diethoxy-4-morpholino-aniline. The nitro-group was reduced by the protocol shown in the scheme:

Reaction scheme:



Chemicals:

4-(2,5-diethoxy-4-nitrophenyl)-morpholine	500 mg	1.687 mmol	1 eq
Sn granula	2.83 g	24 mmol	14.2 eq
Ethanol	4.16 mL		
HCl (30%)	6.13 mL	28 mmol	16.6 eq

Preparation of 2,5-diethoxy-4-morpholino-aniline:

4-(2,5-diethoxy-4-nitrophenyl)-morpholine (500 mg, 1.687 mmol) were mixed with 4.16 mL of Ethanol. 2.83 g of Sn granula and 6.13 mL of HCl (30%) were added. The mixture was stirred under reflux for 60 minutes. The formation of the product was checked with HPLC.

Work up:

The mixture was diluted with 25 mL H_2O. Then 100 mL of dichloromethane were added and basified to pH 10 with NaOH. Then the product was extracted four times with 50 mL dichloromethane and the organic phase was dried with $MgSO_4$. Finally the product was obtained by evaporating the organic solvent under reduced pressure.

Outcome:

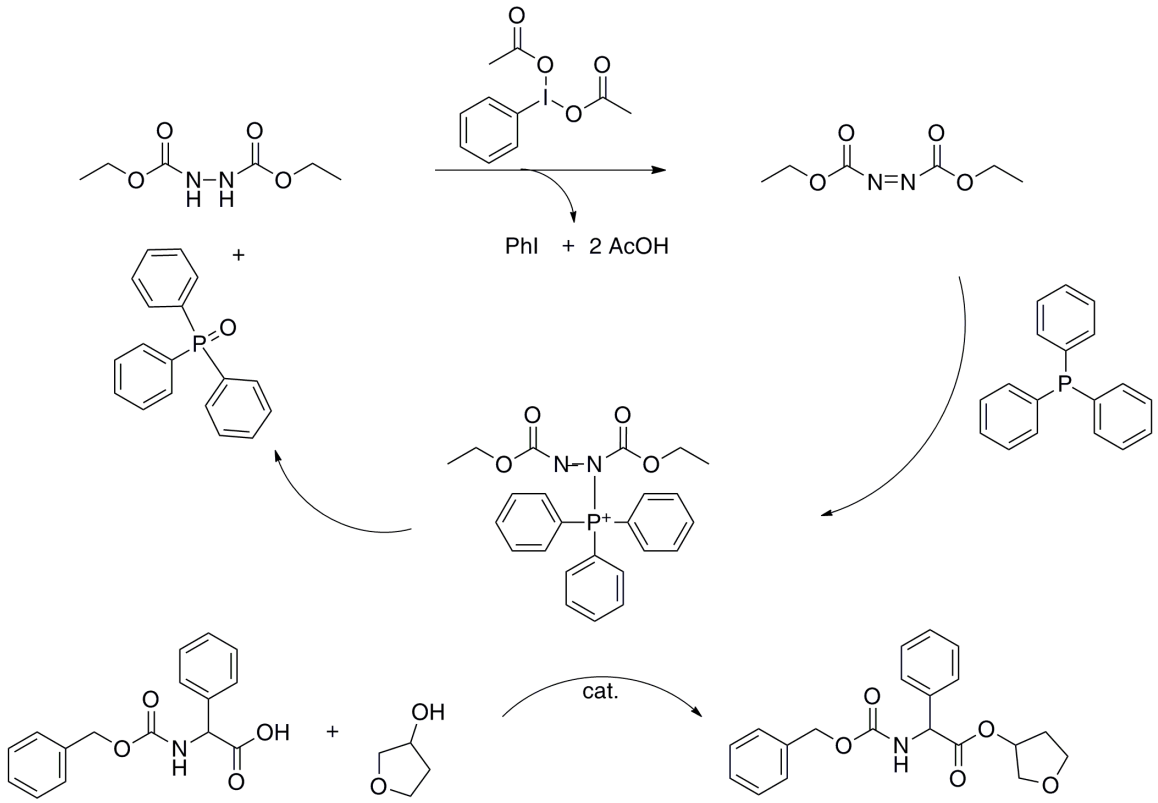
A violet solid was obtained which cleared out a little bit while drying in the vacuum exsiccator at high vacuum for two days.

The absolute yield was 271.5 mg, which corresponds to a theoretical yield of 60%.

2.2.4.3 Preparation of Tetrahydrofuran-3-yl 2[{(Benzyloxy)carbonyl}amino]-2-phenylacetate (BZ-2)

The molecule corresponding to the DOGS code **alpha 0.8 molecule 217** later referred to as **BZ-2** was prepared according to But and Toy. (But & Toy 2007)

Reaction scheme:



Chemicals:

Z-D-phenylglycine		285.29 mg	1 mmol	2.2 eq
(S)-(+)-3-Hydroxytetrahydrofurane	72.8 µL	79.3 mg	0.9 mmol	2 eq
Diethylazodicarboxylate	14 µL	15.6 mg	0.09 mmol	0.2 eq
Triphenylphosphine		119.08 mg	0.454 mmol	1 eq
Iodobenzene diacetate		146.23 mg	0.454 mmol	1 eq
Tetrahydrofurane as solvent	10mL			

Preparation of BZ-2:

3-Hydroxytetrafurane (0.9 mmol, 79.3 mg, 72.8 µL) and Z-D-Phenylglycine (1 mmol, 285.29 mg) were primed in anhydrous THF (10 mL). Next, triphenylphosphine (0.454 mmol, 119.08 mg), diethyl azodicarboxylate (0.09 mmol, 15.6 mg, 14 µL) and iodobenzene diacetate (0.454 mmol, 146.23 mg) were added.

The reaction was set up in a microwave vial. The probe was irradiated with 120 W for 30 minutes at 50°C.

Work up:

The organic phase with the product was washed with $NaHCO_3$ (2x 20 mL) and brine (20 mL) and dried over $MgSO_4$. The solvent was evaporated under reduced pressure. A yellowish oil was obtained. It was dissolved in H_2O/ACN 1:2 mL (3 mL) and purified in the preparative HPLC. The compound was separated over a linear gradient of 50-75% ACN/H_2O (0.1% TFA).

Outcome:

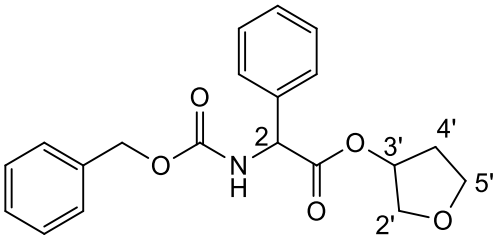
After the purification H_2O and ACN were removed and light yellowish crystals were obtained after cooling the product in the fridge (-20°C, 2 days). The crystals were dried in the vacuum exsiccator for two days.

The absolute yield after purification was 24.8 mg, which is 8% from the theoretical yield.

Analysis:

The melting interval was found to be between 92.1-93.1°C. The expected product could be proven with HR-MS ESI $[M+Na]^+$ ($C_{20}H_{21}NO_5$ + Na^+) (expected mass: 378.1312 determined mass: 378.1308).

Tetrahydrofuran-3-yl 2-[{(Benzyloxy)carbonyl}amino]-2-phenylacetate (BZ-2):



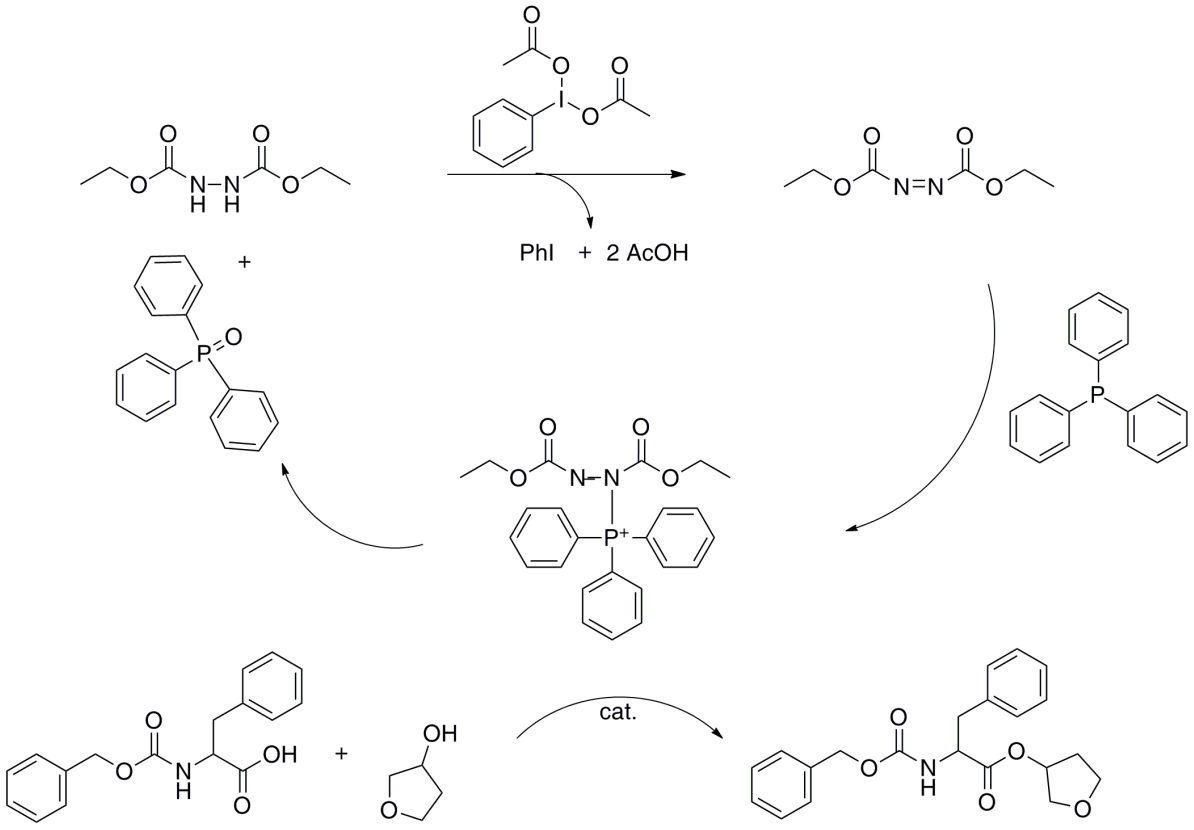
^{1}H NMR (400 MHz, $CDCl_3$): δ = 1.75 – 1.78 (m, 1H), 2.03 – 2.08 (m, 2H), 2.17 – 2.20 (m, 1H), 3.58 – 3.61 (m, 1H), 3.69 – 3.73 (m, 1H), 3.76 – 3.92 (m, 6H), 5.09 – 5.17 (m, 4H), 5.37–5.38 (m, 3H), 5.76 – 5.78 (m, 1H), 7.38 ppm (m, 10H)

^{13}C NMR (100 MHz, $CDCl_3$): δ = 32.79, 58.06, 66.84, 67.19, 72.72, 76.35, 127.04, 128.21, 128.55, 129.04 ppm, signals of COO and NCO hidden by noise.

2.2.4.4 Preparation of Tetrahydrofuran-3-yl-2-[{(benzyloxy)carbonyl}amino]-3-phenyl-propanoate (BZ-3)

The molecule corresponding to the DOGS code **alpha 0.8 molecule 14** later referred to, as BZ-3 was also prepared according to But and Toy. (But & Toy 2007)

Reaction scheme:



Chemicals:

N-(Carbobenzyloxy)-L-phenylalanine		299.32 mg	1 mmol	2.2 eq
(S)-(+)-3-Hydroxytetrahydrofurane	72.8 µL	79.3 mg	0.9 mmol	2 eq
Diethylazodicarboxylate	14 µL	15.6 mg	0.09 mmol	0.2 eq
Triphenylphosphine		119.08 mg	0.454 mmol	1 eq
Iodobenzene diacetate		146.23 mg	0.454 mmol	1 eq
Tetrahydrofurane as solvent	10 mL			

Preparation of BZ-3:

Triphenylphosphine (0.454 mmol, 119.08 mg), diethyl azodicarboxylate (0.09 mmol, 15.6 mg, 14 µL) and iodobenzene diacetate (0.454 mmol, 146.23 mg) were added to 3-Hydroxytetrafurane (0.9 mmol, 79.3 mg, 72.8 µL) and (S)-2-(Benzyloxycarbonylamino)-3-phenyl propanoic acid (1 mmol, 300 mg) in anhydrous THF (10 mL).

The reaction was set up in a microwave vial. The probe was irradiated with 120 W for 30 minutes at 50°C.

Work up:

The organic phase with the product was washed with $NaHCO_3$ (2x 20 mL) and brine (20 mL) and dried over $MgSO_4$. The solvent was evaporated under reduced pressure. A yellowish oil was obtained which was dissolved in H_2O/ACN 1:1.5 mL (3 mL) and purified by preparative HPLC. The compound was separated over a linear gradient of 50-75% ACN/H_2O (0.1% TFA).

After the purification H_2O and ACN were removed and the remaining oil was dried in the vacuum exsiccator for two days.

Outcome:

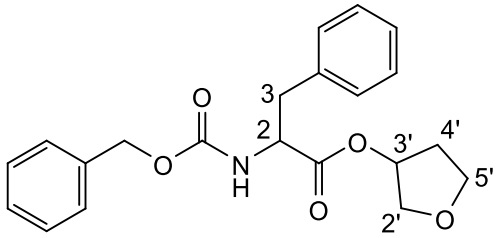
The absolute yield after purification was 18.8 mg which is 5.3% from the theoretical yield.

As product a yellowish oil was obtained which could not be crystalized, therefore no melting point was obtained.

Analysis:

The expected product could be proven with HR-MS ESI $[M+Na]^+$ ($C_{21}H_{23}NO_5 + Na^+$) (expected mass: 392.1468 determined mass: 392.1463).

Tetrahydrofuran-3-yl 2-[{(benzyloxy)carbonyl}amino]-3-phenylpropanoate (BZ3):



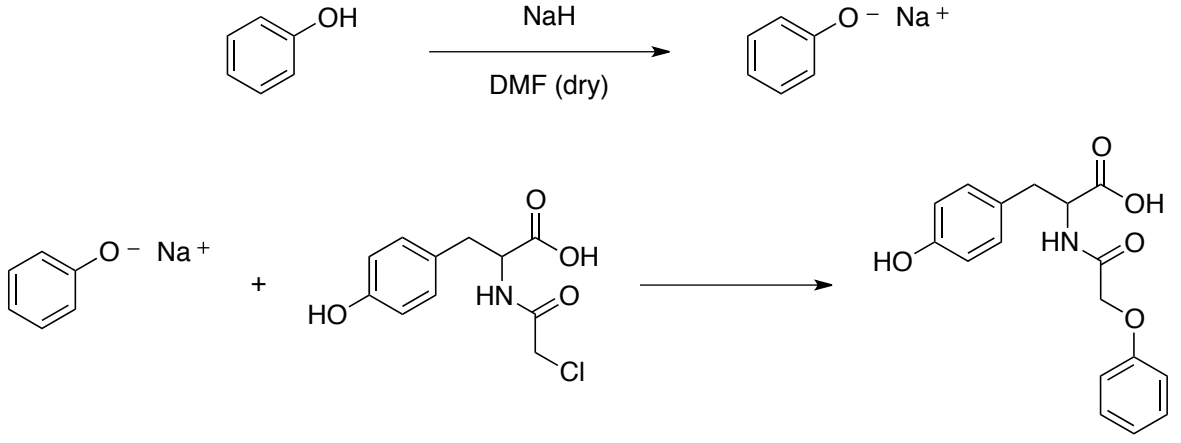
^{1}H NMR (400 MHz, $CDCl_3$) 3.11 (d, *J* = 6.0 Hz, 2 H), 3.66–3.88 (m, 5H), 4.64 (dd, *J* = 6.4 and 14 Hz, 1 H), 5.08 (s, 2 H), 5.14–5.28 (m, 2 H), 7.13–7.29 ppm (m, 10 H)

^{13}C NMR (100 MHz, $CDCl_3$): major isomer: δ = 32.65, 38.21, 54.79, 66.89, 72.82, 75.94, 127.21, 128.11, 128.53 128.62, 129.31, 206.92 ppm, signal of NCO hidden by noise; minor isomer: δ = 32.59, 38.31, 54.85, 67.01, 72.68, 76.09 ppm

2.2.4.5 Preparation of Tetrahydrofuran-3-yl 3-(4-hydroxyphenyl)-2-(2-phenoxyacetamido)propanoate (BZ-4)

The molecule corresponding to the DOGS code **alpha 0.9 molecule 5** later referred to, as BZ-4 was also prepared according to McBurney and co-workers. (McBurney et al. 2004)

Reaction scheme:



Chemicals:

Phenol		43.8 mg	0.4657 mmol	1 eq
NaH		12.1 mg	0.5045 mmol	1.1 eq
N-chloracetyl-L-tyrosine		110.0 mg	0.427 mmol	0.9 eq
DMF (dry) as solvent	2 mL			
Na_2CO_3 (alternative to NaH)		53.47 mg	0.5045 mmol	1.1 eq

Preparation of BZ-4 first protocol:

Phenol (43.8 mg, 0.4657 mmol), and sodium hydride (12.1 mg, 0.5045 mmol) were added to 2mL of dry DMF as solvent. The mixture was stirred for one hour at room temperature. Then N-chloracetly-L-tyrosine (110.0 mg, 0.427 mmol) was added and the mixture was stirred over night (12 h). The progress was monitored with HPLC.

Work up:

The mixture was diluted with H_2O (5 mL) and subsequently washed tree times with ethylacetate (20 mL). The organic solvent was evaporated under reduced pressure after being dried over $MgSO_4$.

Outcome:

There was no product formed and the reaction conditions were changed.

Preparation of BZ-4 second protocol:

The same amounts and educts were used but NaH was changed to Na_2CO_3 (53.47 mg, 0.5045 mmol). The reaction was heated to 70°C. The work up stayed the same.

Outcome:

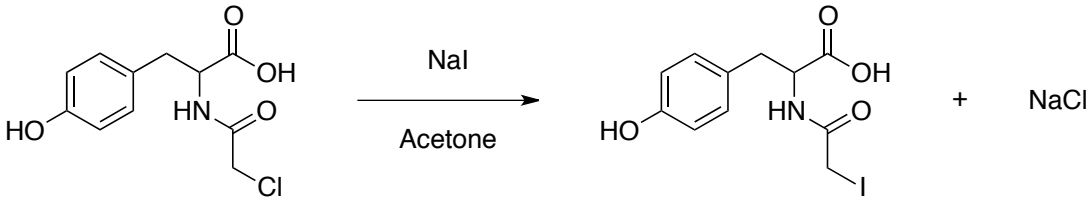
Unfortunately also with the changed conditions no product was formed.

Preparation of BZ-4 third protocol:

For the third trial to synthesize BZ-4 instead of N-chloracetyl-L-tyrosine we intended to use N-iodacetyl-L-tyrosine. The desired product was synthesized in our laboratory. Unfortunately the yield was very low and it was decided that the compound BZ-4 was not pursued any further.

2.2.4.6 Preparation of N-iodacetyl-L-tyrosine for the preparation of BZ-4

Reaction scheme:



Chemicals:

N-chloracetyl-L-tyrosine		80 mg	0.310 mmol	1 eq
NaI		55.8 mg	0.3725 mmol	1.2 eq
Acetone as solvent	1 mL			

Preparation of N-iodacetyl-L-tyrosine:

80 mg of N-chloracetyl-L-tyrosine (0.310 mmol) and 55.8 mg of NaI (0.3725 mmol) were dissolved in 1 mL of acetone. The mixture was stirred for 1.5 days at room temperature.

Work up:

The mixture was diluted with H_2O (5 mL) and extracted with dichloromethane (3x 10 mL). The organic phase was dried over $MgSO_4$ and the organic solvent was evaporated under reduced pressure.

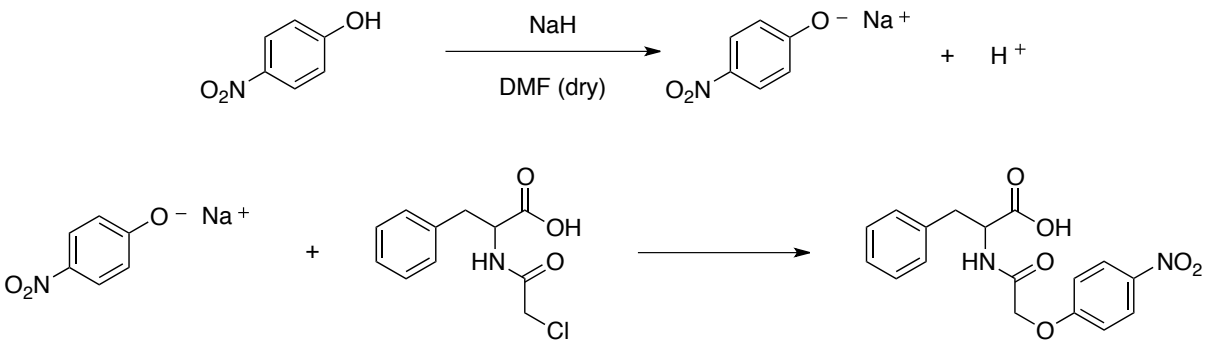
Outcome:

The residue in the flask was weighed giving an absolute yield of 8.4 mg of N-iodacetyl-L-tyrosine which corresponds to 2.4% of the theoretical yield.

2.2.4.7 Preparation of Tetrahydrofuran-3-yl 2-(2-(4-aminophenoxy)acetamido)-3-phenylpropanoate (BZ-5)

The molecule corresponding to the DOGS code **alpha 0.9 molecule 35** later referred to, as BZ-5 was also prepared according to McBurney and co-workers. (McBurney et al. 2004)

Reaction scheme:



Chemicals:

Nitrophenol		69.08 mg	0.496 mmol	1 eq
NaH		12.9 mg	0.537 mmol	1.1 eq
N-Chloracetyl-DL-Phenylalanine		110.0 mg	0.455 mmol	0.9 eq
DMF (dry) as solvent	2 mL			
Cs_2CO_3 (alternative to NaH)		118.63 mg	0.3641 mmol*	0.7 eq

*see second protocol amounts adapted

Preparation of BZ-5 first protocol:

Nitrophenol (69.08 mg, 0.496 mmol), and sodium hydride (12.9 mg, 0.537 mmol) were added to 2 mL of dry DMF as solvent. The mixture was stirred for one hour at room temperature. Then N-Chloracetyl-DL-Phenylalanine (110.0 mg, 0.455 mmol) was added and the mixture was stirred until the reaction was completed. The progress was monitored with HPLC.

Work up:

The mixture was diluted with H_2O (5 mL) and subsequently washed three times with ethylacetate (10 mL). The organic solvent was evaporated under reduced pressure after being dried over $MgSO_4$.

Outcome:

After the mixture was stirred over night the reaction was stopped. There was no product formed and the reaction conditions were changed.

Preparation of BZ-5 second protocol:

NaH was changed to Cs_2CO_3 (118.63 mg, 0.3641 mmol) and the reaction was carried out only with 75 mg of N-Chloracetyl-DL-Phenylalanine (0.310 mmol). All the other amounts were adapted.

Preparation of BZ-5 third protocol:

Instead of N-Chloracetyl-DL-Phenylalanine N-Iodacetyl-DL-Phenylalanine was used. The amounts were adjusted to the amount of N-Iodacetyl-DL-Phenylalanine prepared by Tiago Rodriguez.

Nitrophenol		14.46 mg	0.103 mmol	
NaOH (8%)	2mL		0.4 mmol	
N-Iodacetyl-DL-Phenylalanine		38.25 mg	0.115 mmol	
Acetone as solvent	2 mL			

2 mL of NaOH (8%, 0.4 mmol), Nitrophenol (14.46 mg, 0.103 mmol) and acetone (2 mL) were mixed and heated under reflux for 90 minutes. Then the N-Iodacetyl-DL-Phenylalanine (38.25 mg, 0.115 mmol) was added and the mixture was transferred to the microwave. It was heated to 50°C and irradiated to with 250 W for 10 minutes.

Outcome:

Also under these conditions no product was formed. Due to the narrow time budget, it was decided not to pursue synthesis of the desired product any further.

2.2.5 Discussion of the reactions

2.2.5.1 BZ-1

The reaction Wang is an Ytterbium triflate catalysed one pot reaction that was carried out with anthranilic acid, triethoxymethane and the 2,5-diethoxy-4-morpholino-aniline building block. The synthesis method describes that reaction as solvent-free, but in the present case a solvent was needed because none of the building blocks were liquid. The reaction was adapted to the laboratory microwave.

As marked in the list of building blocks (Table 13) the 2,5-diethoxy-4-morpholinoaniline compound was not obtained as ordered. Instead, the 4-(2,5-diethoxy-4-nitrophenyl)morpholine was ordered. In order to start with the correct educt, the nitro group was first reduced as described in the protocol 2.2.4.2.

The work up was carried out as described in the original protocol (L. Wang et al. 2003). For the gradients on the preparative HPLC the solubility of the compound in ACN/H_2O was checked and the separation was carried out as described in section 2.2.4.1. After the purification it could be proved with HPLC and NMR that the desired final product was obtained.

2.2.5.2 BZ-2 and BZ-3

BZ-2 and BZ-3 were synthesized under the same conditions because they only differed in one C-atom in one of the building blocks (Z-D-Phenylglycine for BZ-2 and N-(Carbobenzyloxy)-L-phenylalanine for BZ-3). This reaction was first successful under reflux using a modified protocol by But & Toy for organic Mitsunobu coupling reactions (Protocol A). The reaction was completed, but the educts were not very reactive under the given conditions so the protocol was adapted for microwave synthesis, which was successful. The yield could be increased and the reaction time was significantly shorter. For both reactions and both products the same work up was applied. The products obtained in the reflux reaction were analysed with HPLC in our laboratory and in the mass spectrometer coupled with the HPLC. For the biological testing in section 3 at first only the products obtained in the microwave were tested. For the last protocol also the compounds obtained under reflux were analysed. These compounds were purified under the same conditions and titled with BZ-2A and BZ-3A.

2.2.5.3 BZ-4 and BZ-5

For the molecules BZ-4 and BZ-5 (mol_5_a09 and mol_35_a09) the building blocks according to the suggestion by DOGS were ordered.

The synthesis of BZ-4 and BZ-5 was tried several times but the first intermediate could not be formed. Generally BZ-4 and BZ-5 only differed a little in the building blocks but the reaction was to be performed in the same way. The diverse protocols that were used are listed above. Generally they differ in the reducing agent. In the second modification the nucleophiles were changed from chlorides to iodides to obtain better leaving groups in the reactions. That proved unsuccessful either and still the intermediate was not formed. For time reasons we stopped pursuing these products and focused on the other compounds.

These two molecules were good examples for the limits of the *de novo* design software DOGS with respect to the suggested syntheses.

The reactions would probably require protections groups for the carbonic acids but the building blocks suggested by DOGS were not varied according to these synthetic needs. Therefore it was rather clear that the reactions for BZ-4 and BZ-5 could not work without being further investigated and adapted. Probably a different synthesis route could yield the desired products in an easier way.

2.2.5.4 Molecule mol_136_a01

The last molecule mol_136_a01 that was picked from the computational experiments has a completely different structure to all the known protease inhibitors and also to the other molecules chosen in our selection. It contains a sugar part and a 4a,8a-dihydroquinazolin-4(3H)-one structure appearing very often in the molecules that DOGS has suggested. But due to the sugar part of the molecule one would have to face several problems in the synthesis. Each of the alcohol groups could react in the coupling of the building blocks and very specific protection groups would be needed. But before finding a possible synthesis route the availability of the building blocks was checked. This specific binding block was no longer available in the Sigma Aldrich vendor catalogue. Therefore, it was decided not to synthesize the desired compound.

2.2.6 Conclusion

In general the synthesis routes suggested by DOGS are feasible and even with comparatively low yields it is possible to obtain the desired product in a normally equipped laboratory with a general practical knowledge of organic synthesis. As shown in BZ-4 and BZ-5 the reactivity of the building blocks for the explicit combination of fragments is not always sure and there is also the possibility of different coupling reactions. Therefore a synthetic chemist is needed to check the feasibility and sometimes needs to decide whether different educts can make the compounds more reactive. The program DOGS does not take into consideration that some substituents can drastically change the reactivity.

2.3 Biological Evaluation

2.3.1 Fluorescence Kit

The activity of the synthesized compounds **BZ-1**, **BZ-2**, **BZ-3** was tested with a SensoLyte® 490 HIV-1 protease assay kit fluorometric (Z. Yu et al. 2010). The system is based on the Förster resonance energy transfer principle FRET. (Förster 1948)

A fluorescent fragment EDANS and the fluorescence suppressive DABCYL fragment are attached to a synthetic HIV protease substrate imitating the cleavage site p17/24 of the polypeptide Gag. The fluorescence of EDANS is quenched by DABCYL. If the synthetic peptide is cleaved by the protease, the distance between EDANS and DABCYL increases. The fluorescence is no longer quenched and can therefore be detected. (Figure 24)

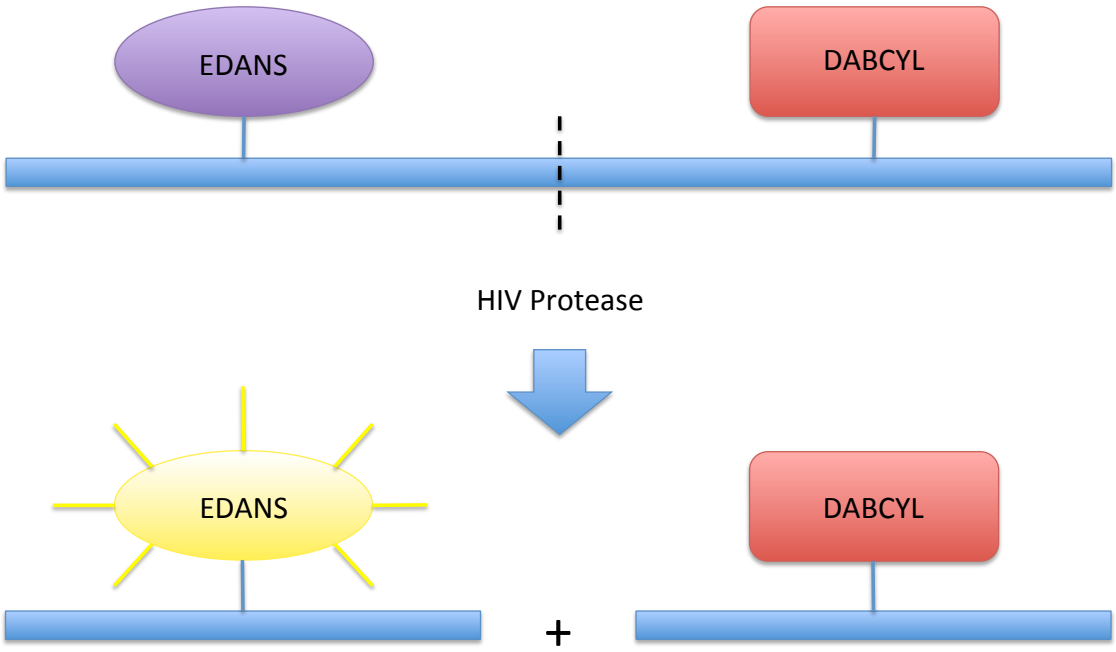


Figure 24: FRET peptide cleaved by HIV protease into fluorecent EDANS and quencher DABCYL

The increase of fluorescence over time is proportional to the substrate turnover of the protease over time.

If the protease is less productive e.g. because it is inhibited by a test compound, the increase of the fluorescence over time is lower or if the inhibition is absolute no fluorescence can be detected. If the fluorescence signal is detected in continuous time steps, the influence of the tested compounds on the protease can be assessed.

2.3.2 Biological evaluation of the synthesized compounds

The synthesized compounds were tested in three completely independent measurements. For each substance and concentration triplicates were tested. The final concentrations of the tested compounds were 100 µM, 10 µM, 1 µM, 100 nM and 10 nM.

The following controls were measured according to the protocol included in the test kit and treated equally to the probes: (from the AnaSpec Protocol)

- Positive control: HIV-protease diluent
- Inhibitor control: pepstatin A in assay buffer
- Vehicle control: assay buffer + DMSO
- Autofluorescence control: test compound + assay buffer
- Substrate control: only assay buffer

The protease was used in the concentration of 9.3 nM. As blank without inhibitor the assay buffer was used. As positive control the strong HIV protease inhibitor Pepstatin A was used. It inhibits the HIV protease to 100% at the concentration of 0.6 µM in which it was used in the assay. The wells with the negative control contained HIV protease and 1% DMSO without inhibitor or fluorescent substrate. As auto-fluorescent control the substances were measured alone in a 100 µM concentration in assay buffer.

All measurements were performed at room temperature in black microplates with a flat-bottom, 96 wells and a non-binding surface (Greiner 96 Flat Bottom Black Polystyrol) (Greiner Bio One GmbH, Frickenhausen, Germany) with a capacity of 100 µL per well.

The fluorescence was detected every 3 minutes (for the first protocol every 5 minutes) with a Tecan Infinite M 1000 platereader device (TECAN, Männedorf, Swiss) at 340 +/- 20 nm excitation and 490 +/- 20 nm emission.

SensoLyte® 490 HIV-1 protease assay kit fluorometric:

Table 15: Chemicals provided in the testkit

Component	Description	1 Amount
Component A	HIV-1 protease substrate EDANS/DABCYL FRET peptide	600 µL
Component B	EDANS, fluorescence standard reference	100 µM DMSO solution, 20 µL
Component C	Pepstatin A	27,4 µg powder
Component D	2x assay buffer	50 mL
Component E	Stop solution	30 mL
Component F	DMSO	100 µL
Component G	DTT	1 M, 200 µL

Further, the recombinant HIV-1 protease (protein) is required to perform the measurements. From Anaspec the amount of 5 mg was ordered 3 times for the diverse measurements (15 mg in total).

The test protocol provided in the assay kit was adapted especially with respect to the amounts and is listed below.

2.3.2.1 First test protocol: (Plate 1)

Preparation of the needed solutions:

- 1x assay buffer fresh for the experiment:

Table 16: Preparation of assay buffer

Components	Volume
2x assay buffer (component D)	5 mL
1 M DTT (1000x, component G)	10 µL
Deionized water	5 mL
Total volume	10 mL

- HIV-1 protease substrate solution:

Table 17: HIV-1 protease substrate solution

Components	Volume
HIV-1 protease substrate (50x, Component A)	100 µL
1x assay buffer	4,9 mL
Total volume	5 mL

- HIV-1 protease diluent: The provided 5 mg of HIV-1 protease (200 µM) were diluted in 500 µL assay buffer. In the tests the final concentration 9.3 nM (2 µL) and 23 nM (5 µL) were used.
- Reference ligand Pepstatin A as control inhibitor: Into the vial of Pepstatin A (component C) 20 µL of DMSO (component F) were added to get a concentration of 2 mM. The compounds were dissolved and mixed by vortexing. Then the solution was further diluted in assay buffer in a different vial to get a concentration of 60 µM. 1 µL was used as reference ligand, which resulted in a final concentration of 0.6 µM in the well.
- The test compounds were diluted to a concentration of 10 mM in DMSO and in a second step to the appropriate concentrations for testing (100 µM final

concentration in the well). 1 µL of test compound in DMSO was used for testing the inhibitory activity of the compound.

- The DMSO concentration was 1% in each well.

The first test series were composed as shown in the following table:

Table 18: First measurement – Pipetting scheme

	1	2
A	**Positive control:** 2 µL Protease 50 µL Substrate 1 µL DMSO 47 µL Buffer	**BZ-2 100 µM blank:** 1 µL BZ-2 100 µM 99 µL Buffer
B	**Positive control:** 2 µL Protease 50 µL Substrate 1 µL DMSO 47 µL Buffer	**BZ-2 100µM:** 2 µL Protease 50 µL Substrate 1 µL BZ-2 100 µM 47 µL Buffer
C	**Positive control:** 2 µL Protease 50 µL Substrate 1 µL DMSO 47 µL Buffer	**BZ-2 100 µM:** 2 µL Protease 50 µL Substrate 1 µL BZ-2 100 µM 47 µL Buffer
D	**Assay blank:** 100 µL Buffer	**BZ-2 100 µM:** 2 µL Protease 50 µL Substrate 1 µL BZ-2 100 µM 47 µL Buffer
E	**Pepstatin A pure:** 1 µL Pepstatin A 99 µL Buffer	**BZ-3 100 µM blank:** 1 µL BZ-3 100 µM 99 µL Buffer
F	**Pepstatin A:** 2 µL Protease 50 µL Substrate 1 µL Pepstatin A 47 µL Buffer	**BZ-3 100 µM:** 2 µL Protease 50 µL Substrate 1 µL BZ-3 100 µM 47 µL Buffer

G	**Pepstatin A:** 2 µL Protease 50 µL Substrate 1 µL Pepstatin A 47 µL Buffer	**BZ-3 100 µM:** 2 µL Protease 50 µL Substrate 1 µL BZ-3 100 µM 47 µL Buffer
H	**Pepstatin A:** 2 µL Protease 50 µL Substrate 1 µL Pepstatin A 47 µL Buffer	**BZ-3 100 µM:** 2 µL Protease 50 µL Substrate 1 µL BZ-3 100 µM 47 µL Buffer

The probes were measured in regular time steps of 5 minutes each for in total between 30 to 90 minutes. The first experiment was measured for 110 minutes to really reach the kinetic endpoint of the reaction.

BZ-1 was not measured in that test, because the purification was not completed at that time.

2.3.2.2 Second test protocol: (Plate 2)

The second test series was tested similar to the first protocol, but the protease concentration was raised to 23 nM (5 µL in the assay). Furthermore, BZ-1 was tested instead of BZ-2 which was not promising as an active inhibitor in the first test. The two potential inhibitors BZ-1 and BZ-3 were tested in different concentrations (100 µM, 10 µM, 1 µM, 100 nM, 10 nM) to estimate the K_D values of the compounds. The kinetic interval was set to only 3 minutes and 37 kinetic cycles were measured. For the results only the first 22 were used because then the plateau / end point of the reaction was reached. The measurements were carried out in triplicates. The composition of the wells is listed in

Table 19.

Table 19: Second plate – Pipeting scheme

	1	2	3	4	5	6
A	Positive control: 5µL Protease 50µL Substrate 1µL DMSO 44µL Assay buffer			Inhibitor control: 5µL Protease 50µL Substrate 1µL Pepstatin A 44µL Assay buffer		
B	BZ-3 100µM: 5µL Protease 50µL Substrate 1µL BZ-3 44µL Assay buffer			BZ-1 100µM: 5µL Protease 50µL Substrate 1µL BZ-1 44µL Assay buffer		
C	BZ-3 10µM: 5µL Protease 50µL Substrate 1µL BZ-3 44µL Assay buffer			BZ-1 10µM: 5µL Protease 50µL Substrate 1µL BZ-1 44µL Assay buffer		
D	BZ-3 1µM: 5µL Protease 50µL Substrate 1µL BZ-3 44µL Assay buffer			BZ-1 1µM: 5µL Protease 50µL Substrate 1µL BZ-1 44µL Assay buffer		

E	BZ-3 100nM: 5µL Protease 50µL Substrate 1µL BZ-3 44µL Assay buffer	BZ-1 100nM: 5µL Protease 50µL Substrate 1µL BZ-1 44µL Assay buffer
F	BZ-3 10nM: 5µL Protease 50µL Substrate 1µL BZ-3 44µL Assay buffer	BZ-3 blank: 1µL BZ-3 99µL Assay buffer
G	BZ-3 blank: 1µL BZ-3 99µL Assay buffer	Vehicle blank: 1µL DMSO 99µL Assay buffer

2.3.2.3 Third protocol

Even though the second protocol showed different results compared to the first and the inhibition in smaller concentrations seemed better than in higher concentrations the used protocol was changed and evaluated with a different manual suitable for fluorescent assays with HIV protease and the ordered test kit from AnaSpec. (Furfine 2001)

All the measurements were carried out at 340 +/- 20 nm excitation and 490 +/- 20 nm emission.

First, the fluorescence intensity of EDANS was measured to calibrate the plate reader.

Therefore, the assay buffer was freshly mixed according to the table below and the substrate solution was prepared. Then, the EDANS was diluted directly in the microplate and 50 µL of the substrate solution was added to each well to simulate the background fluorescence of the substrate comparable with the assay conditions.

Table 20: Dilution of the assay buffer

Assay buffer (2x)	5000µL
DTT	10µL
Deionized water	5000µL
Total	**10mL**

Table 21: Substrate solution

1x assay buffer	980µL
Substrate (200µM in DMSO)	20µL
Total	**1mL**

Concentration of substrate solution: 4µM

4 µL of EDANS were pipetted into well A1 and diluted with 46 µL of assay buffer dilute (see 1.1). This results in an 8 µM concentration in well A1. Directly in the plate EDANS is diluted further by pipetting 50 µL of well A1 to the next well (A2) according to the scheme in Table 22. By adding the 50 µL of substrate solution to simulate the background fluorescence of the substrate (inner filter effect) comparably to the assay conditions the concentration of EDANS in every well is again 2-fold resulting in the final concentrations in the right column. To have the same volume in every well 50 µL were removed from the last well.

Table 22: Dilution scheme of EDANS: Concentrations before and after adding the substrate

Dilution in the wells			Concentration	Conc. after adding substrate
4µL EDANS	+ 46µL buffer	→	8µM (A1)	4µM
50µL (A1)	+ 50µL buffer	→	4µM (A2)	2µM
50µL (A2)	+ 50µL buffer	→	2µM (A3)	1µM
50µL (A3)	+ 50µL buffer	→	1µM (A4)	0.5µM
50µL (A4)	+ 50µL buffer	→	0.5µM (A5)	0.25µM
50µL (A5)	+ 50µL buffer	→	0.25µM (A6)	0.125µM
50µL (A6)	+ 50µL buffer	→	0.125µM (A7)	0.0625µM
50µL (A7)	+ 50µL buffer	→	0.0625nM (A8)	0.03125µM

The fluorescence intensity was measured at extinction/emission 340/490 +/- 20 nm. The results are plotted in Figure 25 (EDANS concentration vs. relative fluorescence units RFU).

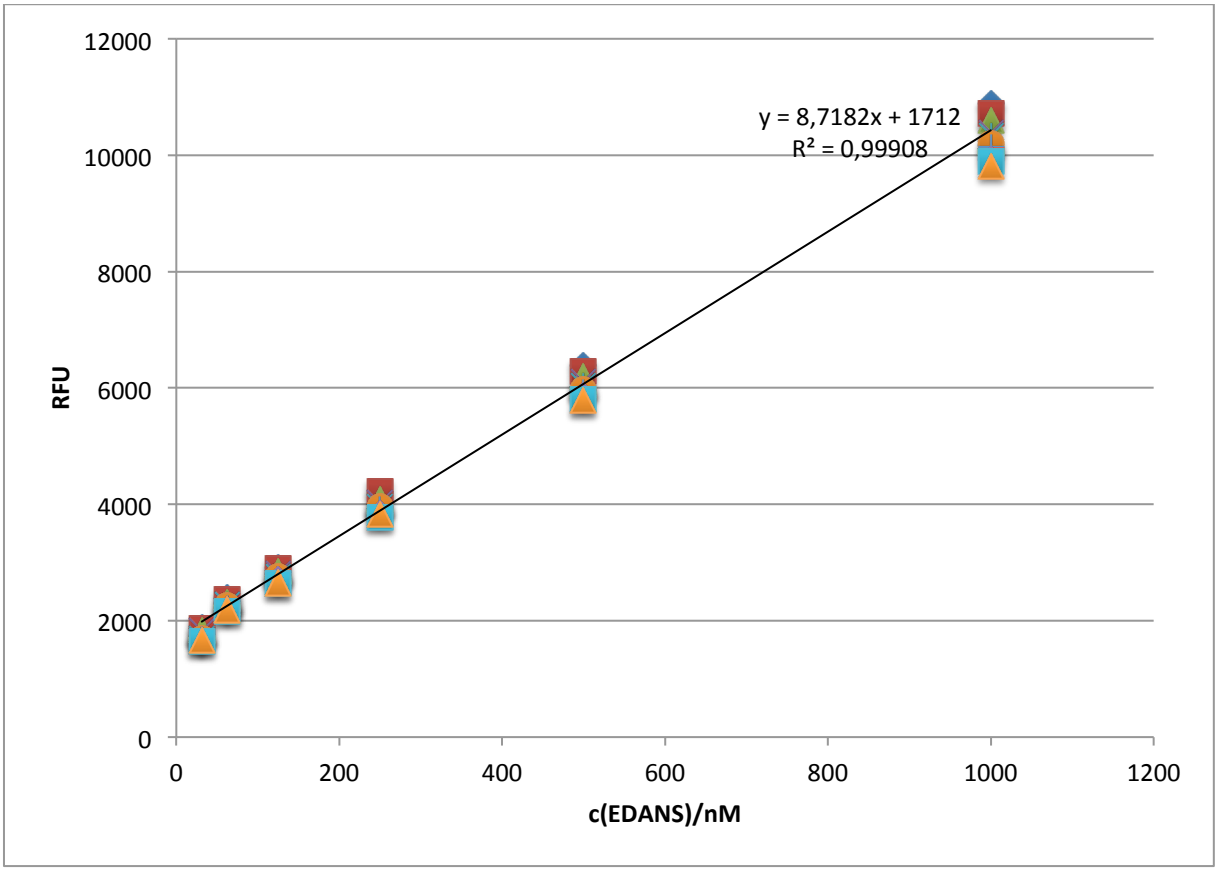


Figure 25: Calibration of the platereader with EDANS

In the second step the working enzyme concentration was determined.

Table 23: Product information HIV-1 protease (J Schneider & Kent 1988; Seelmeier et al. 1988)

MW (Dimer)	21584 g/mol
Distributed concentration	0.2mg/mL
Distributed volume	25µL
Distributed amount	5µg
Distributed concentration (Mol/L)	9µM

98 µL of assay buffer were pipetted into well A1 and always 50 µL into the wells A2-A12. Then 2 µL of HIV-protease stock solution were added to well A1 and mixed carefully with the

assay buffer using the pipettor. 50 µL of well A1 were pipetted into A2 and mixed. The step was repeated until well A12 is filled. The last 50 µL from A12 were removed and wasted.

Then 50 µL of substrate solution were very carefully pipetted into the wells A1-A12. This step dilutes the protease concentration 2-fold. The final enzyme concentrations are listed in the table below (Table 24).

Table 24: Concentration of HIV-protease in the wells

well	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
Conc [nM]	90	45	22.5	11.25	5.6	2.8	1.4	0.7	0.35	0.18	0.09	0.045

The fluorescence was measured for 2h at extinction/emission 340/490 +/- 20 nm. The data was plotted t[min] vs. RFU (Figure 26).

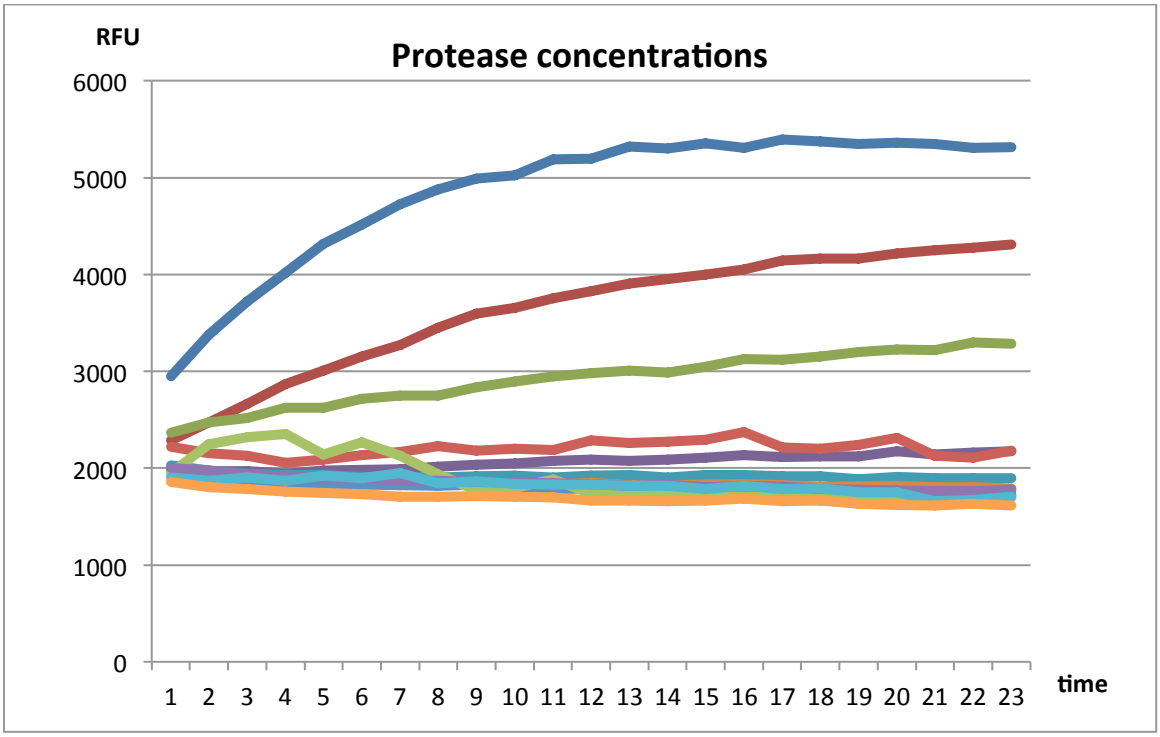


Figure 26: Different concentrations of HIV-protease measured to determine the working enzyme concentration

For the inhibitor testing the highest enzyme concentration yielding a linear substrate cleavage over 2 hours was used. For the measurements resulting in non-linear curves only the time range with linear characteristics of the curve was used to calculate the slope.

The final protease concentration used in the inhibitor assays was 22.5 nM.

Whit the determined enzyme concentration the potential inhibitors where subsequently measured again to clarify the results from the two previous measurements.

For each inhibitor 4 concentrations were measured: 100 µM, 10 µM, 1 µM, 100 nM. The inhibitors BZ-1 (and BZ-1.1) and BZ-2 (and BZ-2A) were only checked for activity. BZ-3 (and BZ-3A) which had already shown better results in the first test were tested in duplicates. For 100 µM every compound was tested in duplicates. In addition to the probes the following controls were tested:

- Pepstatin A (2 µM final concentration in the well) as reference inhibitor
- Blank
- DMSO control
- Substrate control (50 µL buffer, 50 µL substrate solution)

The test compounds were diluted in a well plate and contained 1% of DMSO each. Each well contained 10 µL of inhibitor in the desired final concentration, 40 µL of enzyme working solution and 50 µL of substrate solution (4 µM).

The probes were measured with excitation/emission 340/490 +/- 20 nM in a time interval of 3 minutes for 40 minutes in total.

2.3.3 Results and Discussion

On the first plate the substances BZ-2 and BZ-3 were tested for their general activity. They were both tested at the concentration of 100 µM in triplicates. Figure 27 shows the inhibition in per cent compared to the inhibition of Pepsatin A which is a very strong inhibitor of HIV protease. BZ-3 showed some activity in the test and it was considered to test BZ-3 together with BZ-1 in different concentrations to be able to estimate the K_D values of the two substances. BZ-2 showed less inhibition than BZ-3 (Figure 27) so we decided not to test it further.

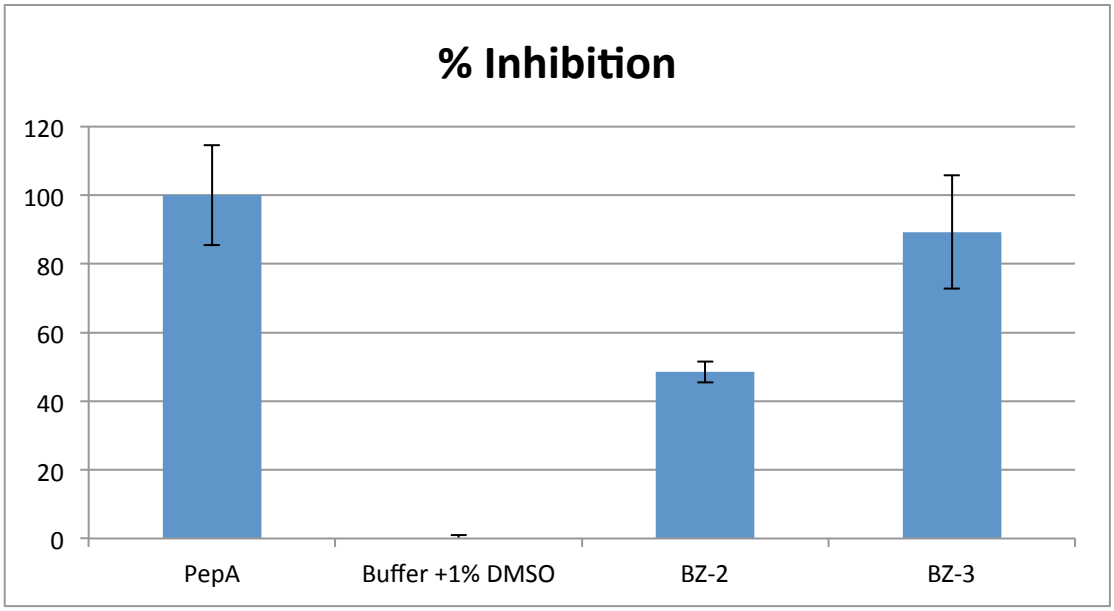


Figure 27: Per cent Inhibition according to the first test protocol measured in triplicates

The second measurement was then carried out with BZ-1 and BZ-3 in concentrations ranging from 100 µM to 10 nM. The results in per cent inhibition can be found in Figure 28. Several problems occurred during this test run evoking some questions about the reliability of the first test and the applicability of the test kit for the tested compounds in general.

The measurements were carried out in triplicates. At least two of the simultaneously tested wells showed the expected gain of fluorescence and the components showed some inhibition. But while analysing the data more closely it was found out that the inhibition did not correlate with the concentration used in the tests. The compound with 10 µM concentration showed less inhibition than the same compound in a 10 nM concentration which cannot be explained.

Some difficulties in the handling of the components in the test kit were obvious. While pipetting the substrate into the prepared wells already containing assay buffer, the mixture of the substrate solution and the assay buffer easily formed bubbles. It was difficult to get the wells free of bubbles.

For two wells containing BZ-1 in 10 nM concentration there was not enough HIV-protease left. The data for BZ-1 in 10 nM concentration is therefore only measured once.

In Figure 28 the per cent inhibition of that test run is depicted. There is obviously no correlation between the concentrations of the compound and their inhibition.

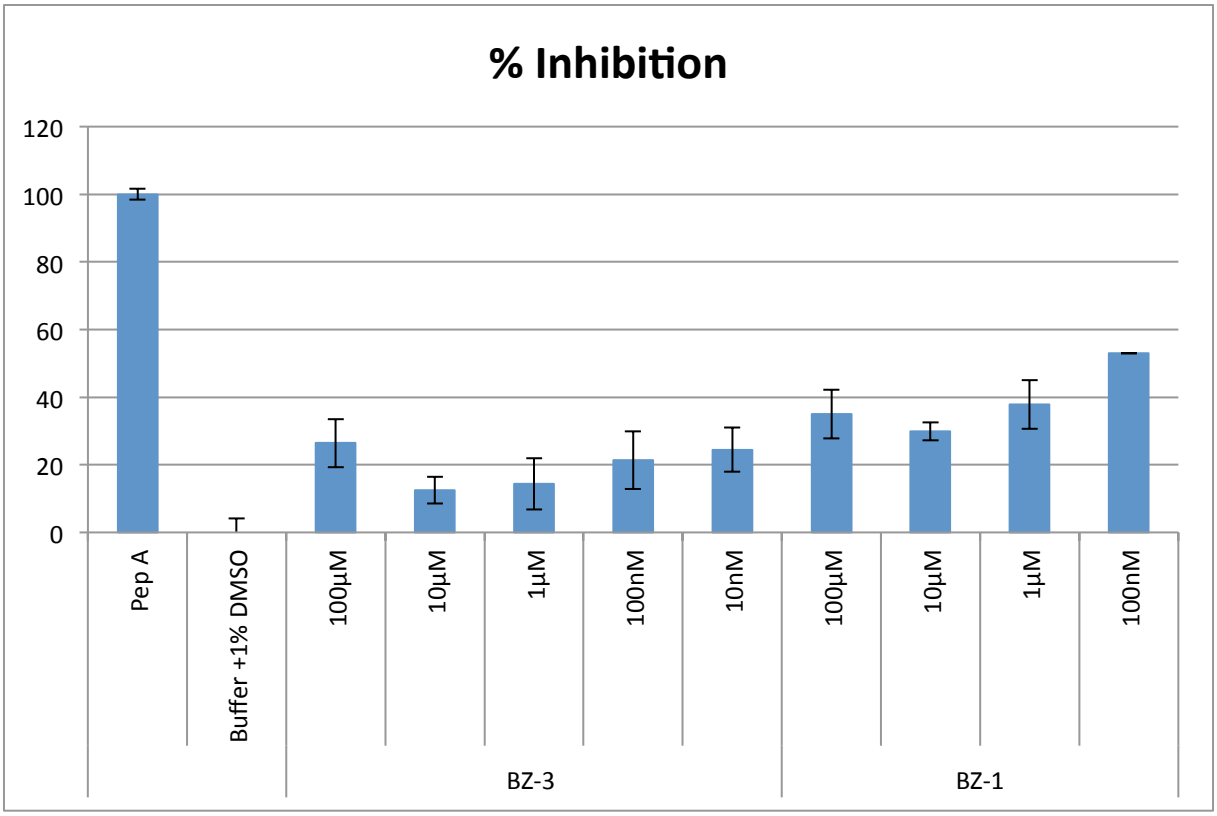


Figure 28: Per cent inhibition second test

Due to these results we checked whether the component might aggregate in the higher concentrations, which was actually very unlikely but could not be excluded at first.

The aggregation measurement was carried out by the group of Prof. Dr. Jean-Christophe Leroux with a Beckman Coulter DelsaNano C device.

No aggregation could be detected in any of the concentrations.

In a last step it was decided to modify the test protocol of the assay kit according to a paper that describes different standard protocols for HIV protease assays.

In preparation for the modified protocol the plate reader was calibrated. Then the adequate protease concentration could be determined as well as the working substrate concentration. The detailed composition of these solutions is listed in section 2.3.2.3.

The tests were then carried out with a protease concentration of 22.5 nM and a substrate concentration of 4 µM. It was also ensured to have the exact DMSO concentration of 1 % in each well. The substances were pipetted very carefully to keep the wells free of bubbles.

While setting up the measurement, the well with the highest RFU gain was chosen incorrectly and the first 3 kinetic cycles could not be used as valuable data. But after changing the reference well the measurement was continued normally.

For BZ-1, which was a violet substance, no activity could be detected in the higher concentrations. It is possible that the colour interfered with the fluorescence.

The measurement for BZ-2 worked out without problems, but the activity of the compound was very low. Further, the activity did not correlate with the tested concentrations.

The best compound in general was BZ-3 where the concentration and the per cent inhibition correlated with the concentration in the last experiment. (Figure 31) It was also the most promising compound in the first measurement. (Figure 27)

Still, according to the results of the tests we could not clearly declare one of the compounds as active or partially active against HIV protease. To be indicative for that, the test runs were not comparable enough and the results were too diverse.

In Figure 29, Figure 30 and Figure 31 each compound was only measured once.

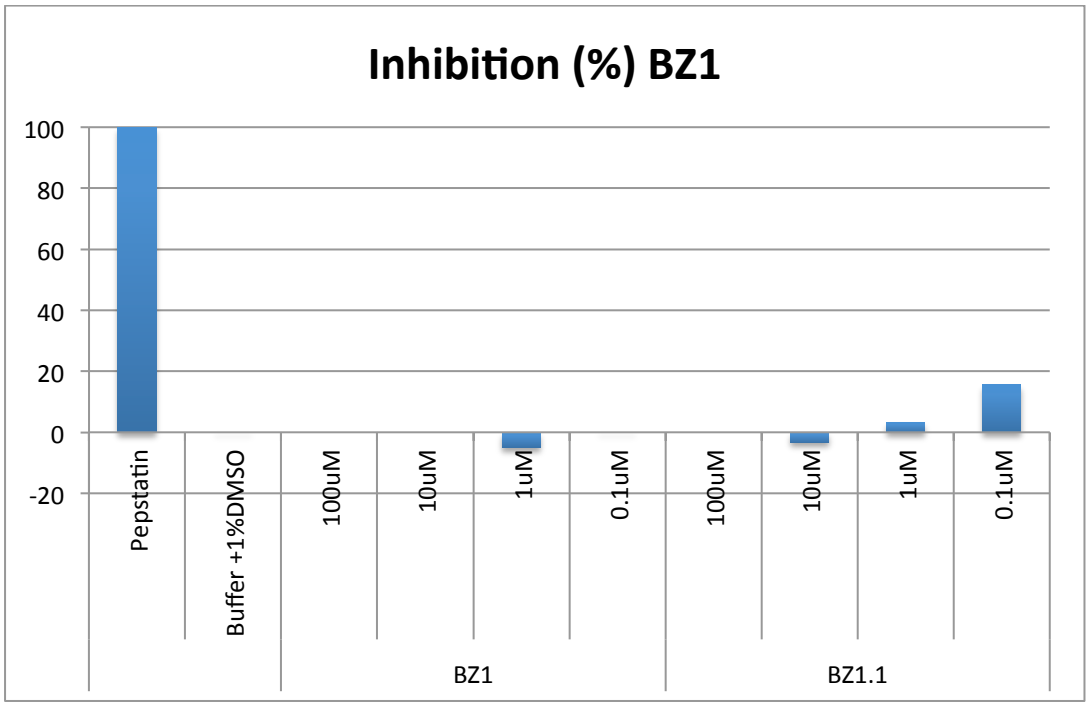


Figure 29: Per cent inhibition of BZ-1 according to the third protocol

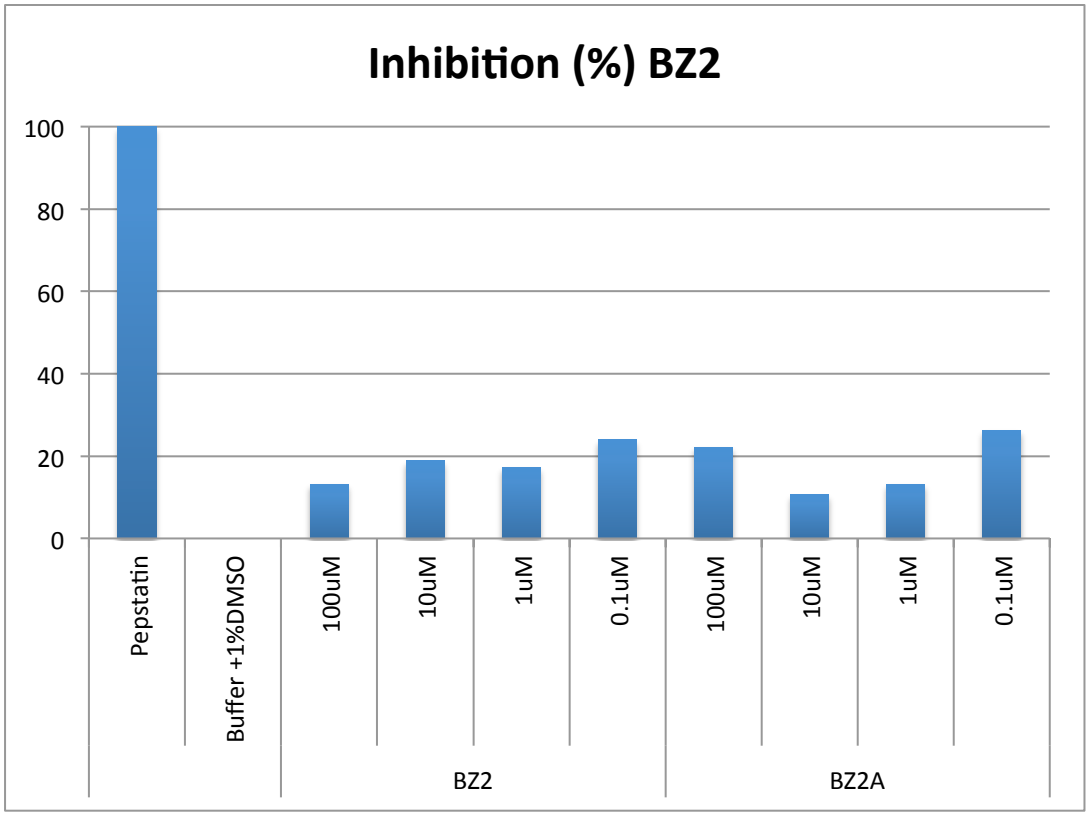


Figure 30: Per cent inhibition of BZ-2 according to the third protocol

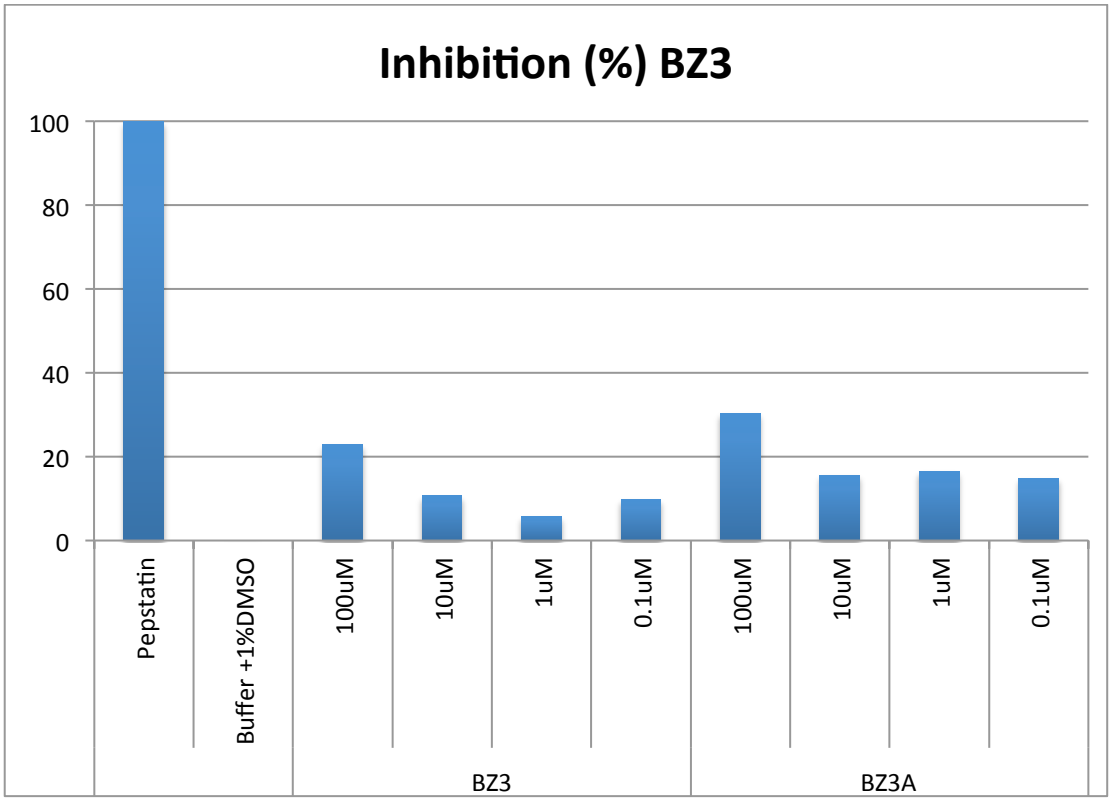


Figure 31: Per cent inhibition of BZ-3 according to the thrid protocol

In one additional step some derivatives were synthesised by the colleagues in the group and together with BZ-3 they were sent to a company to test the activity again under clearly reproducible conditions. None of the tested substances showed significant activity.

3 Summary, Conclusion and Outlook

The goal of this diploma thesis was to identify drug-like HIV protease inhibitors based on structures generated with the *de novo* design software DOGS. As a starting point several protease inhibitors were investigated more closely to find an optimal reference structure for the software. These include Darunavir, Nelfinavir, Brecanavir and Amprenavir, which served as starting points for generating molecules with DOGS. All the generated molecules were analysed according to their MW and their SlogP. Finally, Amprenavir was considered as reference molecule because of its low molecular weight and its diverse functional groups. Thus, compounds generated on the basis of Amprenavir were transformed into 3D conformers and docked into the binding pocket of HIV protease.

Due to time limitations and to obtain synthesizable molecules within the allocated time for the Diploma thesis only a subset of compounds was investigated in the docking experiment. The settings were chosen by default and the results can only give an idea about the possible interactions of the compounds with the protein, but not a reliable prediction. For precise predictions the interacting side-chains of HIV protease would have to be monitored more closely with respect to orientation and protonation state. In addition, all water molecules were removed from the binding pocket although some interactions of known inhibitors are only possible over the conserved water molecules in the binding cavity.

The molecules to be synthesised were chosen manually and with respect to synthetic feasibility and low molecular weight. The suggested reactions encoded in the files created by DOGS were checked and the building blocks were ordered. The molecules were not modified and the synthesis of three out of the six selected compounds was carried out strictly according to the reactions suggested by DOGS.

For testing the substances a Sensolyte FRET assay kit was ordered. This assay kit is helpful to test to synthesized structures for activity against the HIV protease. Due to time limitations it was not possible to measure the K_D values for the compounds because this would have required a different assay design. Further we got contradicting results for the compounds. It would be necessary to confirm the results with another assay or another technique to measure activity. After having the compounds tested in a company it was confirmed that the structures synthesized during the project were all not active against HIV protease.

As a resumée the project was partly successful because we were able to synthesize three compounds as suggested by DOGS. Considering the very limited number of compounds synthesized one cannot say whether the software was not suitable for the type of reference inhibitor or the reference was not a good template to generate HIV protease inhibitors. Probably some of the generated structures would be active and only the selection failed. Thus, further investigations will have to focus on a deeper investigation of the power of this new method.

4 Appendix

4.1 References

Adamson, C.S. & Freed, Eric O, 2007. Human Immunodeficiency Virus Type 1 Assembly , Release , and Maturation. *Advances in Pharmacology*, 55(07).

Agniswamy, J. et al., 2012. Terminal interface conformations modulate dimer stability prior to amino terminal autoprocessing of HIV-1 protease. *Biochemistry*, 51(5), pp.1041–50. Available at: http://europepmc.org/articles/PMC3287067 [Accessed April 12, 2013].

Aidshilfe Wien, 2012. Aidshilfe Wien. Available at: http://www.aids.at/ [Accessed April 1, 2012].

Aktories, K. et al., 2009. *Allgemeine und Spezielle Pharmakologie und Toxikologie: Begründet von W. Forth, D. Henschler, W. Rummel - mit Zugang zum Elsevier-Portal*, Urban & Fischer Verlag/Elsevier GmbH; Auflage: 10. Available at: http://www.amazon.com/Allgemeine-Spezielle-Pharmakologie-Toxikologie-Elsevier-Portal/dp/3437425226 [Accessed April 12, 2013].

Alterman, M. et al., 1998. Design and synthesis of new potent C2-symmetric HIV-1 protease inhibitors. Use of L-mannaric acid as a peptidomimetic scaffold. *Journal of medicinal chemistry*, 41(20), pp.3782–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9748353.

Barré-Sinoussi, F. et al., 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science (New York, N.Y.)*, 220(4599), pp.868–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6189183 [Accessed April 3, 2013].

Bemis, G W & Murcko, M A, 1996. The properties of known drugs. 1. Molecular frameworks. *Journal of medicinal chemistry*, 39(15), pp.2887–93. Available at: http://dx.doi.org/10.1021/jm9602928 [Accessed March 27, 2013].

Berman, H.M., 2000. The Protein Data Bank. *Nucleic Acids Research*, 28(1), pp.235–242. Available at: http://nar.oxfordjournals.org/content/28/1/235.abstract [Accessed February 28, 2013].

Bohacek, R.S. & McMartin, C., 1994. Multiple Highly Diverse Structures Complementary to Enzyme Binding Sites: Results of Extensive Application of a de Novo Design Method Incorporating Combinatorial Growth. *Journal of the American Chemical Society*, 116(13), pp.5560–5571. Available at: http://dx.doi.org/10.1021/ja00092a006 [Accessed April 13, 2013].

Brik, A. & Wong, C.-H., 2003. HIV-1 protease: mechanism and drug discovery. *Organic & Biomolecular Chemistry*, 1(1), pp.5–14. Available at: http://pubs.rsc.org/en/content/articlehtml/2003/ob/b208248a [Accessed April 12, 2013].

Brodt, H., 1998. HIV postexposure prophylaxis. German-Austrian recommendations. Deutsche AIDS-Gesellschaft (DAIG) and Osterreichische AIDS- Gesellschaft (OAG). *European journal of medical research*, 3(10), pp.485–500. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9753707 [Accessed April 12, 2013].

Brown, J., 1987. Approval of AZT. *Food and Drug Administration, U.S. Department of Health and Human Services*, ((202) 245), p.6867.

Brown, N. et al., 2004. A graph-based genetic algorithm and its application to the multiobjective evolution of median molecules. *Journal of chemical information and computer sciences*, 44(3), pp.1079–87. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15154776 [Accessed April 13, 2013].

Burke, D.S., 1997. Recombination in HIV: an important viral evolutionary strategy. *Emerging infectious diseases*, 3(3), pp.253–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2627633&tool=pmcentrez&rendertype=abstract [Accessed April 13, 2013].

But, T.Y.S. & Toy, P.H., 2007. The Mitsunobu reaction: origin, mechanism, improvements, and applications. *Chemistry, an Asian journal*, 2(11), pp.1340–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17890661 [Accessed March 15, 2012].

Böhm, H.J., 1993. A novel computational tool for automated structure-based drug design. *Journal of molecular recognition : JMR*, 6(3), pp.131–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8060670 [Accessed April 12, 2013].

Böhm, H.J., 1992a. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *Journal of computer-aided molecular design*, 6(6), pp.593–606. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1291628 [Accessed April 12, 2013].

Böhm, H.J., 1992b. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *Journal of computer-aided molecular design*, 6(1), pp.61–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1583540 [Accessed March 17, 2013].

Böhm, H.J., 1994. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *Journal of Computer-Aided Molecular Design*, 8(3), pp.243–256. Available at: http://link.springer.com/10.1007/BF00126743 [Accessed April 12, 2013].

Caflisch, A., Miranker, Andrew & Karplus, Martin, 1993. Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. *Journal of Medicinal Chemistry*, 36(15), pp.2142–2167. Available at: http://dx.doi.org/10.1021/jm00067a013 [Accessed April 13, 2013].

Campbell, N.A., 2009. *Biologie - Der neue Campbell*, Addison-Wesley Verlag; Auflage: 8., aktualisierte Auflage. Available at: http://www.amazon.de/Biologie-neue-Campbell-Pearson-Studium/dp/3827372879 [Accessed April 13, 2013].

CDC, C. of D.C., 1981. Karposi's Sarcoma and Pneumonia Among Homosexual Men - New York City and California. *MMWR Morbidity and Mortality Weekly Report*, 30, pp.305–308.

Chen, Z. et al., 1997. Human immunodeficiency virus type 2 (HIV-2) seroprevalence and characterization of a distinct HIV-2 genetic subtype from the natural range of simian immunodeficiency virus-infected sooty mangabeys. *Journal of virology*, 71, pp.3953–3960.

Clapham, P.R. & Weiss, R.A., 1997. Immunodeficiency viruses: Spoilt for choice of co-receptors. *Nature*, 388, pp.230–231.

Clark, D E et al., 1995. PRO-LIGAND: an approach to de novo molecular design. 1. Application to the design of organic molecules. *Journal of computer-aided molecular design*, 9(1), pp.13–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7751867 [Accessed April 13, 2013].

Clark, David E. & Murray, Christopher W., 1995. PRO_LIGAND: An Approach to de Novo Molecular Design. 5. Tools for the Analysis of Generated Structures. *Journal of Chemical Information and Modeling*, 35(5), pp.914–923. Available at: http://dx.doi.org/10.1021/ci00027a020 [Accessed April 13, 2013].

De Clercq, E., 2009. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *International journal of antimicrobial agents*, 33(4), pp.307–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19108994 [Accessed October 14, 2012].

Clumeck, N. et al., 1984. Acquired Immunodeficiency Syndrome in African Patients. *New England Journal of Medicine*, 310(8), pp.492–497. Available at: http://dx.doi.org/10.1056/NEJM198402233100804.

Coffin, J. et al., 1986. Human immunodeficiency viruses. *Science*, 232, p.697.

Dalby, A. et al., 1992. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *Journal of Chemical Information and Computer Sciences*, 32(3), pp.244–255. Available at: http://dx.doi.org/10.1021/ci00007a012.

Damewood, J.R., Lerman, C.L. & Masek, B.B., 2010. NovoFLAP: A ligand-based de novo design approach for the generation of medicinally relevant ideas. *Journal of chemical information and modeling*, 50(7), pp.1296–303. Available at: http://dx.doi.org/10.1021/ci100080r [Accessed April 13, 2013].

Dandache, S. et al., 2007. In vitro antiviral activity and cross-resistance profile of PL-100, a novel protease inhibitor of human immunodeficiency virus type 1. *Antimicrobial agents and chemotherapy*, 51(11), pp.4036–43. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2151432&tool=pmcentrez&rendertype=abstract [Accessed March 20, 2012].

Danziger, D.J. & Dean, P M, 1989. Automated site-directed drug design: a general algorithm for knowledge acquisition about hydrogen-bonding regions at protein surfaces. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)*, 236(1283), pp.101–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2565575 [Accessed April 12, 2013].

Davies, D.R., 1990. The structure and function of the aspartic proteinases. *Annual review of biophysics and biophysical chemistry*, 19, pp.189–215. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev.bb.19.060190.001201 [Accessed April 12, 2013].

Degen, J. & Rarey, M., 2006. FlexNovo: structure-based searching in large fragment spaces. *ChemMedChem*, 1(8), pp.854–68. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16902939 [Accessed March 10, 2013].

DeWitte, R.S. & Shakhnovich, E.I., 1996. SMoG: de Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence. *Journal of the American Chemical Society*, 118(47), pp.11733–11744. Available at: http://dx.doi.org/10.1021/ja960751u [Accessed April 13, 2013].

Dey, F. & Caflisch, A., 2008. Fragment-based de novo ligand design by multiobjective evolutionary optimization. *Journal of chemical information and modeling*, 48(3), pp.679–90. Available at: http://dx.doi.org/10.1021/ci700424b [Accessed April 13, 2013].

Douguet, D, Thoreau, E. & Grassy, G., 2000. A genetic algorithm for the automated generation of small organic molecules: drug design using an evolutionary algorithm. *Journal of computer-aided molecular design*, 14(5), pp.449–66. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10896317 [Accessed April 13, 2013].

Douguet, Dominique et al., 2005. LEA3D: a computer-aided ligand design for structure-based drug design. *Journal of medicinal chemistry*, 48(7), pp.2457–68. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15801836 [Accessed April 13, 2013].

Dunn, B.M., 2002. Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chemical reviews*, 102(12), pp.4431–58. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12475196 [Accessed March 1, 2013].

Durrant, J.D., Amaro, R.E. & McCammon, J.A., 2009. AutoGrow: a novel algorithm for protein inhibitor design. *Chemical biology & drug design*, 73(2), pp.168–78. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2724963&tool=pmcentrez&rendertype=abstract [Accessed April 13, 2013].

Eisen, M.B. et al., 1994. HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins*, 19, pp.199–221.

Eldridge, M.D. et al., 1997. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes.

Journal of computer-aided molecular design, 11(5), pp.425–45. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9385547 [Accessed March 17, 2013].

Emerman, M., Vazeux, R. & Peden, K., 1989. The rev Gene Product of the Human Immunodeficiency Virus Affects Envelope-Specific RNA Localization. *Cell*, 57, pp.1155–1165.

Fechner, U. & Schneider, Gisbert, 2006. Flux (1): a virtual synthesis scheme for fragment-based de novo design. *Journal of chemical information and modeling*, 46(2), pp.699–707. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16563000 [Accessed April 13, 2013].

Fechner, U. & Schneider, Gisbert, 2007. Flux (2): comparison of molecular mutation and crossover operators for ligand-based de novo design. *Journal of chemical information and modeling*, 47(2), pp.656–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17315990 [Accessed April 13, 2013].

Feher, M. et al., 2008. The use of ligand-based de novo design for scaffold hopping and sidechain optimization: two case studies. *Bioorganic & medicinal chemistry*, 16(1), pp.422–7. Available at: http://dx.doi.org/10.1016/j.bmc.2007.09.026 [Accessed March 10, 2013].

Franchetti, P. et al., 1998. Potent and Selective Inhibitors of Human Immunodeficiency Virus Protease Structurally Related to L-694,- 746. *Antiviral Chemistry & Chemotherapy*, 9, pp.303–309.

Freed, E O, 2001a. HIV-1 replication. *Somatic cell and molecular genetics*, 26(1-6), pp.13–33. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2815006&tool=pmcentrez&rendertype=abstract.

Freed, E O, 2001b. HIV-Replication. *Somatic cell and molecular genetics*, 26(1/6), pp.13–33.

Freed, E O & Martin, M.A., 2001. HIVs and their replication. In D. Knipe et al., eds. *Fields Virology, 4th ed.* Philadelphia: Lippincott, Wiliams and Wilkins, pp. 1971–2041.

Freed, Eric O, 1998. MINIREVIEW HIV-1 Gag Proteins : Diverse Functions in the Virus Life Cycle. *Virology*, 251, pp.1–15.

Frenkel, David et al., 1995. PRO_LIGAND: An approach to de novo molecular design. 4. Application to the design of peptides. *Journal of Computer-Aided Molecular Design*, 9(3), pp.213–225. Available at: http://link.springer.com/10.1007/BF00124453 [Accessed April 13, 2013].

Furfine, E.S., 2001. HIV protease assays. *Current protocols in pharmacology / editorial board, S.J. Enna (editor-in-chief) ... [et al.]*, Chapter 3, p.Unit3.2. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21971798 [Accessed April 13, 2013].

Förster, T., 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik*, 437(1-2), pp.55–75. Available at: http://doi.wiley.com/10.1002/andp.19484370105 [Accessed February 28, 2013].

G J Moyle,D Back, 2001. Principles and practice of HIV-protease inhibitor pharmacoenhancement. *HIV medicine*, 2(2), pp.105 – 13. Available at: http://www.researchgate.net/publication/11619829_Principles_and_practice_of_HIV-protease_inhibitor_pharmacoenhancement [Accessed April 12, 2013].

Gao, F. et al., 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature*, 397, pp.436–441.

Garnier, L., Bowzard, J.B. & Wills, J.W., 1998. Recent advances and remaining problems in HIV assembly. *Aids*, 12 Suppl.A, pp.S5–16.

Gasteiger, J., Rudolph, C. & Sadowski, J., 1990. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Computer Methodology*, 3(6), pp.537–547. Available at: http://dx.doi.org/10.1016/0898-5529(90)90156-3 [Accessed April 13, 2013].

Ghosh, A.K. et al., 2008. Design of HIV protease inhibitors targeting protein backbone: an effective strategy for combating drug resistance. *Accounts of chemical research*, 41(1), pp.78–86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17722874 [Accessed April 13, 2013].

Gillet, V.J. et al., 1990. Automated structure design in 3D. *Tetrahedron Computer Methodology*, 3(6), pp.681–696. Available at: http://dx.doi.org/10.1016/0898-5529(90)90167-7 [Accessed April 12, 2013].

Gillet, Valerie J et al., 1993. SPROUT: A program for structure generation. *Journal of Computer-Aided Molecular Design*, 7(2), pp.127–153. Available at: http://link.springer.com/10.1007/BF00126441 [Accessed April 12, 2013].

Glen, R.C. & Payne, A.W., 1995. A genetic algorithm for the automated generation of molecules within constraints. *Journal of computer-aided molecular design*, 9(2), pp.181–202. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7608749 [Accessed April 13, 2013].

Globus, A., Lawton, J. & Wipke, T., 1999. Automatic molecular design using evolutionary techniques. *Nanotechnology*, 10(3), pp.290–299. Available at: http://stacks.iop.org/0957-4484/10/i=3/a=312.

Gottlieb, M.S. et al., 1981. Pneumocystis Carinii Pneumonia and Mucosal Candidiasis in Previously Healthy Homosexual Men. *N. Engl. J. Med*, 305, pp.507–8.

Greene, W.C., 1993. Aids and the immune-system. *Scientific AmScientific American*, 269, pp.98–105.

Grzybowski, B.A. et al., 2002. Combinatorial computational method gives new picomolar ligands for a known enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3), pp.1270–3. Available at: http://www.pnas.org/content/99/3/1270.short [Accessed April 13, 2013].

Gulnik, S., Afonina, E. & Eissenstat, M., 2010. HIV-1 Protease Inhibitors as Antiretroviral Agents. In *Enzyme inhibition in Drug Discovery and Development: The Good and the Bad*. pp. 749–810.

Harper, M.E. et al., 1986. Detection of lymphocytes expressing human T-lymphotropic virus type III in lymph nodes and peripheral blood from infected individuals by in situ hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, 83, pp.772–776.

Hartenfeller, M, 2010. Development of a computational method for reaction-driven de novo design of druglike compounds. *Frankfurt am Main: Goethe …*. Available at: http://publikationen.stub.uni-frankfurt.de/frontdoor/deliver/index/docId/21104/file/Dissertation_Markus_Hartenfeller.pdf [Accessed August 21, 2013].

Hartenfeller, Markus et al., 2008. Concept of combinatorial de novo design of drug-like molecules by particle swarm optimization. *Chemical biology & drug design*, 72(1), pp.16–26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18564216 [Accessed March 7, 2013].

Hartenfeller, Markus et al., 2012. DOGS: Reaction-Driven de novo Design of Bioactive Compounds. *PLoS computational biology*, 8(2), p.e1002380. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3280956&tool=pmcentrez&rendertype=abstract.

Hartenfeller, Markus & Schneider, Gisbert, 2011a. De novo drug design. *Methods in molecular biology (Clifton, N.J.)*, 672, pp.299–323. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20838974 [Accessed April 13, 2013].

Hartenfeller, Markus & Schneider, Gisbert, 2011b. Enabling future drug discovery by de novo design. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5), pp.742–759. Available at: http://doi.wiley.com/10.1002/wcms.49 [Accessed March 19, 2012].

Hecht, D. & Fogel, G.B., 2009. A novel in silico approach to drug discovery via computational intelligence. *Journal of chemical information and modeling*, 49(4), pp.1105–21. Available at: http://dx.doi.org/10.1021/ci9000647 [Accessed April 13, 2013].

Herold, G., 2011. *Innere Medizin 2011 [Taschenbuch]*, Herold,; Auflage: 2011. Available at: http://www.amazon.de/Innere-Medizin-2011-Gerd-Herold/dp/B00442K7RQ [Accessed April 13, 2013].

Hiss, J. a, Hartenfeller, Markus & Schneider, Gisbert, 2010. Concepts and applications of "natural computing" techniques in de novo drug and peptide design. *Current*

pharmaceutical design, 16(15), pp.1656–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20222857.

Ho, C.M.W. & Marshall, G.R., 1993. SPLICE: A program to assemble partial query solutions from three-dimensional database searches into novel ligands. *Journal of Computer-Aided Molecular Design*, 7(6), pp.623–647. Available at: http://link.springer.com/10.1007/BF00125322 [Accessed April 12, 2013].

Ho, D.D., 1995. Time to hit HIV, early and hard. *New England Journal of Medicine*, 333(7), pp.450–1.

Hodge, C.N. et al., 1996. Improved cyclic urea inhibitors of the HIV-1 protease: synthesis, potency, resistance profile, human pharmacokinetics and X-ray crystal structure of DMP 450. *Chemistry & biology*, 3(4), pp.301–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8807858.

Hornak, V. & Simmerling, C., 2007. Targeting structural flexibility in HIV-1 protease inhibitor binding. *Drug Discovery Today*, 12(75), pp.132–138.

Huang, Q., Li, L.-L. & Yang, S.-Y., 2010. PhDD: a new pharmacophore-based de novo design method of drug-like molecules combined with assessment of synthetic accessibility. *Journal of molecular graphics & modelling*, 28(8), pp.775–87. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20206562 [Accessed March 11, 2013].

Hyland, L.J., Tomaszek, T.A. & Meek, T.D., 1991. Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry*, 30(34), pp.8454–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1883831 [Accessed April 12, 2013].

Irwin, J.J. & Shoichet, B.K., 2005. ZINC--a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, 45(1), pp.177–82. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1360656&tool=pmcentrez&rendertype=abstract [Accessed March 13, 2013].

Ishchenko, A. V & Shakhnovich, E.I., 2002. SMall Molecule Growth 2001 (SMoG2001): an improved knowledge-based scoring function for protein-ligand interactions. *Journal of medicinal chemistry*, 45(13), pp.2770–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12061879 [Accessed April 13, 2013].

Johnson, M.A., Maggiora, G.M. & Meeting, A.C.S., 1990. *Concepts and applications of molecular similarity*, Available at: http://books.google.at/books/about/Concepts_and_applications_of_molecular_s.html?id=iGbJaz1ziWkC&pgis=1 [Accessed August 20, 2013].

Jones, G. et al., 1997. Development and validation of a genetic algorithm for flexible docking. *Journal of molecular biology*, 267(3), pp.727–48. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9126849 [Accessed March 7, 2013].

Jones, K.A. & Peterlin, B.M., 1994. Control of RNA initiation and elongation at the HIV-1 promotor. *Annual review of biochemistry*, 63, pp.717–743.

Joyce, C. & Anderson, I., 1985. Us licenses blood test for AIDS. *New Scientist*, 105, pp.3–4.

Kempf, D.J. et al., 1995. ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 92(7), pp.2484–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=42242&tool=pmcentrez&rendertype=abstract [Accessed April 12, 2013].

Kennedy, J. & Eberhart, R., 1995. Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, 4, pp.1942–1948. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=488968.

King, N.M. et al., 2012. Extreme entropy-enthalpy compensation in a drug-resistant variant of HIV-1 protease. *ACS chemical biology*, 7(9), pp.1536–46. Available at: http://www.researchgate.net/publication/227394209_Extreme_Entropy-Enthalpy_Compensation_in_a_Drug-Resistant_Variant_of_HIV-1_Protease?ev=prf_pub [Accessed April 12, 2013].

Klebe, G & Böhm, H.J., 1997. Energetic and entropic factors determining binding affinity in protein-ligand complexes. *Journal of receptor and signal transduction research*, 17(1-3), pp.459–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9029508 [Accessed August 21, 2013].

Klebe, Gerhard, 2006. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today*, 11(13-14), pp.580–594. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1359644606001784 [Accessed October 25, 2012].

Klebe, Gerhard, 2009. Wirkstoffdesign. Available at: http://ebooks.ub.uni-muenchen.de/17690/ [Accessed April 13, 2013].

Kohl, N.E. et al., 1988. Active human immunodeficiency virus protease is required for viral infectivity. *Proceedings of the National Academy of Sciences of the United States of America*, 85(13), pp.4686–90. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=280500&tool=pmcentrez&rendertype=abstract.

Kutchukian, P.S., Lou, D. & Shakhnovich, E.I., 2009. FOG: Fragment Optimized Growth algorithm for the de novo generation of molecules occupying druglike chemical space. *Journal of chemical information and modeling*, 49(7), pp.1630–42. Available at: http://dx.doi.org/10.1021/ci9000458 [Accessed April 13, 2013].

Lalezari, J.P. et al., 2007. Preliminary safety and efficacy data of brecanavir, a novel HIV-1 protease inhibitor: 24 week data from study HPR10006. *The Journal of antimicrobial chemotherapy*, 60(1), pp.170–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17491001 [Accessed March 22, 2012].

Lamarre, D. et al., 1997. Antiviral properties of palinavir, a potent inhibitor of the human immunodeficiency virus type 1 protease. *Antimicrobial agents and chemotherapy*, 41(5), pp.965–71. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=163834&tool=pmcentrez&rendertype=abstract.

Lapatto, R. et al., 1989. X-ray analysis of HIV-1 proteinase at 2.7 A resolution confirms structural homology among retroviral enzymes. *Nature*, 342(6247), pp.299–302. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2682266 [Accessed April 13, 2013].

Lewis, R.A. et al., 1992. Automated site-directed drug design using molecular lattices. *Journal of Molecular Graphics*, 10(2), pp.66–78. Available at: http://dx.doi.org/10.1016/0263-7855(92)80059-M [Accessed April 12, 2013].

Lewis, R.A. & Dean, P M, 1989a. Automated site-directed drug design: the concept of spacer skeletons for primary structure generation. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)*, 236(1283), pp.125–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2565577 [Accessed April 12, 2013].

Lewis, R.A. & Dean, P M, 1989b. Automated site-directed drug design: the formation of molecular templates in primary structure generation. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)*, 236(1283), pp.141–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2565578 [Accessed April 12, 2013].

Liu, H. et al., 1999. Structure-based ligand design by dynamically assembling molecular building blocks at binding site. *Proteins*, 36(4), pp.462–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10450088 [Accessed April 13, 2013].

Louis, J M, Clore, G.M. & Gronenborn, A.M., 1999. Autoprocessing of HIV-1 protease is tightly coupled to protein folding. *Nature structural biology*, 6(9), pp.868–75. Available at: http://dx.doi.org/10.1038/12327 [Accessed April 12, 2013].

Louis, John M et al., 2007. HIV-1 protease: structure, dynamics, and inhibition. *Advances in pharmacology (San Diego, Calif.)*, 55(07), pp.261–98. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17586318 [Accessed March 22, 2012].

Luban, J., 1996. Absconding with the chaperone: essential cyclophilin-Gag interaction in HIV-1 virions. *Cell*, 87(7), pp.1157–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8980222.

Luo, Z., Wang, R & Lai, L, 1996. RASSE: a new method for structure-based drug design. *Journal of chemical information and computer sciences*, 36(6), pp.1187–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8941995 [Accessed April 13, 2013].

Markgren, P. et al., 2002. Relationships between structure and interaction kinetic for HIV-1 Protease inhibitors. *Journal of medicinal chemistry*, 45, pp.5430–5439.

Mata, Paulina et al., 1995. SPROUT: 3D Structure Generation Using Templates. *Journal of Chemical Information and Modeling*, 35(3), pp.479–493. Available at: http://dx.doi.org/10.1021/ci00025a016 [Accessed April 12, 2013].

McBurney, B., Foss, P. & Reed, E., 2004. 1-Ethoxy-4-nitrobenzene. *Acta Crystallographica Section E: Structure Reports*, (3), pp.2179–2181. Available at: http://scripts.iucr.org/cgi-bin/paper?ac6133 [Accessed April 13, 2013].

Meek, T.D., Rodriguez, E.J. & Angeles, T.S., 1994. Use of steady state kinetic methods to elucidate the kinetic and chemical mechanisms of retroviral proteases. *Methods in enzymology*, 241, pp.127–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7854175 [Accessed April 12, 2013].

Miranker, A & Karplus, M, 1991. Functionalitiy maps of binding sites: a multiple copy simultaneous search method. *Proteins*, 11, pp.29–34.

Mitsuya, H et al., 1985. 3'-Azido-3'-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 82(20), pp.7096–100. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=391317&tool=pmcentrez&rendertype=abstract.

MMWR Weekly, 1983. Current Trends Aquired Immunodeficiency Syndrome (AIDS) Update - United States. *MMWR Morbidity and Mortality Weekly Report*, 32(24), pp.309–11.

Moriaud, F. et al., 2009. Computational fragment-based approach at PDB scale by protein local similarity. *Journal of chemical information and modeling*, 49(2), pp.280–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19434830 [Accessed March 10, 2013].

Murray, C W et al., 1997. PRO_SELECT: combining structure-based drug design and combinatorial chemistry for rapid lead discovery. 1. Technology. *Journal of computer-aided molecular design*, 11(2), pp.193–207. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9089436 [Accessed April 13, 2013].

Murray, Christopher W., Clark, David E. & Byrne, D.G., 1995. PRO_LIGAND: An approach to de novo molecular design. 6. Flexible fitting in the design of peptides. *Journal of Computer-Aided Molecular Design*, 9(5), pp.381–395. Available at: http://link.springer.com/10.1007/BF00123996 [Accessed April 13, 2013].

Nachbar, R.B., 1998. Molecular evolution: A hierarchical representation for chemical topology and its automated manipulation. In *Proceedings of the Third Annual Conference of Genetic Programming 1998*. pp. 246–253. Available at: http://130.203.133.150/showciting;jsessionid=8891F80B19D72FF4980002EE8FC01461?cid=1219405 [Accessed April 13, 2013].

Nachbar, R.B., 2000. Molecular Evolution: Automated Manipulation of Hierarchical Chemical Topology and Its Application to Average Molecular Structures. *Genetic Programming*

and Evolvable Machines, 1(1-2), pp.57–94. Available at: http://link.springer.com/article/10.1023/A:1010072431120 [Accessed April 13, 2013].

Navia, M.A. et al., 1989. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature*, 337(6208), pp.615–20. Available at: http://dx.doi.org/10.1038/337615a0 [Accessed April 13, 2013].

Nicolaou, C.A., Apostolakis, J. & Pattichis, C.S., 2009. De novo drug design using multiobjective evolutionary graphs. *Journal of chemical information and modeling*, 49(2), pp.295–307. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19434831 [Accessed March 10, 2013].

Niederman, T.M., Thielan, B.J. & Ratner, L., 1989. Human immunodeficiency virus type 1 negative factor is a transcriptional silencer. *Proceedings of the National Academy of Sciences of the United States of America*, 86, pp.1128–1132.

Nikitin, S. et al., 2005. A very large diversity space of synthetically accessible compounds for use with drug design programs. *Journal of computer-aided molecular design*, 19(1), pp.47–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16059666 [Accessed April 13, 2013].

Nisius, B. & Rester, U., 2009. Fragment shuffling: an automated workflow for three-dimensional fragment-based ligand design. *Journal of chemical information and modeling*, 49(5), pp.1211–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19413347 [Accessed March 10, 2013].

Northrop, D.B., 2001. Follow the Protons: A Low-Barrier Hydrogen Bond Unifies the Mechanisms of the Aspartic Proteases. *Accounts of Chemical Research*, 34(10), pp.790–797. Available at: http://dx.doi.org/10.1021/ar000184m [Accessed April 12, 2013].

Ohno, M., Fornerod, M. & Mattaj, I.W., 1998. Nucleocytoplasmic transport the last 200 nanometers. *Cell*, 92, pp.327–336.

Pearlman, D.A. & Murcko, Mark A., 1993. CONCEPTS: New dynamic algorithm forde novo drug suggestion. *Journal of Computational Chemistry*, 14(10), pp.1184–1193. Available at: http://doi.wiley.com/10.1002/jcc.540141008 [Accessed April 12, 2013].

Pegg, S.C., Haresco, J.J. & Kuntz, I D, 2001. A genetic algorithm for structure-based de novo design. *Journal of computer-aided molecular design*, 15(10), pp.911–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11918076 [Accessed April 13, 2013].

Pellegrini, E. & Field, M.J., 2003. Development and testing of a de novo drug-design algorithm. *Journal of computer-aided molecular design*, 17(10), pp.621–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15068363 [Accessed April 13, 2013].

Pfeffer, P. et al., 2010. GARLig: a fully automated tool for subset selection of large fragment spaces via a self-adaptive genetic algorithm. *Journal of chemical information and modeling*, 50(9), pp.1644–59. Available at: http://dx.doi.org/10.1021/ci9003305 [Accessed April 13, 2013].

Pierce, A.C., Rao, G. & Bemis, Guy W, 2004. BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, p38, and HIV protease. *Journal of medicinal chemistry*, 47(11), pp.2768–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15139755 [Accessed April 13, 2013].

Piot, P. et al., 1984. Acquired immunodeficiency syndrome in a heterosexual population in Zaire. *The Lancet*, 2((8394)), pp.65–9.

Polgar, L., Szeltner, Z. & Boros, I., 1994. Substrate-Dependent Mechanisms in the Catalysis of Human Immunodeficiency Virus Protease. *Biochemistry*, 33(31), pp.9351–9357. Available at: http://dx.doi.org/10.1021/bi00197a040 [Accessed April 12, 2013].

Prabakaran, P. et al., 2007. Structure and function of the HIV envelope glycoprotein as entry mediator, vaccine immunogen, and target for inhibitors. *Advances in pharmacology (San Diego, Calif.)*, 55(07), pp.33–97. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17586312 [Accessed March 22, 2012].

Proschak, E., Zettl, H., et al., 2009. From molecular shape to potent bioactive agents I: bioisosteric replacement of molecular fragments. *ChemMedChem*, 4(1), pp.41–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19058279 [Accessed March 1, 2013].

Proschak, E., Sander, K., et al., 2009. From molecular shape to potent bioactive agents II: fragment-based de novo design. *ChemMedChem*, 4(1), pp.45–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19058280 [Accessed April 13, 2013].

Quinn, T.C. et al., 1986. AIDS in Africa: An Epidemiologic Paradigm. *Science*, 234, pp.955–963.

Quiñones-Mateu, M.E. et al., 2008. Viral drug resistance and fitness. *Advances in pharmacology (San Diego, Calif.)*, 56(07), pp.257–96. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18086415 [Accessed March 22, 2012].

Ratner, L. et al., 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature*, 313, pp.277–284.

Ratner, L., 1993. HIV life cycle and genetic approaches. *Prospectives in Drug Discovery and Design*, 1, pp.3–22.

Rechenberg, I., 1973. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*, Available at: http://books.google.de/books/about/Evolutionsstrategie.html?hl=de&id=-WAQAQAAMAAJ&pgis=1 [Accessed August 20, 2013].

Rotstein, S.H. & Murcko, Mark A., 1993. GroupBuild: a fragment-based method for de novo drug design. *Journal of Medicinal Chemistry*, 36(12), pp.1700–1710. Available at: http://dx.doi.org/10.1021/jm00064a003 [Accessed April 12, 2013].

Rupp, M., Proschak, E. & Schneider, Gisbert, 2007. Kernel approach to molecular similarity based on iterative graph similarity. *Journal of chemical information and modeling*, 47(6), pp.2280–6. Available at: http://dx.doi.org/10.1021/ci700274r [Accessed March 2, 2013].

Sanchez-Pescador, R. et al., 1985. Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). *Science*, 227, pp.484–492.

Schechter, I. & Berger, A., 1967. On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications*, 27(2), pp.157–162. Available at: http://dx.doi.org/10.1016/S0006-291X(67)80055-X [Accessed April 12, 2013].

Schneider, G, Lee, M.L., et al., 2000. De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks. *Journal of computer-aided molecular design*, 14(5), pp.487–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10896320 [Accessed April 13, 2013].

Schneider, G, Clément-Chomienne, O., et al., 2000. Virtual Screening for Bioactive Molecules by Evolutionary De Novo Design Special thanks to Neil R. Taylor for his help in preparation of the manuscript. *Angewandte Chemie (International ed. in English)*, 39(22), pp.4130–4133. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11093229 [Accessed April 13, 2013].

Schneider, Gisbert, 2008. Evolutionary Molecular Design in Virtual Fitness Landscapes. In *Virtual Screening for Bioactive Molecules*. Wiley-VCH Verlag GmbH, pp. 161–186. Available at: http://dx.doi.org/10.1002/9783527613083.ch8.

Schneider, Gisbert & Baringhaus, K.-H., 2008. *Molecular Design: Concepts and Applications*, Wiley-VCH. Available at: http://www.amazon.com/Molecular-Design-Applications-Gisbert-Schneider/dp/3527314326 [Accessed April 13, 2013].

Schneider, Gisbert & Fechner, U., 2005. Computer-based de novo design of drug-like molecules. *Nature reviews. Drug discovery*, 4(8), pp.649–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16056391 [Accessed March 5, 2013].

Schneider, Jens & Kent, S.B.H., 1988. Enzymatic activity of a synthetic 99 residue protein corresponding to the putative HIV-1 protease. *Cell*, 54(3), pp.363–368. Available at: http://www.sciencedirect.com/science/article/pii/0092867488901997.

Schubert, U. et al., 1996. The two biological activities of human immunodeficiency virus type-1 Vpu protein involve two separable structural domains. *Journal of virology*, 70, pp.809–819.

Schubert-Zsilavecz, M. & Steinhilber, H.J.R.D., 2010. *Medizinische Chemie*, Deutscher Apotheker Vlg. Available at: http://www.amazon.com/Medizinische-Manfred-Schubert-Zsilavecz-Hermann-Steinhilber/dp/3769250028 [Accessed April 12, 2013].

Seelmeier, S. et al., 1988. Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. *Proceedings of the National Academy of Sciences of the*

United States of America, 85(18), pp.6612–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=282027&tool=pmcentrez&rendertype=abstract.

Sharp, Paul M & Hahn, Beatrice H, 2008. Prehistory of HIV-1. *Nature*, 455(October), pp.605–606.

Shaw, G.M. et al., 1984. Molecular characterization of human t-cell leukemia (lymphotropic) virus type III in the aquired immune deficiency syndrome. *Science*, (226), pp.1165–1171.

Shen, C.-H. et al., 2010. Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters. *The FEBS journal*, 277(18), pp.3699–714. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2975871&tool=pmcentrez&rendertype=abstract [Accessed March 22, 2012].

Shuman, C.F., Hämäläinen, M.D. & Danielson, U.H., 2004. Kinetic and thermodynamic characterization of HIV-1 protease inhibitors. *Journal of molecular recognition : JMR*, 17(2), pp.106–19. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15027031 [Accessed March 4, 2012].

Shuman, C.F., Vrang, L. & Danielson, U.H., 2004. Improved structure-activity relationship analysis of HIV-1 protease inhibitors using interaction kinetic data. *Journal of medicinal chemistry*, 47(24), pp.5953–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15537350.

Stahl, Martin et al., 2002. A validation study on the practical use of automated de novo design. *Journal of computer-aided molecular design*, 16(7), pp.459–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12510880 [Accessed April 13, 2013].

Todorov, N P & Dean, P M, 1998. A branch-and-bound method for optimal atom-type assignment in de novo ligand design. *Journal of computer-aided molecular design*, 12(4), pp.335–49. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9777492 [Accessed April 13, 2013].

Tschinke, V. & Cohen, N.C., 1993. The NEWLEAD program: a new method for the design of candidate structures from pharmacophoric hypotheses. *Journal of Medicinal Chemistry*, 36(24), pp.3863–3870. Available at: http://dx.doi.org/10.1021/jm00076a016 [Accessed April 12, 2013].

Turner, B.G. & Summers, M.F., 1999. Structural biology of HIV. *Journal of molecular biology*, 285(1), pp.1–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9878383.

Vinkers, H.M. et al., 2003. SYNOPSIS: SYNthesize and OPtimize System in Silico. *Journal of medicinal chemistry*, 46(13), pp.2765–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12801239 [Accessed April 9, 2013].

Wain-Hobson, S. et al., 1985. Nucleotide sequence of the AIDS virus, LAV. *Cell*, 40, pp.9–17.

Wang, L. et al., 2003. Yb(OTf) 3 -CatalyzedOne-Pot Synthesis of Quinazolin-4(3 H)-onesfrom Anthranilic Acid, Amines and Ortho Esters (or FormicAcid) in Solvent-Free Conditions. *Synthesis*, 2003(08), pp.1241–1247. Available at: https://www.thieme-connect.de/ejournals/pdf/10.1055/s-2003-39397.pdf [Accessed April 13, 2013].

Wang, Renxiao, Gao, Ying & Lai, Luhua, 2000. LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design. *Journal of Molecular Modeling*, 6(7-8), pp.498–516. Available at: http://link.springer.com/10.1007/s0089400060498 [Accessed April 13, 2013].

Waszkowycz, Bohdan et al., 1994. PRO_LIGAND: An Approach to de Novo Molecular Design. 2. Design of Novel Molecules from Molecular Field Analysis (MFA) Models and Pharmacophores. *Journal of Medicinal Chemistry*, 37(23), pp.3994–4002. Available at: http://dx.doi.org/10.1021/jm00049a019 [Accessed April 13, 2013].

Wei, P. et al., 1998. A novel CDK9-associated C-type Cyclin interactis directly with HIV-1 Tat and mediates its High Affinity, Loop-Specific Binding to TAR RNA. *Cell*, 92, pp.451–462.

Weiss, R.A., 1996. Retrovirus classification and cell interactions. *Journal of antimicrobial chemotherapy*, 37, pp.Suppl. B, 1–11.

Westhead, D R et al., 1995. PRO-LIGAND: an approach to de novo molecular design. 3. A genetic algorithm for structure refinement. *Journal of computer-aided molecular design*, 9(2), pp.139–48. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7608745 [Accessed April 13, 2013].

Whitcomb, J.M. & Hughes, S.H., 1992. Retroviral reverse transcription and integration: progress and problems. *Annual review of cell biology*, 8, pp.275–306. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1282352.

Wildman, S.A. & Crippen, G.M., 1999. Prediction of Physicochemical Parameters by Atomic Contributions. *Journal of Chemical Information and Modeling*, 39(5), pp.868–873. Available at: http://dx.doi.org/10.1021/ci990307l [Accessed April 13, 2013].

Wilk, T. et al., 2001. Organisation of immature human immunodeficiency virus type 1. *Journal of virology*, 75, pp.759–771.

Wlodawer, A. et al., 1989. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science (New York, N.Y.)*, 245(4918), pp.616–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2548279 [Accessed April 13, 2013].

Worobey, M. et al., 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature*, 455(7213), pp.661–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18833279 [Accessed March 14, 2012].

www.hivbook.com, 2011. *HIV Book 2011* 19th ed. C. Hoffmann & J. Rockstroh, eds., Hamburg: Medizin Fokus Verlag. Available at: www.hivbook.com [Accessed April 4, 2012].

Xie, D. et al., 1999. Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. *Cancer Research*, pp.1702–1707.

Yu, Z. et al., 2010. *Selective and facile assay of human immunodeficiency virus protease activity by a novel fluorogenic reaction*, Available at: http://journals.ohiolink.edu/ejc/article.cgi?issn=00032697&issue=v397i0002&article=197_safaohabanfr [Accessed April 13, 2013].

Zeldin, R.K. & Petruschke, R.A., 2004. Pharmacological and therapeutic properties of ritonavir-boosted protease inhibitor therapy in HIV-infected patients. *The Journal of antimicrobial chemotherapy*, 53(1), pp.4–9. Available at: http://jac.oxfordjournals.org/content/53/1/4.long [Accessed April 10, 2013].

Zhu, J., Yu, H., et al., 2001. Design of new selective inhibitors of cyclooxygenase-2 by dynamic assembly of molecular building blocks. *Journal of computer-aided molecular design*, 15(5), pp.447–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11394738 [Accessed April 13, 2013].

Zhu, J., Fan, H., et al., 2001. Structure-based ligand design for flexible proteins: application of new F-DycoBlock. *Journal of computer-aided molecular design*, 15(11), pp.979–96. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11989626 [Accessed April 13, 2013].

Ziegler, J.B. et al., 1985. Postnatal transmission of AIDS-associated retrovirus from mother to infant. *The Lancet*, 1, pp.896–898.

4.2 Figures

4.3 Tables

4.4 Abbreviations

ACN	Acetonitrile
AIDS	Aquired immune deficiency syndrome
AZT	Azidothymidine
CDC	Centre of disease control
CYP 450	Cytochrome P450
DMSO	Dimethylsulfoxide
DOGS	Design of genuine structures
EA	Evolutionary algorithm
FDA	Food and Drug Administration
FRET	Föster resonance energy transfer
GA	Genetic algorithm
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
HPLC	High pressure/performance liquid chromatography
HRMS	High resolution mass spectrometrie
HTS	High throughput screening
IN	Integrase
LAS	Lymphadenopathy syndrome
MMWR	Morbidity and Mortality Weekly Reviews
MOE	Molecular optimization environment
MS	Mass spectrometry/spectrometer
MW	Molecular weight
NMR	Nuclear magnetic resonance
NNRTI	Non nucleoside reverse transcriptase inhibitors
NRTI	Nucleoside reverse transcriptase inhibitors
PDB	Protein data base
PEP	Post exposure prophylaxis
PI	Protease inhibitors
PR	Protease
PSO	Particle swarm optimization
QSAR	Quantitative structure activity relationship

RPHPLC	Reverse phase HPLC
RT	Reverse transcriptase
SIV	Simian immunodeficiency syndrome
STD	Standarddeviation
TFA	Trifluoroacetic acid
WHO	World health organization

5 Curriculum Vitae

Name:	Birgit Zonsics
Date of birth:	11.01.1988
Martial status:	Single
Nationality:	Austrian
Parents:	Mag. Erich Zonsics
	Mag. Margit Zonsics

Education

2011-2012	Practical work for the diploma thesis at Eidgenössische Technische Hochschule (ETH) Zürich
2006-2013	Undergraduate student at University of Vienna
1998-2006	Secondary school at Gymnasium Sacré Coeur 1030 Wien
1995-1996	Primary school at École du Présidial in Limoges, 87000 France
1994-1998	Primary school Sankt Franziskus 1030 Wien

Summer schools

2009	TCM summer school in Chengdu, China with Prof. Kopp
2011	EuroPIN summer school of drug design in Vienna, Austria

Working experience

Jun. 2010 – Sep. 2011	Apotheke zum Goldenen Reichsapfel (Mag. Kowarik) 1010 Wien
Summer 2006, 2007, 2008	Regenbogenapotheke am Leberberg (Mag. Portisch) 1110 Wien