

Investigation of BCRP-inhibitors using QSAR and machine learning methods

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Federico Marighetti

aus

Padua (Italien)

Bonn, 2019

Prof. Dr. M. Wiese

Prof Dr. G. Bendas

Tag der Promotion: 23.07.2019

Erscheinungsjahr: 2019

"All models are wrong but some are useful"
George E. P. Box

Contents

1	Introduction	1
1.1	Multidrug resistance	1
1.2	ATP-binding cassette (ABC) transporters	2
1.2.1	P-glycoprotein (P-gp / ABCB1)	4
1.2.2	Multidrug Resistance associated Protein 1 (MRP1 / ABCC1)	8
1.2.3	Breast Cancer Resistance Protein (BCRP / ABCG2)	11
2	Materials and Methods (Molecular Modelling)	19
2.1	Descriptors	19
2.1.1	1D and 2D MOE descriptors	19
2.1.2	3D MOE descriptors	26
2.1.3	RECON/TAE descriptors	28
2.1.4	RDF descriptors	28
2.1.5	Inductive descriptors	29
2.1.6	Dragon descriptors	29
2.2	Machine learning methods	30
2.2.1	Self-Organizing Maps (SOMs)	30
2.2.2	Support vector Machine (SVM)	32
2.2.3	k-Nearest Neighbors (k-NN)	34
2.3	Analysis of data	34
2.4	3D QSAR	35
2.4.1	PLS analysis	37
2.4.2	Validation of 3D QSAR models	38
3	Materials and Methods (Biological testing)	41
3.1	Materials	41
3.1.1	Chemicals	41
3.1.2	Materials	42
3.1.3	Instruments	43
3.2	Cell culture	44
3.2.1	Thawing of cells	44
3.2.2	Freezing of cells	45

3.3	Cell lines	45
3.3.1	MCF-7 and MCF-7 MX cell lines	45
3.3.2	A2780 and A2780 Adr cell lines	45
3.3.3	MDCK II and MDCK II-BCRP cell lines	45
3.4	Preparation of the buffer solutions	46
3.4.1	Preparation of a 5x concentrated Krebs HEPES buffer	46
3.4.2	Preparation a the PBS buffer	46
3.5	Cell counting using the CASY [®] model TT	47
3.6	Accumulation Assays	48
3.6.1	Hoechst 33342 assay	48
3.6.2	Calcein AM assay	49
3.7	MTT assay	50
3.7.1	Analysis of the data using GraphPad Prism [®]	51
4	Investigation of the inhibitory activity of a new class of tariquidar analogues.	55
4.1	Optimization of the Hoechst 33342-assay	55
4.1.1	Optimization of the used Hoechst 33342 concentration	55
4.1.2	Optimization of the cell number.	58
4.1.3	Analysis of the raw data obtained from the Hoechst 33342 assay	61
4.2	Determination of the activity of a new class of BCRP-inhibitors.	62
4.2.1	Tariquidar analogues with modified anthranilic acid partial structure	62
4.2.2	Characterization of the inhibitory activity of the compound WK-X-27	63
4.2.3	Variations of the first linker	65
4.2.4	Variations of the second linker	69
4.2.5	Reduction of the scaffold size	73
4.2.6	Variations on the second aromatic ring	76
4.2.7	Variations on the third aromatic ring	80
4.2.8	Variations on the first aromatic ring	84
4.2.9	Variations of the linker position	91
4.2.10	Compounds containing a tetrahydroisoquinoline moiety	92
4.2.11	The effect of selected compounds on the cytotoxicity of mitoxantrone, SN-38 and Hoechst 33342	97
4.2.12	Determination of the cytotoxicity of selected BCRP inhibitors .	103
5	Development of binary QSAR models for classification of BCRP inhibitors.	111
5.1	Aim of this work	111

5.2	Preparation of the molecular structures and descriptor calculation . . .	112
5.2.1	Description of the used training dataset	112
5.2.2	Structure optimization	112
5.2.3	Calculation of molecular descriptors	114
5.3	Use of SOMs to discriminate between BCRP-inhibitors and non-inhibitors	115
5.3.1	Identification of the optimal parameters of the SOM	115
5.3.2	Analysis of results	116
5.4	Use of SVM to discriminate between BCRP-inhibitors and non-inhibitors	133
5.4.1	Generation of the models	133
5.4.2	Evaluation of the results	134
5.4.3	Use of an external testset	135
5.5	Use of k-NN to discriminate between BCRP-inhibitors and non-inhibitors	136
5.5.1	Generation of the models	136
5.5.2	Evaluation of the results	136
5.5.3	Use of an external testset	136
5.6	Comparison of the results	137
6	Structure-activity relationships of quinazolines derivatives as inhibitors of BCRP.	143
6.1	Structural and activity data	143
6.2	Free-Wilson analysis	145
6.3	Preparation of the dataset	150
6.4	Chemical descriptor based QSAR model	151
6.5	3D-QSAR analysis	152
6.6	Contour plot	157
7	Conclusions	161

Acknowledgements

Schreiben musste ich diese Arbeit zwar allein, doch an ihrem guten Gelingen beteiligt waren jede Menge mir nahestehender Menschen, denen ich Dank schulde.

An erster Stelle, und mögen mir andere verzeihen, muss mein Doktorvater Professor Dr. Wiese genannt werden. Zu jeder Uhrzeit war er für mich da, hat mich aufgebaut und mir immer geholfen, auch wenn es mal eine schwerere Phase gegeben hat. Das habe ich nie als selbstverständlich angesehen. Herzlichen Dank.

Ich danke Herrn Prof. Dr. Bendas für die hilfsbereite und wissenschaftliche Betreuung als Zweitgutachter.

Weiterhin möchte ich mich bei Frau Prof. Dr. König und Herr Prof. Dr. Wägele für die Mitwirkung in der Prüfungskommission bedanken.

Einen herzlichen Dank auch an Herrn Dr. Mathias Weigt. Besonders wertvoll war für mich, dass es ehrliche und offene Kritik gab, die mich immer wenn nötig, geerdet hat. Für seine Ehrlichkeit und die Kollegialität und die nette Zeit als Kollegen im 2. Semester bin ich sehr dankbar.

Das gilt natürlich auch für meine Familie. Wenn ich mit dem Kopf nicht zuhause, sondern bei der Arbeit war, ist sie es gewesen, die mich wachgerüttelt hat und mich dabei immer unterstützt und motiviert hat. Dieser Beistand ist für mich nicht gottgegeben, sondern den habe ich nur euch und eurer unendlichen Geduld zu verdanken. Eure Hilfe bedeutet mir mehr, als sich irgendjemand überhaupt vorstellen kann. Das war schon immer so und beschränkt sich nicht nur auf die Dissertation. Dafür bin ich euch auf ewig dankbar.

Darüber hinaus gilt mein Dank meiner Freundin Katleen, die mein Jammern und Klagen jahrelang ertragen und mich immer wieder aufgerichtet hat: Ohne ihre Unterstützung, hätte ich nie geschafft diese Arbeit zu abschließen.

Abbreviations

ABC transporter	Transporter ATP binding cassette transporter
ACC	Accuracy
ATP	Adenosin-5'-triphosphate
ATS	Autocorrelation of a topological structure
BCRP	Breast cancer resistance protein
Calcein-AM	Calcein acetoxymethyl ester
CoMFA	Comparative molecular field analysis
CoMSIA	Comparative molecular similarity indices analysis
CV	Cross-validation
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
FN	False negative
FP	False positive
HOMO	Highest occupied molecular orbital
HTS	High throughput screening
IC	Inhibitory concentration
k-NN	k-Nearest Neighbors
LMO cross-validation	Leave-many-out cross-validation
LOO cross-validation	Leave-one-out cross-validation
LUMO	Lowest unoccupied molecular orbital
MCF-7 cells	Michigan Cancer Foundation-7 cells
MDCK cells	Madin-Darby canine kidney cells
MDR	Multidrug resistance
MLR	Multiple linear regression
MOE	Molecular Operating Environment
MOPAC	Molecular Orbital PACkage
MRP1	Multidrug resistance associated protein 1
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MX	Mitoxantrone

Contents

NBD	Nucleotide binding domain
P-gp	P-glycoprotein
PBS buffer	Phosphate buffer saline buffer
PCA	Principal component analysis
PEOE	Partial equation of orbital electronegativities method
PLS	Patial least square
QSAR	Quantitative structure activity relationship
RDF	Radial distribution function
RMSD	Root-mean-square deviation
SD	Standard deviation
SOM	Self-Organizing Map
SVM	Support Vector Machine
TAE descriptors	Transferable atom equivalents descriptors
TMD	Transmembrane domain
TMH	Transmembrane spanning helix
TN	True negative
TP	True positive
VSA	van der Waals surface area
WHIM descriptors	Weighted Holistic Invariant Molecular descriptors

1 Introduction

1.1 Multidrug resistance

Cancer is one of the main causes of death in economically developed countries [1]. According to the data reported by the International Agency for Research on Cancer (IARC) for 2008, 12.7 million patients were affected by cancer and 7.6 million cancer deaths were registered worldwide [1, 2]. Those statistics makes cancer the second cause of death after cardiovascular diseases, which are accountable of 17.5 million deaths in 2005 [3].

The multidrug resistance (MDR) is defined as the ability of cells and organisms to resist treatment by structurally unrelated drugs [4] and it represents a critical problem in cancer therapy, especially in acute leukemias [5]. MDR is divided into two forms: the intrinsic drug resistance and the acquired drug resistance. The intrinsic drug resistance is related to the capacity of a tumor to show resistance against chemotherapeutic drugs already from beginning of the therapy. In acquired drug resistance, tumors respond to initial therapy, but later appear to be strongly resistant to the original treatment. The consequence of MDR is the failure of therapy and the intractability of the tumor with chemotherapeutic drugs [6].

Multidrug resistance is related to overexpression of certain ATP binding cassette (ABC) transporter proteins, a family of membrane transport proteins widely expressed in all living organisms. These membrane transporters are characterized by the combination of one or more transmembrane domains and cytoplasmic ATP binding domains [7].

1.2 ATP-binding cassette (ABC) transporters

ABC transporters are membrane proteins which are able to transport several kind of substrates, ranging from ions to small molecules, across the plasma membrane or other intracellular membranes [8]. The physiological roles of these transporters include the transport of metabolites and the protection of the organism from toxic compounds [9]. ABC transporters are also widely distributed in the organism and are expressed in a wide variety of tissues and organs. The tissue localizations of the major human ABC transporters involved in MDR are reported in table 1.1. Several diseases are associated with ABC transporters: The cystic fibrosis (ABCC7) and the Dubin-Johnson syndrome (ABCC2) are only two examples of them [9, 6].

Table 1.1: Distribution of the major human ABC transporters involved in MDR.

ABC transporter	Tissue localization
P-gp / ABCB1	kidney proximal tubule, placenta, adrenal cortex, testis, uterus, lymphocytes, hematopoietic cells [4]
MRP1 / ABCC1	lung, testis, kidney, placenta, skeletal and cardiac muscles [10]
BCRP / ABCG2	brain (BBB), liver, kidney, lactating breast, placenta, testis, small intestine, colon, haematopoietic compartment [11]

At the moment, 48 members of the ABC transporter family are known in humans [5] and are divided into seven subfamilies from ABCA to ABCG, according to their gene structure, sequence similarity and phylogenetic analysis [8, 12, 13]. Among these 48 known humans members of the ABC transporter family, P-glycoprotein (P-gp / ABCB1), the multidrug resistance associated protein 1 (MRP1 / ABCC1) and the breast cancer resistance protein (BCRP / ABCG2) have been recognized to play an important role in MDR [7, 14].

A functional ABC transporter requires the presence of at least two nucleotide binding domains (NBDs) and two transmembrane domains (TMDs) [15]. The nucleotide binding domain of ABC transporters is highly conserved for all members of this family. In particular, four conserved sequence elements and three highly conserved residues are present with minimal variations in all members of the family. The conserved sequence elements consist of the Walker A, the Walker B, the ABC signature sequence, and the D-Loop, while the highly conserved residues are the Q-loop, the A-loop and the H-loop. [16, 17, 18].

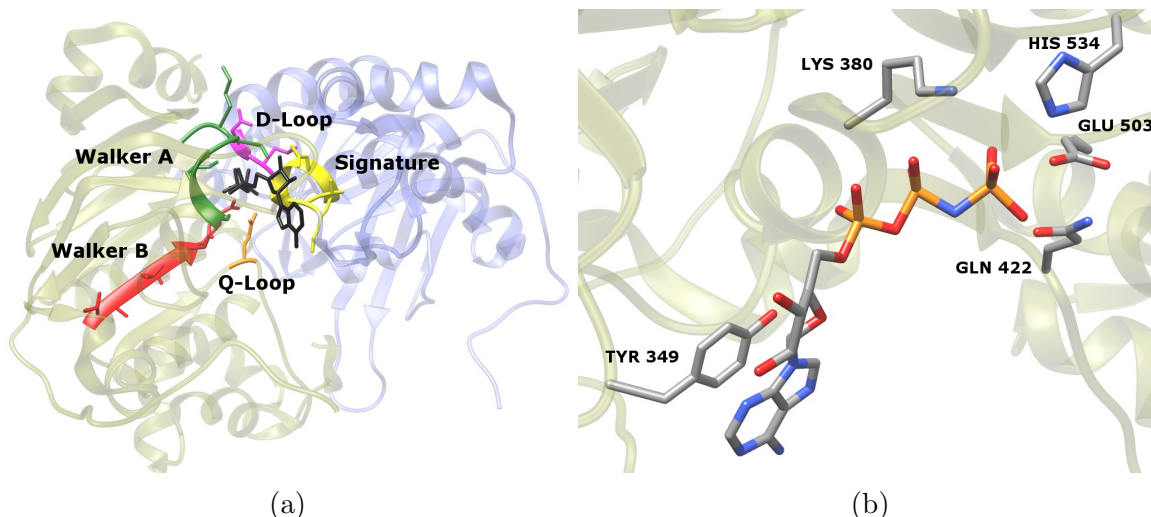


Figure 1.1: (a) Representation of the two NBDs of Sav1866 (PDB-entry 2ONJ). The highly conserved motifs are colored in green (Walker A), red (Walker B), orange (Q-loop), cyan (D-loop) and yellow (signature sequence). The molecular structure of AMP-PNP is colored in black. (b) The interaction of AMP-PNP with the A-loop (Tyr349) the H-loop (His534), the Q-loop (Gln422), the Walker B (Glu503) and the Walker A motif (Lys380) of the bacterial transporter Sav1866. Figures are adopted from the work of Seeger *et al.* [16].

The Walker A and the signature sequence are in contact with the γ -phosphate moiety of the nucleotide, while the A-loop contacts its adenine; the Walker B contains a highly conserved glutamate residue, which is probably involved in the nucleophilic attack on ATP through a water molecule; the Q-loop interacts with the γ -phosphate of the nucleotide via a water molecule and is also involved in the interaction between NBD and TMD; the D-loop could play a role in the communication between the two NBDs and the H-loop has a direct interaction with the γ -phosphate of the nucleotide and is directly involved in the catalytic hydrolysis of the nucleotide [16, 18]. In figure 1.1a are shown the two nucleotide binding domains of the bacterial ABC transporter Sav1866 [19], where the Walker A, the Walker B, the signature sequence, the D-loop and the Q-loop are highlighted in different colors. The 5'-adenylylimidodiphosphate (AMP-PNP), a synthetically prepared non-hydrolyzable analog of adenosine 5'-triphosphate that was co-crystallized with the protein, is presented in black. In figure 1.1b are presented the principal amino acids that interact with the AMP-PNP in Sav1866. In this figure, it is possible to recognize the interaction of the A-loop (Tyr349) with the adenine of the nucleotide and the interactions of the H-loop (His534), of the Q-loop (Gln422), of the Walker B (Glu503) and of the Walker A (Lys380) with the γ -phosphate of the nucleotide.

In contrast to the NBDs, the transmembrane domains of ABC transporters show a low degree of homology. The ABC exporters normally feature two membrane spanning domains (MSDs) with similar amino acid sequences interacting together to form a

1 Introduction

dimer interface. Each membrane spanning domain generally consists of six transmembrane spanning helices (TMHs), with a total of 12 transmembrane spanning helices [16]. An important exception is the human ABC transporter BCRP, which consists of two monomers with 655 amino acids each containing only one MSD and one NBD. These two identical monomers interact together probably through disulfide bonds [20] to form the complete BCRP structure [21].

1.2.1 P-glycoprotein (P-gp / ABCB1)

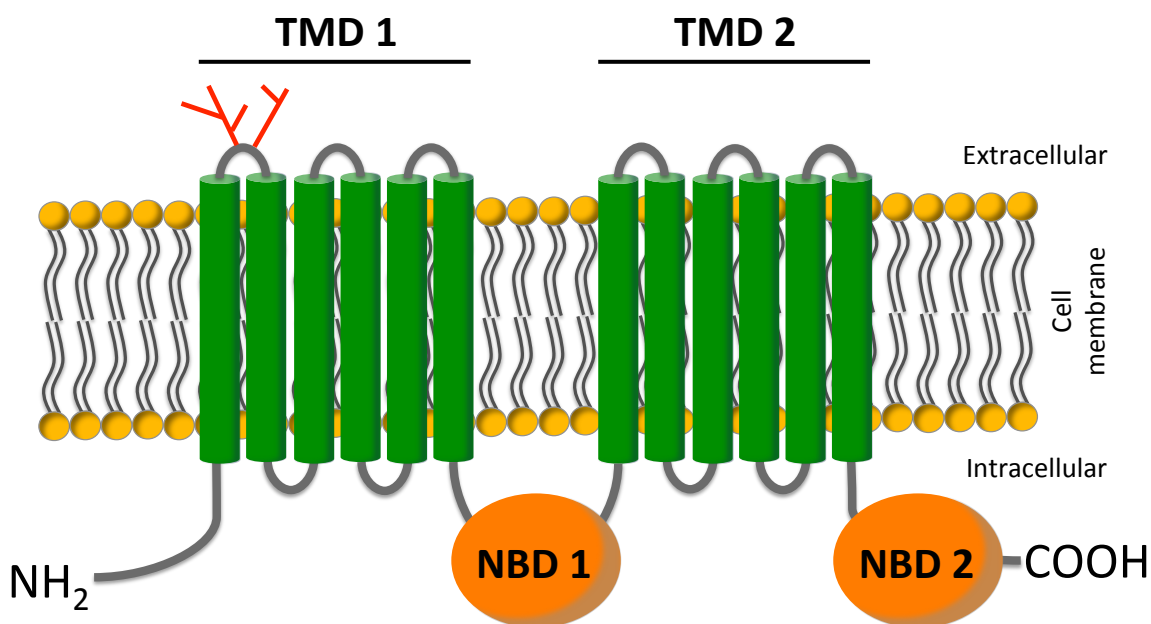


Figure 1.2: Topological model of P-glycoprotein.

P-glycoprotein (P-gp / ABCB1) also known as MDR1 is a member of the ABCB subfamily. It was discovered in 1974 by Juliano and Ling in Chinese hamster ovary cells mutants [22, 23]. Human P-gp is a 1280 amino acids protein, codified by a gene located on the chromosome 7q21 [7]. The molecular weight of human P-gp is about 170 kDa in fully glycosylated form [24]. As illustrated in figure 1.2, P-gp is characterized by two membrane spanning domains (MSDs) and two nucleotide binding domains (NBDs) both on the same amino acidic sequence. Observing the primary sequence from NH₂ to -COOH-termini, the domains appear following the pattern MSD1-NBD1-MSD2-NBD2. The -NH₂ and -COOH-termini as well as the nucleotide binding domains are located in the intracellular space. The first extracellular loop is N-glycosylated [25].

The first information about the three-dimensional structure of P-gp was obtained using electron cryomicroscopy. This preliminary investigation provided evidence for a dimeric association of P-gp in the reconstituted lipid bilayer [26]. Homology models of P-gp, first based on the structure of MsbA [27] and later based on Sav1866 [28] were generated in order to identify and analyze the structure of this transporter and in particular its drug-binding site(s). The X-ray structure of mouse P-gp, that shares 87 % of identity with the human P-gp, was finally obtained at 3.8 Å resolution in 2009 by Aller *et al.* [29]. This structure proved to be wrong and was successively corrected by Jin *et al.* [30, 31]. The obtained structure seems to confirm the hypothesis that this transporter possesses at least two drug-binding sites located in the transmembrane region. This hypothesis was already suggested by Shapiro and Ling more than ten years before [32].

Transport mechanism of P-gp

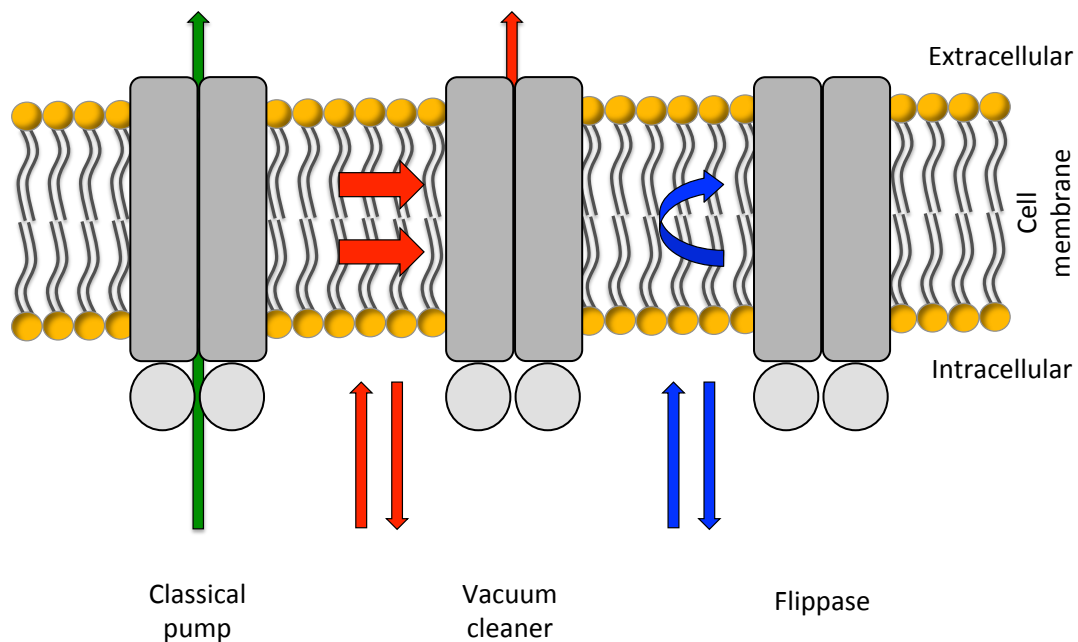


Figure 1.3: The classical pump, vacuum cleaner, and flippase models for substrate transport by P-gp, modified from [33].

The efflux mechanism of P-gp and, in more general, of ABC transporters is not yet clearly explained. Until now, three different models of substrate transport mediated by P-gp were suggested. The first and the most accepted hypothesis about substrate

1 Introduction

transport mediated by P-gp was suggested by Gottesman *et al.*. The results suggested that P-gp works as a vacuum cleaner for hydrophobic molecules, which are embedded in the membrane [34]. Two years later Higgins and Gottesman formulated another theory how substrate transport is mediated by P-gp: They considered P-gp working as a flippase. In this transport model the substrate interacts with the lipid bilayer and enters into the core of the membrane-associated transporter in contact with the inner leaflet of the membrane and either is pumped directly from the inner leaflet of the bilayer to the extracellular space or is flipped from the inner leaflet to the outer leaflet of the membrane. In this second case, the substrate successively diffuses from the outer leaflet to the extracellular space [35]. The third proposed model of transport was suggested in 1994 by Altenberg *et al.* on the basis of data obtained by observations of the transport of rhodamine 123 mediated by P-gp. Altenberg *et al.* hypothesized that P-gp extracts the substrate not from the membrane, as postulated by the flippase and by the vacuum cleaner theories, but directly from the cytosol [36]. The three proposed transport models are summarized in figure 1.3.

P-gp substrates

P-gp substrates vary enormously in size and structure, ranging from small organic cations to large molecules, like actinomycin D [37]. The first pharmacophore model that describes the P-gp substrates was proposed by Seelig *et al.* [38]. In this work was reported that many P-gp substrates have two or three electron-donor groups in their structure, separated by 2.5 Å from each other and separated by 4.6 Å from the other electron-donor group pair. Cianchetta *et al.* [39] suggested a pharmacophore model with two hydrophobic groups 16.5 Å apart and two hydrogen-bond-acceptor groups 11.5 Å apart. The molecular dimension was also found to play an important role in recognition as P-gp substrate. Support Vector Machine was used by Huang *et al.* to discriminate between P-gp substrates and non-substrates, with an accuracy of higher than 90 % [40]. Other studies reported that some chemical characteristics are important but not essential to be a P-gp substrate. Those characteristics include a logP value higher than 2.92, a molecular axis longer than 18 atoms, a high value of E_{homo} , the presence of at least one tertiary basic nitrogen atom, and molecular weight lower than 800 Dalton [25]. Today, many drugs are recognized to be P-gp substrates and belong to several drug classes such anticancers, cardiacs, antimicrobials, immunosuppressants, gastrointestinal agents, HIV proteases, and neurologic agents [41]. A short list of well-known P-gp substrates is given in table 1.2.

Table 1.2: P-gp substrates.

Drug class	P-gp substrates
Anticancers	actinomycin D, colchicine, daunorubicin, doxorubicin, etoposide, imatinib, irinotecan, lapatinib, mitomycin C, nilotinib, paclitaxel, taxol, topotecan, vinblastine, vincristine
Cardiacs	antiarrhythmic agents (digoxin and verapamil), Anticoagulant agents (warfarin), Antihypertensive agents (diltiazem, losartan, propranolol), Antiplatelet agents (clopidogrel, ticagrelor), Statins (atorvastatin, lovastatin)
Antimicrobials	erythromycin, ivermectin, posaconazol, quinolones, rifampicin
Immunosuppressants	cyclosporine A, everolimus, tacrolimus
Gastrointestinal agents	cimetidine, domperidone, loperamide, ondansetron
Fluorescent dyes	calcein AM, BCECF AM, rhodamine 123, Hoechst 33342

P-gp inhibitors

P-gp inhibitors are able to reverse MDR mediated by P-gp. Like substrates, they are structurally diverse. However, the majority of them are weakly amphipathic and hydrophobic and often they contain a tertiary nitrogen atom. Many P-gp inhibitors are also themselves transported by P-gp [42]. Inhibitors of P-gp are classified into three generations according to their specificity, affinity, and toxicity. First generation inhibitors are drugs normally used in pharmacological therapies that are also able to inhibit P-gp. The usage of these compounds as P-gp inhibitors is limited due to their low potency and high toxicity. Second generation inhibitors increase the inhibitory potency against P-gp, and lack the pharmacological activities against other targets. PSC 833 (a non-immunosuppressive analogue of cyclosporine A) and dexverapamil are two members of this inhibitor generation. These compounds, however, are also inhibitors and substrates of CYP3A4 enzyme and other ABC transporters. These additional pharmacological effects complicate the pharmacokinetic of the administered cytostatics. Third generation P-gp inhibitors are potent and selective P-gp inhibitors with low toxicity, developed using structure activity relationship methods. XR 9576 also known as tariquidar is a member of this inhibitors generation [43]. A list of P-gp inhibitors divided on the base of their generation is given in table 1.3.

Table 1.3: P-gp inhibitors categorized on the basis of their generation. The table is based on Amin *et al.* [43].

Generation	Inhibitors	Specificity and limitations
First generation	verapamil, cyclosporine A, reserpine, quinidine, yohimbine, tamoxifen, toremifena	Non-selective, partially also P-gp substrates, pharmacologically active
Second generation	dexverapamil, dexniguldipine, valsopodar (PSC833), dofequidar fumarate (MS-209)	Increased specificity with respect to first generation inhibitors, also substrates and inhibitors of CYP3A4 enzyme and other ABC transporters
Third generation	zosuquidar (LY335979), ianiquidar (R101933), mitotane (NSC-38721), biricodar (VX-710), elacridar (GF 120918), ONT-093, tariquidar (XR 9576), and HM30181	Potent and selective P-gp inhibitors, no limitations

1.2.2 Multidrug Resistance associated Protein 1 (MRP1 / ABCC1)

Multidrug Resistance associated Protein 1 (MRP1) is the first member of the ABCC family. MRP1 is expressed in most tissues throughout the body, especially in lung, testis, kidneys, skeletal muscle, and peripheral blood mononuclear cells [44]. Also MRP1 is involved in multidrug resistance and is overexpressed in many solid tumors such hepatocellular carcinoma and breast cancer [45]. MRP1 was discovered in H69AR cells, obtained from the lung cancer cell line H69 by stepwise selection in media containing doxorubicin [10]. The human MRP1 gene is localized on the gene 16p13.1 [46]. This transporter consists of 1,531 amino acids, with a molecular weight of about 190 kDa, and contains three membrane spanning domains and two nucleotide binding domains (MSD0-MSD1-NBD1-MSD2-NBD2). The first membrane spanning domain contains only five transmembrane helices, while each of the other two membrane spanning domains contain six transmembrane helices as all the other ABC transporters [44]. Unlike P-gp and BCRP, the -NH₂ terminus is extracellular. MRP1 presents two sites of N-glycosilation, localized near to the -NH₂ terminus and on the MSD2, involving the residues Asn19, Asn23 (-NH₂ terminus) and Asn1006 (MSD2) [47]. The topology of MRP1 is shown in figure 1.4.

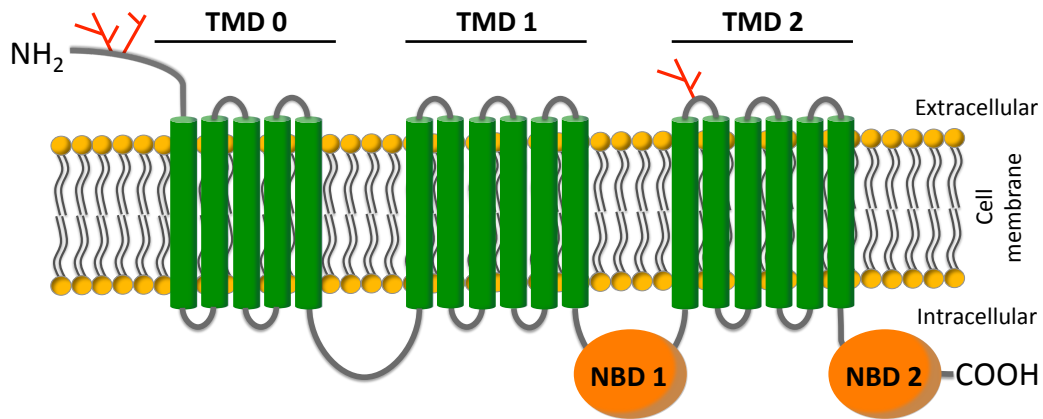


Figure 1.4: Topological model of Multidrug Resistance associated Protein 1.

MRP1 substrates

MRP1 substrates include a wide variety of anticancer drugs like doxorubicin, daunomycin, epirubicin, menogaril, colchicine, vincristine, vinblastine, etoposide, gramicidin D, idarubicin and heavy metal ions, such as arsenite, arsenate and antimonite [10]. Other xenobiotics transported by MRP1 are antivirals like saquinavir and ritonavir, and antibiotics like difloxacin and grepafloxacin [48]. In contrast to P-gp, MRP1 is able to actively transport GSH-, glucuronate-, and sulfate-conjugated organic anions. This active efflux prevents the accumulation of conjugated toxic compounds, which could react back to the parental compound either spontaneously or by enzymatic reaction [44]. Fluorescent substrates transported by MRP1 include calcein AM and BCECF AM. Concerning a common pharmacophore for MRP1 substrates, it is only known that the majority of them contain an hydrophilic and an hydrophobic portion [10]. Some MRP1 substrates are listed in table 1.4.

Table 1.4: MRP1 substrates.

Drug class	Substrates
Folate-based antimetabolites	methotrexate
Anthracyclines	doxorubicin
Plant alkaloids	etoposide, vincristine, vinblastine, paclitaxel, irinotecan
Antiandrogens	flutamide, hydroxyflutamide
Antivirals	saquinavir, ritonavir
Antibiotics	difloxacin, grepafloxacin
Metalloids	arsenate, antimoniate
Toxicants	aflatoxin B1, methoxychlor
GSH- Glucuronide- and sulphate- conjugates	etoposide-gluc, SN-38-gluc, 2,4-Dinitrophenyl-SG, doxorubicin-SG, cyclophosphamide-SG, atraxine-SG, aflatoxin B ₁ -epoxide-SG, 4-nitroquinoline 1-oxide-SG, 14 β -estradiol-17- β -D-gluc, glucuronosylbilirubin, leucotriene-C ₄ , prostaglandin A ₂ -SG, 15-Deoxy- Δ^{12-14} prostaglandin J ₂ -SG, estrone 3-sulphate, sulphatolithocholate
Folates	folic acid, L-leucovorin
Other metabolites	GSSG, GSH, bilirubin
Fluorescent substrates	calcein AM, BCECF AM

MRP1 inhibitors

MRP1 inhibitors could be divided into non-specific MRP1 inhibitors and specific MRP1 inhibitors. The non-specific inhibitors include indomethacin, probenecid, P-gp inhibitors like verapamil and cyclosporine A, tricyclic isoxazoles like LY402913 and LY465806, and several flavonoids [49]. Until today, the only specific MRP1 inhibitors discovered are the leukotriene receptor antagonists MK571, ONO-1078 [50, 51], and the 2-thioureidobenzo[*b*]thiophene-3-carboxylic acid derivatives synthesized and evaluated by Häcker *et al.* [52]. The pharmacophore model for MRP1 inhibitors proposed by Pajeva and Wiese [53] consists of at least one hydrogen bond acceptor group of limited flexibility, an additional hydrophobic center in a flexible side chain, which is connected to a planar hydrophobic ring, and a tertiary protonated nitrogen. Those features have to be at a distance of about 5 Å.

1.2.3 Breast Cancer Resistance Protein (BCRP / ABCG2)

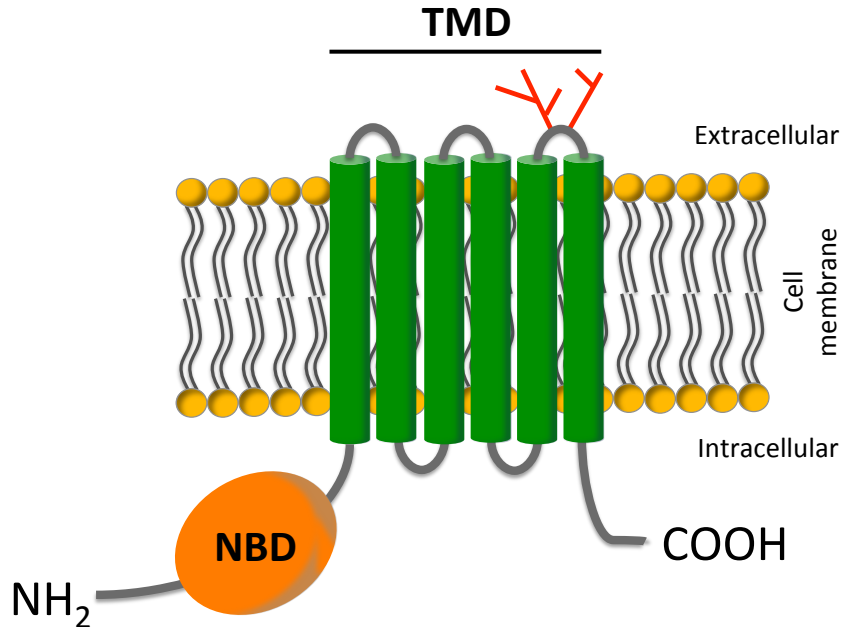


Figure 1.5: Topological model of Breast Cancer Resistance Protein.

Breast Cancer Resistance Protein (BCRP / ABCG2) is the second member of the ABCG family. This transporter was discovered by Doyle *et al.* in 1998 from highly doxorubicin resistant cells MCF-7/AdrVp [54]. The name breast cancer resistant protein was given due to the origin of the MCF-7 cell line (breast cancer cells) [55]. The human BCRP gene is located on chromosome 4q22 and codifies a 655 amino acid protein with molecular weight of about 72 kDa [54, 55]. In contrast to other ABC transporters, the primary sequence of BCRP and of all members of the ABCG family contains only one N-terminal NBD and one C-terminal TMD. The most accepted hypothesis suggested that BCRP could work as an homodimer, with two single BCRP monomers connected each other with disulfide bridges that involve the cysteine 603 residue [56, 20]. An intramolecular disulfide bridge seems to be formed between the two cysteine residues Cys-592 and Cys-608., located in the third extracellular loop [57]. The transmembrane domain (residues $\sim 394 - 655$) seems to consists of six transmembrane spanning helices, according to the hydropathy analysis of the primary sequence, with predictions based on the Hidden Markov Model algorithm (HMM), and immunofluorescence experiments [21]. In the third extracellular loop are also located two asparagines at positions 557 and 596, which are possible sites of N-glycosilation

1 Introduction

[58]. The topology of the breast cancer resistance protein is shown in figure 1.5. To date, there are not crystal structures of BCRP available, therefore its detailed structure remains unknown. The 3D structural analysis of BCRP was performed for the first time in 2006 by McDevitt *et al.* [59]. His research group obtained the 3D structures of the transporter at a resolution of about 18 Å using cryonegative stain electron microscopy. They interpreted the observed tetrameric complex to be comprised of four homodimeric BCRP complexes. Unfortunately, homology modeling is also not helpful to understand the three-dimensional structure of BCRP, due to the very low identity values between the primary sequence of this transporter and that of the other ABC transporters for which crystal structures are available [60].

Localization of BCRP in human tissues

BCRP is present in a wide variety of human tissues where it plays an important role in protection against xenobiotics. It has been identified in the apical membrane of placental syncytiotrophoblasts, in the bile canalicular membrane of hepatocytes, in the luminal membranes of villous epithelial cells in the small intestine and colon, in cardiac muscle, in the endocrine cells of the pancreas, in sebaceous glands, in the endothelium of the nervous system, in the zona reticularis of the adrenal gland, in the Sertoli cells of the testis, and in the venous and capillary endothelial cells of almost all tissues [61, 62]. Expression of BCRP was also observed in stem cells. Its role in stem cells differentiation and protection against xenobiotics is object of investigations [63].

Physiological role of BCRP

BCRP is recognized to be important for protection against xenobiotics and in regulation of oral bioavailability. This transporter seems to be also involved in other physiological roles, as the extrusion of porphyrins (conjugated and non-conjugated) from hematopoietic cells or liver and the harderian gland, and the secretion of vitamins into breast milk [64]. BCRP is believed to protect the fetus from toxins as well as to remove toxins from the fetal space. Presence of BCRP in intestine limits the oral bioavailability of xenobiotics. The role of BCRP in kidney is supposed to be related to renal drug excretion [62]. The physiological roles of BCRP in human tissues are summarized in table 1.5.

Table 1.5: Physiological role of BCRP in human tissues [61].

Tissue	Physiological role
Placenta	Protection of the fetus
Liver	Hepato-biliary excretion
Stem cells	SP-phenotype and protection against hypoxia
Intestine	Reduction of xenobiotics absorption
Brain	Protection against xenobiotics
Breast	Transport of vitamins into breast milk [64]
Kidney	Urinary excretion of drugs

BCRP substrates

Due to its protective role in human organism, BCRP is able to transport a wide variety of substrates with very different structures. Most cell lines over expressing BCRP are highly resistant against mitoxantrone, even if they were never treated with mitoxantrone. Furthermore, the accumulation of mitoxantrone in cells transfected with BCRP cDNA is significantly lower than in cells transfected with empty vectors. These data suggest that BCRP is able to transport mitoxantrone with high efficiency [65, 55].

Derivatives including topotecan, irinotecan and SN-38 (the active metabolite of irinotecan) are also important substrates of BCRP. The accumulation of camptothecin derivatives is also low in cells not previously selected with topotecan and in cells transfected with BCRP cDNA [66].

Flavopiridol is a cyclin-dependent kinase inhibitor under clinical trials. Robey *et al.* [67] exposed the human breast cancer cell line MCF-7 with incrementally increasing concentrations of flavopiridol. The resulting resistant cell line MCF-7 FLV1000 resulted to be resistant not only against flavopiridol, but also against mitoxantrone and topotecan.

BCRP-overexpressing cell lines MCF-7/AdrVp3000 and S1-M1-80 are reported to be highly resistant against anthracyclines, while the other BCRP-overexpressing cell lines are not [68]. Further investigations revealed that these two cell lines express two different mutants of BCRP, the mutant R482T (Thr at position 482) in the MCF-7/AdrVp3000 cell line and the mutant R482G (Gly at position 482) in the S1-M1-80 cell line. BCRP with Arg at position 482 is considered as wild-type.

The antifolate drug methotrexate seems to be transported by the wild-type BCRP, while it is not a substrate of BCRP mutants. Chen *et al.* [69] reported that methotrexate and methotrexate polyglutamates are transported by wild-type BCRP and also that this transporter has a high-capacity but low-affinity for these substrates.

Several fluorescent dyes like Hoechst33342, BODIPY-prazosine, BCECF-AM and

pheophorbide A has been reported to be BCRP substrates [55, 70, 71, 72].

The BCRP substrates are summarized in table 1.6. In figure 1.6 the structures of selected BCRP substrates are given.

Table 1.6: BCRP substrates (Based on Mao *et al.* [65]).

Drug class	Substrate
Anthracyclines ^a	daunorubicin, doxorubicin
Anthracenes	mitoxantrone, bisantrene, aza-anthrapyrazole
Camptothecin derivates	topotecan, SN-38, 9-amino-camptothecin, irinotecan, diflomotecan
Antifolate ^b	methotrexate, methotrexate polyglutamate
Nucleoside analogs	AZT, AZT 5'-monophosphate, lamivudine
Conjugates	estrone-3-sulfate, 4-MUS, E3040S, TLC-S, 4-MUG, E3040G, E ₂ 17 β G, DNP-SG
Other drugs	prazosin, indolocarbazole, popoisomerase I inhibitors (NB-506, J-107088), flavopiridol, ErbB1 tyrosine kinase inhibitor (CI1033), imatinib mesylate (STI1571), pantoprazole
Fluorescent dyes	Hoechst 33342, BODIPY-prazosine, BCECF-AM, pheophorbide A

^aSubstrate only of the BCRP mutants R482T and R482G.

^bSubstrate only of the wild-type BCRP.

BCRP inhibitors

BCRP inhibitors are of interest as chemosensitizers for clinical drug resistance and for improving the pharmacokinetics of chemotherapeutic drugs [73]. Today, several BCRP inhibitors are known and they may be classified into four categories [74]:

- BCRP specific inhibitors
- Broad spectrum inhibitors
- Flavonoids and derivatives
- Inhibitors of other targets that are also BCRP-inhibitors (e.g. tyrosine kinase inhibitors (TKIs) and HIV protease inhibitors)

Fumitremorgin C, a toxin extracted from *Aspergillus fumigatus*, is the first identified specific BCRP inhibitor [66]. Due to its high neurotoxicity in mice, the use *in vivo* is not practicable. Tryprostatine A, a further *A. fumigatus* secondary metabolite was also recognized to be a selective BCRP inhibitor. Contrary to fumitremorgin

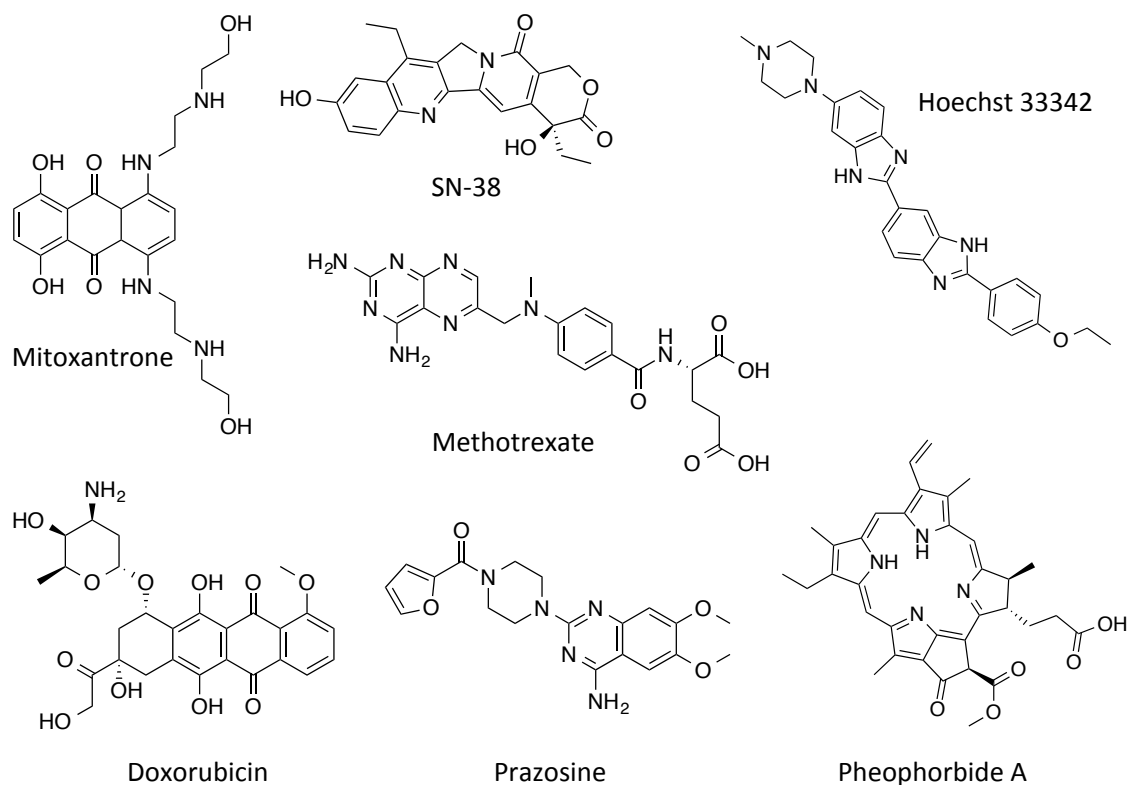


Figure 1.6: Structures of selected BCRP substrates.

C, tryprostatin A does not show any cytotoxicity at effective concentrations but its low inhibitory potency limits the possibilities of application in clinical therapy [75]. The new BCRP selective inhibitor Ko143, a synthetic fumitremorgin C analogue, was developed in order to decrease the cytotoxicity of fumitremorgin C conserving its inhibitory activity and selectivity for BCRP [73]. Novobiocin, a coumermycin antibiotic, was also recognized to be a non-toxic and selective BCRP inhibitor. Its inhibitory potency is however very low: indeed, novobiocin has an IC_{50} value of only about $50 \mu\text{M}$ [76].

The broad spectrum BCRP inhibitors include the high affinity third generation P-gp inhibitors elacridar (GF 120918) and tariquidar (XR 9576) [77, 78]. Both compounds are able to reverse the BCRP mediated resistance in MCF-7 MX cells with IC_{50} values barely lower than $1 \mu\text{M}$. An other broad spectrum BCRP inhibitor is the pipercolinate derivative VX-710 (biricodar), which is able to inhibit only the wild-type BCRP, but not the mutant R482T. Its inhibitory activity is however lower than of fumitremorgin C [79]. Calcium antagonists like nifedipine and nifedipine are also able to non-selectively inhibit BCRP. The best BCRP inhibitor among them, nifedipine, is able to inhibit BCRP with an IC_{50} value of $5 \mu\text{M}$ in HeLa/SN100 cells [80].

Flavonoids have been found to be also non-selective BCRP inhibitors. Morris *et al.*

1 Introduction

[81] evaluated the effect of 20 flavonoids on the cellular accumulation and cytotoxicity in the resistant MCF-7 MX cell line and in the parental MCF-7 cell line. They have reported that the two most potent tested compounds, chrysin and biochanin A were able to increase in MCF-7 MX cells the mitoxantrone accumulation at concentrations of 0.5 and 1.0 μM and the mitoxantrone toxicity at a concentration of 2.5 μM . Phytoestrogens like genistein, naringenin, acacetin, and kaempferol were reported as BCRP inhibitors by Imai *et al.* [82]. In this work it has been also reported that genistein is also a substrate of BCRP and its transport mediated by BCRP in BCRP-transduced LLC-PK1 cells is abolished by fumitremorgin C. In the last years, several studies regarding the interaction of natural flavonoids, synthetic flavonoids derivatives, and other compounds with structures derived from the flavonoid structure with BCRP were performed [83, 84, 85, 86]. The structure-activity relationship of flavonoids as BCRP inhibitors [87] has revealed the important contribution of some structural features to BCRP inhibition. Finally, Juvalle *et al.* [88] have synthesized a 7,8-benzoflavone derivative able to selectively inhibit BCRP, presenting also low cytotoxicity at higher concentrations.

Canetinib (CI-1033, PD 183805) was the first Tyrosine Kinase Inhibitor (TKI) reported to reverse the resistance against SN-38 of BCRP expressing cells [89]. Since this discovery, other tyrosine kinase inhibitors like gefitinib [90] and imatinib mesylate [91] were reported to be also BCRP inhibitors. Further investigations revealed also that imatinib mesylate is not only inhibitor but also substrate of BCRP and its transport mediated by this transporter could be blocked by addition of the specific BCRP inhibitor Ko143 [92].

HIV protease inhibitors were also reported to be BCRP inhibitors by Gupta *et al.* [93]. In particular, ritonavir, saquinavir, and nelfinavir inhibit wild-type BCRP (R482) with IC_{50} values between about 4 and 20 μM . They are also able to inhibit the two mutants R482T and R482G with IC_{50} values about 2-times greater than for the wild-type BCRP. However, HIV protease inhibitors are not specific BCRP inhibitors, but they are also able to inhibit other ABC transporters involved in MDR [94]. The structures of selected BCRP inhibitors are given in figure 1.7.

The first pharmacophore model for BCRP inhibitors was proposed by Chang *et al.* [95]. It was built using four potent BCRP inhibitors (GF120918, Ko143, nelfinavir, and nifedipine) and consists of three hydrogen bond acceptors and three hydrophobic features. Another pharmacophore model was more recently proposed by Matsson *et al.* [96]. It consists of only two hydrophobic centers and one hydrogen bond acceptor feature. The two hydrophobic centers are separated by a distance of 6.75 Å and they are separated from the hydrogen bond acceptor feature by 9.84 Å and 3.47 Å.

1.2 ATP-binding cassette (ABC) transporters

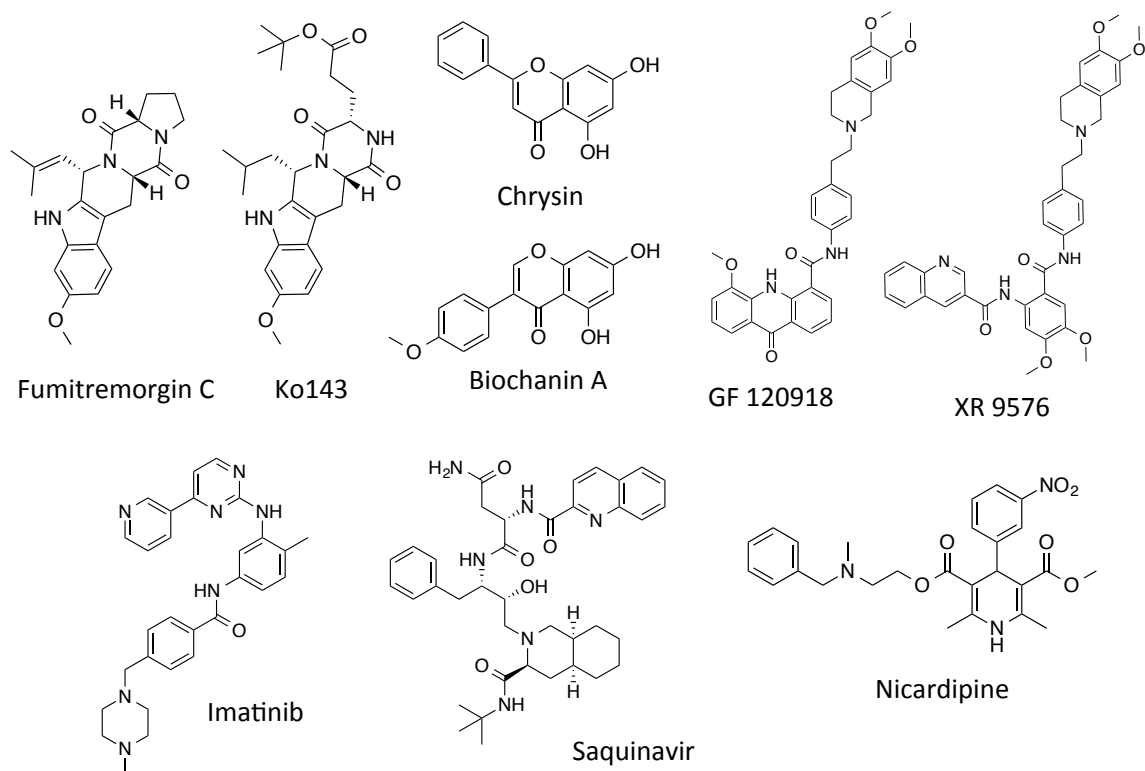


Figure 1.7: Structures of selected BCRP inhibitors.

2 Materials and Methods (Molecular Modelling)

2.1 Descriptors

The purpose of QSAR descriptors is to represent physical properties of molecules by numerical values. These numbers can be successively used in mathematical models to predict chemical or biological properties of codified molecules. In the present work, several molecular descriptors were calculated using the software Molecular Operating Environment (MOE). The descriptors treated in this chapter are divided into classes, according to their calculation and physical meaning. The following classification and description of the used molecular descriptors is based on the manual of the software MOE [97] and on the book *Molecular Descriptors for Chemoinformatics* by Todeschini and Consonni [98].

2.1.1 1D and 2D MOE descriptors

Partial Charge Descriptors

Partial charge descriptors are calculated using partial charges of each atom of the molecule. This class of descriptors is divided into two subgroups, according to the method used to calculate the partial charges. The first subgroup includes descriptors that are calculated using the partial equalization of orbital electronegativities (PEOE) method. This method was developed by Gasteiger in 1979 [99]. It consists of calculating the amount of charge transferred ($q^{<\alpha>}$) between two bonded atoms (with $\chi_B > \chi_A$) until equilibrium. The convergence is guaranteed by an exponentially decreasing scale factor, that damps down the amount of charge transferred at each iteration. The amount of charge transferred at each iteration is calculated as:

$$q^{<\alpha>} = \frac{\chi_A - \chi_B}{\chi_A^+} 0.5^\alpha$$

where χ_A^+ is the electronegativity of the positive ion of atom A, χ_A is the electronegativity of atom A, that depends quadratically on partial charge, χ_B is the

electronegativity of atom B and α (with $\alpha = 1, 2, 3, \dots, \infty$) is the number of iteration steps. The use of a damping factor assures the reaching of convergence. The total charge of an atom is calculated as the sum of the single charge transfers [99].

Descriptors using the PEOE charges are indicated in MOE as *PEOE_x* charge descriptors. Descriptors using charges calculated by an external program are indicated as *Q_x* charge descriptors.

Physical properties

Table 2.1: Physical properties.

Code	Descriptor name
apol	Sum of the atomic polarizabilities
bpol	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule. This descriptor is calculated as follows: $\sum_i p_i - p_j $, where p_i and p_j are the polarizabilities of two bonded atoms of a molecule
FCharge	Total charge of the molecule
mr	Molecular refractivity calculated from a 11 descriptors linear model [97]
SMR	Molecular refractivity calculated as described in <i>Scott et al.</i> [100]
Weight	Molecular weight
logP (o/w)	Log of the octanol/water partition coefficient. This descriptor is calculated from a linear atom type model [97]
SLogP	Log of the octanol/water partition coefficient calculated as described in <i>Scott et al.</i> [100]
vdw_vol	van der Waals volume
density	Molecular mass density
vdw_area	Area of van der Waals surface

In table 2.1 are reported the molecular descriptors describing simple physical properties. Their calculation does not require knowledge of molecular conformations and only the connections of the atoms within the molecule are important. The first two descriptors, *apol* and *bpol* use the atomic polarizability for their calculation, as reported in the *Handbook of Chemistry and Physics (1994)*. *FCharge* is the total charge of the molecule, obtained as the sum of each atomic formal charge. This descriptor takes values equal to 0, ± 1 , ± 2 , ... $\pm \infty$ only. *mr* is the molecular refractivity calculated from a linear model obtained by 1,947 small molecules. *SMR* is the molecular refractivity obtained from an atomic contribution model. *Weight* is the molecular

weight. $\log P(o/w)$ is the octanol/water partition coefficient calculated from a linear model obtained by 1,847 molecules. $S\log P$ is the octanol/water partition coefficient obtained from an atomic contribution model. vdw_vol is the van der Waals volume calculated using a connection table approximation: it means that for calculation of the descriptor is considered only how the atoms in the molecule are connected, not the molecular conformation. $density$ is the ratio of $Weight$ to vdw_vol . vdw_area is the area of van der Waals surface calculated using a connection table approximation.

VSA-Type descriptors

The nature and calculation of van der Waals surface area (VSA) descriptors is well described by Paul Labute, the author of this set of descriptors [101]. The VSA-type descriptors are based on the van der Waals surface area (VSA). The VSA for each atom in a molecule is obtained from the surface area of the atom, approximated as a sphere with radius equal to the van der Waals radius, which is not contained in any other atom of the molecule. The molecular VSA is then calculated as the sum of each single VSAs in the molecule. The calculation of the descriptors occurs in two steps: Firstly, the VSA for each atom of a molecule is calculated. The calculated atomic VSAs are successively combined with physicochemical properties as molar refractivity (MR), lipophilicity ($\log P$) (both calculated using the Wildman and Crippen’s method) and atomic partial charges (calculated using the Gasteiger method) taken only in a certain range, as described in the following equation:

$$P_VSA(u, v) = \sum_i V_i \delta(P_i \in [u, v])$$

where V_i is the atomic VSA, and P is the calculated atomic property in the predefined range. The result is a set of 10 descriptors for $\log P$, 8 descriptors for MR and 14 descriptors for the atomic partial charges. The interval boundaries for $\log P$ are: $-\infty$, -0.4, -0.2, 0, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, ∞ . The interval boundaries for MR are: 0, 0.11, 0.26, 0.35, 0.39, 0.44, 0.485, 0.56, ∞ . The interval boundaries for the atomic partial charges are: $-\infty$, -0.3, -0.20, -0.15, -0.10, -0.05, 0, 0.10, 0.15, 0.20, 0.25, 0.30, ∞ . Together, the VSA-Type descriptors describe a $10 + 8 + 14 = 32$ dimensional chemistry space.

Atom Counts and Bond Counts

Table 2.2: atom counts and bound counts descriptors.

Code	Descriptor name
a_aro	Number of aromatic atoms
a_count	Number of atoms
a_heavy	Number of heavy atoms
a_ICM	Atom information content (mean)
a_IC	Atom information content (total)
a_nH	Number of hydrogen atoms
a_nB	Number of boron atoms
a_nC	Number of carbon atoms
a_nN	Number of nitrogen atoms
a_nO	Number of oxygen atoms
a_nF	Number of fluorine atoms
a_nP	Number of phosphorus atoms
a_nS	Number of sulfur atoms
a_nCl	Number of chlorine atoms
a_nBr	Number of bromine atoms
a_nI	Number of iodine atoms
b_1rotN	Number of rotatable single bonds
b_1rotR	Fraction of totale single bonds
b_ar	Number of aromatic bonds
b_count	Number of bonds
b_double	Number of double bonds
b_heavy	Number of bonds between heavy atoms
b_rotN	Number of rotatable bonds
b_rotR	Fraction of rotatable bonds
b_single	Number of single bonds
b_triple	Number of triple bonds
VAdjMa	Vertex adjacency informations (magnitude)
VAdjEq	Vertex adjacency informations (equality)

The atom counts and bond count descriptors describe the number of particular atom type and bound type in a molecule or properties directly connected with them. The atom counts and bond count descriptors calculated with MOE are reported in table 2.2.

Kier and Hall Connectivity and Kappa Shape Indices

The Kier-Hall connectivity indices calculated with MOE are divided in:

- Atomic connectivity indices (order 0 and 1)
- Carbon connectivity indices (order 0 and 1)
- Atomic valence connectivity indices (order 0 and 1)
- Carbon valence connectivity indices (order 0 and 1)

They consider each heavy atom (or only carbon atoms, in the case of carbon connectivity indices) of a molecule as a vertex. The connectivity δ for each vertex represents the number of direct connections of this vertex with the others. The connectivity index is then defined as:

$${}^0\chi = \sum_i \delta_i^{-1/2}$$

For connectivity indices of zero order and:

$${}^1\chi = \sum_i (\delta_i \cdot \delta_j)^{-1/2}$$

For connectivity indices of first order. In the second case, δ_i and δ_j are connectivity values of vertices with topological distance equal to 1. The atomic (or carbon) valence connectivity indices are obtained by replacing the vertex degree δ with the valence vertex degree δ^v in the formulas reported above. The valence vertex degree is defined as:

$$\delta^v = \frac{p_i - h_i}{z_i - p_i - 1}$$

where p_i is the number of valence electrons in the i atom, h_i is the number of hydrogens attached to the i heavy atom and z_i is the total number of electrons in the i atom.

The Kappa shape indices compare the molecular graph with the minimal and maximal molecular graphs that are possible to be obtained with the same atoms number. The first, second and third order κ shape indices are calculated as:

$${}^1\kappa = 2 \cdot \frac{{}^1P_{max} \cdot {}^1P_{min}}{({}^1P)^2} = \frac{A(A-1)^2}{({}^1P)^2}$$

$${}^2\kappa = 2 \cdot \frac{{}^2P_{max} \cdot {}^2P_{min}}{({}^2P)^2} = \frac{(A-1)(A-2)^2}{({}^2P)^2}$$

$${}^3\kappa = 4 \cdot \frac{{}^3P_{max} \cdot {}^3P_{min}}{({}^3P)^2} = \begin{cases} \frac{(A-3)(A-2)^2}{({}^3P)^2} & \text{if } A \text{ is even} \\ \frac{(A-1)(A-3)^2}{({}^3P)^2} & \text{if } A \text{ is odd} \end{cases}$$

where P_{min} and P_{max} are the minimal and maximal molecular graphs and A is the number of atoms in the molecule. The Kier alpha-modified shape indices take also the covalent radius of atoms and their hybridization state into account. It can be insert into the original calculation a value α defined as:

$$\alpha = \frac{r_i}{r_C} - 1$$

where r_i is the covalent radius of atom i and r_C is the covalent radius of a carbon atom. The first, second and third order of Kier alpha-modified shape indices are calculated as:

$${}^1\kappa = \frac{(A + \alpha)(A + \alpha - 1)^2}{({}^1P + \alpha)^2}$$

$${}^2\kappa = \frac{(A + \alpha - 1)(A + \alpha - 2)^2}{({}^2P + \alpha)^2}$$

$${}^3\kappa = \begin{cases} \frac{(A + \alpha - 3)(A + \alpha - 2)^2}{({}^3P + \alpha)^2} & \text{if } A \text{ is even} \\ \frac{(A + \alpha - 1)(A + \alpha - 3)^2}{({}^3P + \alpha)^2} & \text{if } A \text{ is odd} \end{cases}$$

The Kier molecular flexibility index *KierFlex* is calculated as:

$$KierFlex = \frac{(KierA1)(KierA2)}{n}$$

where n is the number of atoms in the molecule. The Zagreb index is defined as the sum of the vertex degrees for all the heavy atoms of the molecule.

Adjacency and Distance Matrix Descriptors

The adjacency and distance matrix descriptors are based on the adjacency matrix and the distance matrix of a chemical structure. If a molecule is represented as a graph, the adjacency matrix shows which vertices (atoms) are adjacent (directly connected) to each other. The distance matrix shows on the other hand the topological distance between vertices in the graph representation of a molecule. The adjacency and distance matrix descriptors are summarized in table 2.3.

BCUT and GCUT descriptors

The BCUT and GCUT descriptors [102] are two particular kinds of adjacency matrix descriptors. Their calculation is based on a modified adjacency matrix, called Burden

Table 2.3: Adjacency and distance matrix descriptors.

Code	Descriptor name
balabanJ	Balaban's connectivity topological index
diameter	Topological diameter
petitjean	Graph-theoretical shape coefficient
radius	Topological radius
VDistEq	Vertex distance equality index
VDistMa	Vertex distance magnitude index
wienerPath	Wiener path number
wienerPol	Wiener polarity number

matrix.[103] For BCUT descriptors, the Burden matrix is defined as an adjacency matrix where the diagonal elements are selected atomic properties and the off-diagonal elements B_{ij} , only for adjacent atoms, take the value of $\pi^{-1/2}$, where π is the conventional bond order. Remaining elements take the arbitrary value of 0.001. The matrix eigenvalues are calculated and the smallest, second, third, and largest eigenvalues are used as descriptors. The GCUT descriptors differ to BCUT descriptors only for the definition of the used matrix. For calculation of GCUT descriptors, the values of off-diagonal entries are calculated as $d^{-1/2}$, where d is the graph distance between the two atoms. MOE uses as atomic properties for the matrix diagonal the PEOE partial charges, the atomic contribution to logP and the molar refractivity (both calculated using the Wildman and Crippen SlogP/SMR methods).

Pharmacophore Feature Descriptors

The pharmacophore feature descriptors calculated with MOE can be divided into two subgroups: Descriptors that simply count the number of atoms included in predefined pharmacophoric groups in the molecule and descriptors calculated as the sum of the van der Waals areas of atoms included in the considered pharmacophoric groups. The feature sets used for descriptors calculation are:

- Number of hydrogen bond acceptor atoms
- Number of acid atoms
- Number of basic atoms
- Number of hydrogen bond donor atoms
- Number of hydrophobic atoms

2.1.2 3D MOE descriptors

Potential Energy Descriptors

The potential energy descriptors are calculated using each single component of the potential energy value of a molecular conformation. This class of descriptors comprises the value of the potential energy, descriptors based on bonded forces, and on non-bonded forces. The value of the potential energy is the sum of all single energy components of the force field. Descriptors based on bonded forces calculate the potential energy component of atoms connected by covalent ligands only. They comprise the bond stretching energy, the angle bending energy, the stretch-bend interaction energy, the out-of-plane bending energy, and the torsion energy. Descriptors based on non-bonded forces are the van der Waals energy and the electrostatic energy. [104]

Surface Area and Volume Descriptors

Table 2.4: Surface area and Volume descriptors.

Code	Descriptor name
ASA	Water accessible surface area
dens	Mass density
glob	Globularity
pmi	Principal moment of inertia
pmi(XYZ)	x,y,z components of the principal moment of inertia
rgyr	Molecular weight
std_dim1	Standard dimension1
std_dim2	Standard dimension2
std_dim3	Standard dimension3
vol	van der Waals volume
VSA	van der Waals surface area

These descriptors are correlated with the molecular conformation and are useful to describe dimensional parameters (like volume and surface) of molecules. The water accessible surface area (ASA) and the van der Waals surface area (VSA) quantify the molecular surface using different approaches. The water accessible surface area is calculated using a sphere with diameter of 1.4 Å that simulates a water molecule. The van der Waals surface area is calculated as the sum of the atomic van der Waals surface areas not included in other atoms. Analogously, the van der Waals volume is calculated as the volume included in the van der Waals surface area. The mass density is defined as the molecular weight divided by the van der Waals volume. The value of globularity indicates how is the molecule in the bulk extended. [105] A value

of globularity tending to 1 means that the molecule has a globular form while a value of globularity tending to 0 is associated with linear structures. This descriptor is obviously highly dependent on the molecular conformation. The principal moment of inertia is defined as:

$$pmi = \sum_{i=1}^A m_i \cdot r_i^2$$

where A is the atom number, m_i is the atomic mass and r_i is the perpendicular distance from the axis of the atom i . The three x, y, z components of the principal moment of inertia are defined as:

$$pmiX = \sum_{i=1}^A m_i \cdot (y_i^2 + z_i^2) \quad pmiY = \sum_{i=1}^A m_i \cdot (x_i^2 + z_i^2) \quad pmiZ = \sum_{i=1}^A m_i \cdot (x_i^2 + y_i^2)$$

where x, y and z are the atom coordinates.[98] The three standard dimensions (std_dim1, std_dim2 and std_dim3) calculated by MOE are the largest, second largest and third largest eigenvalues of the covariance matrix of the atomic coordinates.

Conformation Dependent Charge Descriptors

These conformation dependent descriptors are calculated using the stored partial charges of molecules. This class includes the dipole moment and its three separated components, as well as several descriptors calculated combining the partial charges with the water accessible surface area of the molecule.

Vsurf Descriptors

The Vsurf descriptors are the MOE implementation of the VolSurf descriptors developed by Cruciani et. al. [106]. These conformation dependent descriptors are able to numerically represent molecular interaction field (MIFs) informations about the molecular size, the molecular shape, the distribution of hydrophobic and hydrophilic regions and their balance [98]. Interactions fields are calculated using a water probe, a hydrophobic probe or other probes, like ionic probes [106].

MOPAC Descriptors

The MOPAC descriptors are conformation dependent descriptors calculated using the MOPAC package implemented in MOE. The semiempirical Hamiltonians PM3, MNDO or AM1 are used in the calculation to obtain the dipole moment (kcal/mol), the total energy (kcal/mol), the electronic energy (kcal/mol), the heat of formation (kcal/mol),

the ionization potential (kcal/mol), the energy (eV) of the Lowest Unoccupied Molecular Orbital (LUMO), and the energy (eV) of the Highest Occupied Molecular Orbital (HOMO).

2.1.3 RECON/TAE descriptors

Molecular electron density can be represented by quantum chemically derived descriptors. These descriptors are useable with a wide kind of molecules and are easily obtained by ab-initio calculations. However, an inconvenience of these descriptors is the considerable computational effort expended to calculate them using quantum chemical methods [107]. This problem can be bypassed using Transferable Atom Equivalents (TAEs). TAEs are electron density fragments containing a charged nucleus and are delimited by discrete boundaries. It is possible to assembly TAE fragments to represent molecular electron densities. The RECON algorithm firstly combines TAE atoms together to generate the electron density distribution of molecules. The following step is the adjustment of surface electronic properties used for generating QSAR descriptors [107]. Finally, the light reflection algorithm is used to generate the TAE/Shape descriptors [108].

2.1.4 RDF descriptors

Formally, the Radial Distribution Function (RDF) of an ensemble of N atoms can be interpreted as the probability distribution to find an atom in a spherical volume of radius r [109]. The general form of the radial distribution function is given in the following equation:

$$g(r) = f \sum_i^{N-1} \sum_{j>i}^N A_i A_j e^{-B(r-r_{ij})^2}$$

where f is a scalar vector, A are atomic properties of the atoms i and j , r_{ij} is the interatomic distance between the atoms i and j , and B is a smoothing parameter that could be interpreted as a temperature factor that defines the movement of atoms.

The RDF descriptors calculated with MOE are divided into two groups, according to the used distance: the 3D-RDF descriptors are calculated using the geometrical distance, while the 2D-RDF descriptors are calculated using the topological distance. Each molecule is codified by either an RDF vector of 30 values (if 3D-RDF descriptors are used) or by a RDF vector of 15 values (if 2D-RDF descriptors are used). The step size of r is equal to 0.5 Å for 3D-RDF descriptors, while the 2D-RDF descriptors use a step size for r equal to 1 Å. The B value is fixed to 100 Å⁻².

The used atom properties for the calculation of RDF descriptors are the electronegativity (e), the atomic mass (m), the refractivity (p), and the atomic volume (v). RDF

descriptors were also calculated without the use of atomic properties (unweighted, u).

2.1.5 Inductive descriptors

Inductive descriptors basically quantify inductive and steric interactions between substituents and reaction centers. Other parameters that can be quantified with the same mathematical apparatus are the partial atomic charges, analogues of chemical hardness-softness and electronegativity [110].

2.1.6 Dragon descriptors

The Dragon descriptors used in this work were calculated using the free version of the program (E-Dragon), available at the internet address <http://www.vcclab.org/lab/edragon>.

Autocorrelation descriptors (DRAGON)

The Autocorrelation of a Topological Structure (ATS) also known as Moreau-Broto autocorrelation [111] is the most famous spatial autocorrelation, defined as:

$$ATS_k = \frac{1}{2} \sum_{i=1}^A \sum_{j=1}^A w_i \cdot w_j \cdot \delta(d_{ij}; k)$$

where w are atomic properties, A is the number of atoms of the molecule, k is the lag, and d_{ij} is distance (geometrical or topological) between the atoms i and j . $\delta(d_{ij}; k)$ is the Kronecker delta function: it is equal to 1 if $d_{ij} = k$, while it is equal to 0 if $d_{ij} \neq k$.

The ATS descriptors are divided into topological and geometrical ATS descriptors based on the used distance matrix (topological or geometrical, respectively). The δ values range from 1 to 15 for the topological ATS descriptors, while the geometrical ATS descriptors use δ values between 1.5 and 15.5 Å, with a step size of 0.5 Å. The atomic properties used for the calculation of the ATS descriptors are the VSA values (a), the electronegativity (e), the hydrogen bond acceptors (ha), the hydrogen bond donors (hb), the SLogP increments (L), the atomic mass (m), the SMR increments (s), and the atomic volume (v). Like the RDF descriptors, also the ATS descriptors were calculated without the use of atomic properties (unweighted, u).

3D-MoRSE descriptors (DRAGON)

The 3d-MORSE (3d-Molecule Representation of Structures based on Electron diffraction) descriptors represent the 3D structure of a molecule by a fixed number of variables independent on the number of atoms of the molecule [112]. 3D-MoRSE descriptors

are calculated using different atomic properties such as atomic mass, atomic volume, atomic electronegativity, and atomic polarizability [113].

WHIM descriptors (DRAGON)

The WHIM (Weighted Holistic Invariant Molecular) descriptors codify important molecular 3D informations regarding molecular size, shape, symmetry, and atom distribution [114]. The algorithm performs a principal component analysis on the cartesian coordinates of the molecule by using a weighted covariant matrix [98]. The atomic properties used to calculate WHIM descriptors are atomic mass, the van der Waals volume, the atomic electronegativity (e), the atomic polarizability and the electrotopological state indices of Kier and Hall [113].

GETAWAY descriptors (DRAGON)

The GETAWAY descriptors are calculated from the Molecular Influence Matrix (MIM) defined as:

$$H = M \cdot (M^T \cdot M)^{-1} \cdot M^T$$

where H is the MIM and M is the molecular matrix (the centered cartesian coordinates of the atoms in the molecule). The superscript T refers to the transposed matrix [98, 115].

2.2 Machine learning methods

2.2.1 Self-Organizing Maps (SOMs)

Self-organizing Maps (SOMs), discovered by Teuvo Kohonen 30 years ago, are unsupervised neural networks able to project a multidimensional chemical space to a one- or two-dimensional layer of neurons [116, 117]. Self-organizing Maps are able to solve a wide variety of classification problems as such as speech recognition (Kohonen, 1990) or can be used in rational Drug Design [118]. Basically, a SOM consists of a group of neurons arranged in a one-dimensional array of neurons or, more frequently, a two-dimensional plane. Each neuron is in relationship with a defined number of other neurons of the SOM, named neighbors [117]. According to the topological distance between a neighbor and the considered neuron, the neighbor will be classified as nearest neighbor, second-nearest neighbor, etc. The arrangement of the neighborhood of a neuron is normally squared, rectangular or hexagonal [119]. These three kinds of two-dimensional arrangement have the problem that neurons on the border of the map have fewer neighbors as compared to neurons on the center of the map. This problem

has been solved using a toroidal topology [120]. This solution has been used in the present work. In case of a square SOM, each neuron is codified as a vector:

$$m_i = (m_{i1}, m_{i2}, m_{i3}, \dots, m_{in})$$

Each neuron of a SOM is initialized with a random vector. The input information (in this work, a chemical compound) is also codified as a vector:

$$x_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})$$

Where vector elements consist of molecular descriptors. Now the task of the SOM now is to find the Best-Matching Cell or, in other words, the most similar neuron to the chosen input vector. The similarity between the input vector and the neurons of a SOM is based on the Euclidian distance:

$$d_{(p,q)} = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

The neuron with minimal distance to the input vector is defined as the winning neuron. Once that the winning neuron has been found, the algorithm modifies the neuron vector according to the equation:

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)]$$

Where m_i represent the winning neuron, x is the input vector, t is the training iteration and $h_{ci}(t)$ is the Gaussian kernel, that defines the neighborhood of the winning neuron and how these neurons must be modified. Mathematically, $h_{ci}(t)$ is defined as:

$$h_{ci} = \alpha(t) \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right)$$

Where $\alpha(t)$ is the scalar-valued adaptation gain, that monotonically decreases from 1 to 0, $\|r_i - r_c\|$ is the Euclidian distance between the winning neuron and the neighboring neuron and $\sigma(t)$ is the radius of the neighborhood. The self-organizing maps calculated with the Kohonen package implemented in R, start as default with a value of radius that covers 2/3 of all unit-to-unit distances. Thus, in the early iteration steps the SOM is globally modified. Modifications will be more restricted with progressing of the training and, nearly to the end of the training, they will affect only the winning neurons [119, 121]. On the end of the training process, objects with similar input vectors (molecular descriptors, in our case) will be mapped into the same neuron or in closely adjacent neurons [120]. Finally, the algorithm assigns the class of the neurons in the SOM by looking which element in the training set has the closest Euclidian

distance to the considered neuron. If more than one element is associated with the same neuron, the class is calculated by the amount of the training elements of the two classes: the most represented class, in term of training elements on the neuron, also will become the class associated to that neuron. A trained SOM can be used to classify an external (test) dataset. The test dataset has to be projected to the map. The best matching neuron assigns the class label to the object, according to the Euclidian distance between the input vector and the vector associated to the neuron.

2.2.2 Support vector Machine (SVM)

Support vector machines (SVMs) are kernel-based classification methods, proposed by Vapnik in the beginning of 80s and improved in 1995 [122, 123]. SVMs are widely used to solve classification problems in many research fields as computational biology and drug discovery [124, 125, 126, 127]. Like other classification methods, SVMs separate two or more classes of objects with different labels in a multidimensional space through a hyperplane. What distinguishes a SVM from other classification methods is how it chooses the hyperplane. As shown in figure 2.1a, more lines (or hyperplanes) are able to separate two populations of vectors in the training set, but which one will generalize well? The problem to find optimal hyperplanes for separating classes of vectors has been solved using support vectors: If two vector populations are separated without errors by a hyperplane, it is possible to take into account only the nearest training vectors to the hyperplane and to maximize the distance between the hyperplane and these vectors. The hyperplane with the maximized margins is also the hyperplane that provides the best classifier (figure 2.1b) [122, 128]. If we assume that the class labels are 1 and -1, x_i is the i^{th} vector in a dataset and y_i is the label of x_i . The dataset is linearly separable if there exists a vector w (called weight vector) and a scalar b (called bias) such that the inequities

$$\begin{aligned} w \cdot x_i + b &\geq 1 && \text{if } y_i = 1 \\ w \cdot x_i + b &\leq -1 && \text{if } y_i = -1 \end{aligned}$$

are valid for all dataset elements. The optimal hyperplane, which minimizes the margins, is defined as [122]:

$$w_0 \cdot x + b = 0$$

Unfortunately, data coming from biological assays are not ideal, they contain errors. In this case it is impossible to cleanly separate two clusters using a line or a hyperplane. The SVM algorithm has to be modified to consider this problem and to

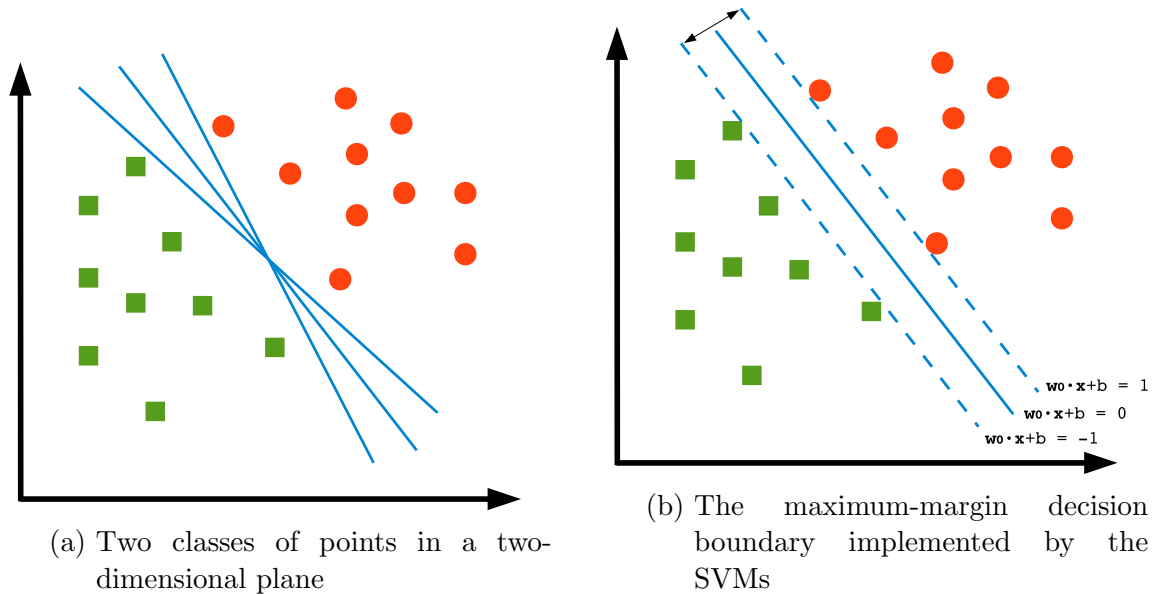


Figure 2.1: Linear separation with hyperplanes (reproduced from Tarca et. al., 2007)

be able to separate the dataset with the minimal number of errors. The presence of errors can be allowed by introduction of a positive variable ξ called *slack variable* that permits to exclude the elements in the training dataset that seem to be experimental errors. In other words, the hard margin in the original form of the algorithm has been substituted with a soft margin. This idea can be expressed in mathematical form as the minimization of the function:

$$\frac{1}{2}w^2 + CF \left(\sum_{i=1}^l \xi_i \right)$$

where C is a constant parameter which sets the weight of the slack variables and F is a monotonic convex function. Large values of C mean that a large penalty is assigned to errors [124].

In many practical cases, nonlinear classifiers generate better accuracy than linear classifiers. It is possible to extend the linear SVMs to nonlinear SVMs, if it is required by the analyzed problem. This extension is provided by using a kernel transformation, which adds an additional dimension to the dataset. In the transformed feature space, the kernel transformation enables to linearly separate a dataset that was not separable in the original feature space [129]. Two examples of kernels, implemented in R and used in the present work are the polynomial:

$$k(x,z) = (xz^T + 1)^d$$

And the radial basis function (RBF):

$$k(x,z) = \exp(-\gamma||x - z||^2)$$

2.2.3 k-Nearest Neighbors (k-NN)

k-Nearest Neighbors (k-NN) is a simple and fast algorithm that assigns the class label of an input vector according to the majority class of its closest k-nearest neighbors. The distance between the input vector and the neighborhood is generally the Euclidian distance, defined as:

$$d_{(p,q)} = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

The parameter k is the number of neighbors considered for assigning the class label. Normally this value is odd, to avoid undecideds.

2.3 Analysis of data

In the two class case with classes 1 (positive) or 0 (negative), the single prediction has four different outcomes. The *true positive* (TP) and *true negative* (TN) are the two cases of correct classification, while the *false positive* (FP) and the *false negative* (FN) are not. The false positive occurs when the outcome is predicted as positive but it is actually negative. Otherwise, the false negative occurs when the positive outcome is incorrectly predicted as negative. The *true positive rate* is given by [130]:

$$TP \text{ rate} = \frac{TP}{TP + FN}$$

The *true negative rate* is otherwise given by:

$$TN \text{ rate} = \frac{TN}{TN + FP}$$

The overall quality of the prediction is given by the *accuracy*, expressed as follows:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}$$

The *confusion matrix*, an easy representation of the results of a two classes prediction, is given as shown below:

	Positive predicted	Negative predicted
Positive observed (active)	TP	FN
Negative observed (inactive)	FP	TN

2.4 3D QSAR

While the classical QSAR methods relate the biological activity of compounds to their physicochemical parameters, the 3D QSAR methods, particularly comparative molecular field analysis, take in to consideration also the three-dimensional structures and the binding modes of protein ligands [131]. One of the first developed 3D QSAR methods is the Comparative Molecular Field Analysis (CoMFA). This method was described for the first time by *Cramer et. al* [132].

The compounds used for CoMFA should be structurally related and should have the same kind of interaction with the target protein. In contrast to classical QSAR, it is sufficient that they conserve a common pharmacophore and not necessarily the same skeleton, as required for classical QSAR.

The first step on performing a CoMFA analysis consists to generate the low energy 3D structure of the most active compound in the chosen molecules set. In the next step, the other molecules of the dataset are superimposed on to the template structure, according to the common pharmacophore. The superimposed molecules are finally placed into a three-dimensional box that is several Å larger than the combined volumes of all molecules. The box is then divided into a regular lattice with a default distance of 2 Å between the grid points (Figure 2.2).

The neutral carbon atom, charged atoms or hydrogen bond donor or acceptors are used as probe atoms to measure the interaction energies at each grid point (the van der Waals, the Coulomb and the hydrogen bond interactions, respectively). The van der Waals interactions are calculated using the Lennard-Jones potential equation:

$$E_{vdW} = \sum_{i=1}^n (A_{ij}r_{ij}^{-12} - C_{ij}r_{ij}^{-6})$$

where E_{vdW} is the sum of the van der Waals interaction energies, r_{ij} is the distance between the atom i and the grid point j , where the probe atom is localized, and A_{ij} and C_{ij} are constants related to the van der Waals radii of the corresponding atoms.

The Coulomb interactions are calculated using the Coulomb potential equation:

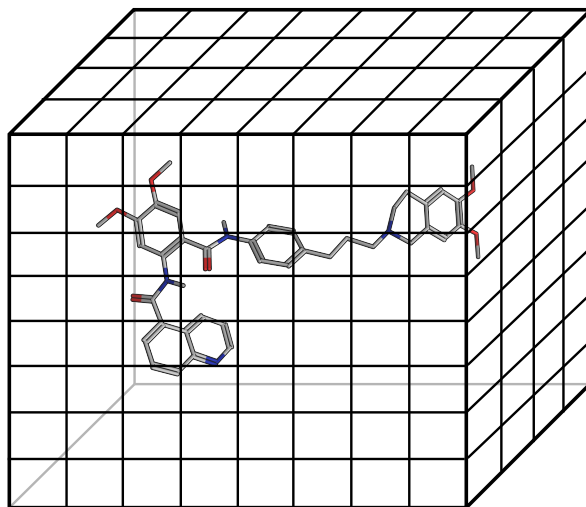


Figure 2.2: Schematic representation of the CoMFA grid. For clearer representation, only one molecule is shown and only the outer lines of the grid are drawn.

$$E_c = \sum_{i=1}^n \frac{q_i q_j}{\mathcal{E} r_{ij}}$$

where E_c is the Coulomb interaction energy, q_i is the partial charge of the atom i , q_j is the charge of the probe atom j , \mathcal{E} is the dielectric constant, and r_{ij} is the distance between the atom i and the grid point j , where the positive charge is localized.

As shown in figure 2.3, the Lennard-Jones and the Coulomb potentials can change their values from very small to very large numbers within an Ångström fraction. In order to prevent the presence of too small and too large values in the CoMFA fields, cut-off values were defined. However, small changes of the molecule orientation could result in very different results of the CoMFA analysis [134].

In contrast to CoMFA, the comparative molecular similarity indices analysis (CoM-SIA) [135] does not calculate interaction energies but 'similarity indices' to probe atoms. These similarity indices are calculated at the intersections of a regular lattice [136]. The similarity indices are calculated using the following equation:

$$A_{F,k}^q(j) = \sum_{i=1}^n W_{probe,k} w_{ik} e^{-\alpha r_{iq}^2}$$

where $A_{F,k}$ is the similarity index between the molecule and a probe atom placed at the intersections of the lattice, i is the summation index over all atoms of the investigated molecule j , $w_{probe,k}$ is the probe atom with charge +1, radius equal to 1 Å, and hydrophobicity equal of +1, w_{ik} is the actual value of the physicochemical property k of atom i , α is the attenuation factor, and r_{iq} is the mutual distance between

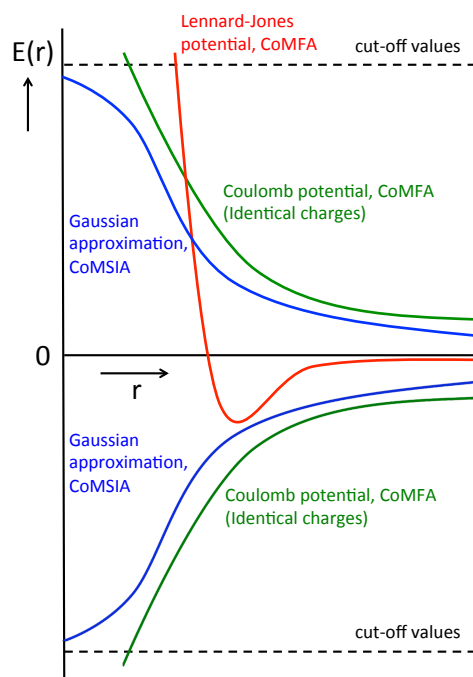


Figure 2.3: The Lennard-Jones and Coulomb functions used in CoMFA studies compared to the Gaussian approximation used in CoMSIA studies [131, 133].

the probe atom at the grid point q and the atom i of the molecule.

The use of the Gaussian potential instead of the Lennard-Jones and the Coulomb potentials leads also to limited energy values also short distances between the atom and the probe. For this reason, cut-off values are not needed in CoMSIA analysis (Figure 2.3). The better uniformity of the potential values in the CoMSIA analysis with respect to the CoMFA analysis allows the calculation of interactions inside the molecule as well as on the molecular surface. Furthermore, the CoMSIA analysis is not highly influenced by the orientation of the molecule in the lattice as in CoMFA analysis. These advantages of the CoMSIA analysis lead generally to better predictivity values in comparison to the CoMFA analysis.

2.4.1 PLS analysis

The matrix obtained from a CoMFA or a CoMSIA analysis consists of several thousand columns, while the number of rows (given by the number of molecules in the analysis) is generally lower than one hundred. For this reason, a normal multiple linear regression is not possible. The Partial Least Squares (PLS) analysis is the method of choice for deriving a linear relationship between the activity column (the Y vector) and the field variables (the X matrix) of CoMSIA and CoMFA analysis. Analogously to PCA, in the PLS analysis the X matrix (the CoMFA and CoMSIA fields) is decomposed

into the product of a score matrix and a loading matrix. The columns of the score matrix are called latent vectors. The latent variables can be considered as linear combination of the original field variables. The number of components to use for the analysis is chosen in order to find the maximal correlation between the X and Y values. The sequential extraction of the components is performed by trying to maximize the variance of the independent variables and the correlation with the Y vector. The next extracted component is orthogonal to the previous one and explains a part of the remaining variance. The extraction of the components is performed until reaching a maximal number of component or until complete explanation of the variance. The use of high numbers of components leads to overfitting by the model: In order to prevent this, the generated 3D QSAR models must be statistically validated.

2.4.2 Validation of 3D QSAR models

The quality of 3D QSAR models is mainly evaluated by using two parameters: the r^2 value (the fraction of explained variance) and the q^2 value (the cross-validated r^2). R^2 measures the ability of the 3D QSAR model to describe the variance of the data, or rather it estimates the fit's quality of the regression model obtained from the training set. R^2 is calculated as:

$$r^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

where y_i is the actual value, \hat{y}_i is the value predicted by the model and \bar{y} is the mean of the y values.

The q^2 value expresses the robustness of the 3D-QSAR model. It is calculated according to the following equation:

$$q^2 = 1 - \frac{PRESS}{SD} = 1 - \frac{\sum_{i=1}^n (Y_a - Y_p)^2}{\sum_{i=1}^n (Y_a - Y_m)^2}$$

where PRESS is the predictive sum of squares, SD is the sum of squared deviations, and Y_p is a predictive value. The values of q^2 can be between $-\infty$ and 1. If the value of q^2 is equal to 1, the model is perfect and each point of the model can be calculated without errors. A value of q^2 equal to 0 means that the model is not more predictive than a model where to each element of the dataset was assigned the average of the observed activity values as predicted value. A negative q^2 value means that the predicted values are worse than the average of the activity values of the dataset.

3 Materials and Methods (Biological testing)

3.1 Materials

3.1.1 Chemicals

Table 3.1: Chemicals used in this work.

Chemicals	Manufacturer	Article number
Calcein AM	Sigma	17783
Calcium chloride dihydrate	Merck	P 4901
Cyclosporine A	Sigma	C 3662
D-glucosemonohydrate	Merck	1040740500
Dimethylsulfoxide (DMSO)	Acros	AC 19773
Disodium hydrogenphosphate	Applichem	A 4732
HEPES (free acid)	Appluchem	A 3707
Hoechst 33342	Sigma	B 2261
Hydrochloric acid (0.5 M)	Grüssing	24204
Ko143	Tocris	3241
Magnesium sulfate heptahydrate	Applichem	A 4101
Melsept SF	Braun	18907
Methanol	Merck	107018
Mitoxantrone	Sigma	M 6545
Potassium chloride	Merck	104936
Potassium dihydrogenphosphate	Applichem	A 3095
SN-38	TCI Europe N.V.	E 0748
Sodium bicarbonate	Merck	106329
Sodium chloride	Merck	106404
Sodium hydroxide solution (1 M)	Grüssing	22195

3.1.2 Materials

Table 3.2: Materials for assays and cell culture.

Material	Manufacturer	Article number
1.0-5.0 mL Bulk, Natural pipette tips	Starlab	11009-5000
1.5 mL Amber microtubes with attached pp cap	Sarstedt	72690004
1.5 mL Neutral microtubes with attached pp cap	Sarstedt	72690001
2.0 mL MaxyClear microtubes	Axygen scientific	MCT-200-C
96 Well, clear PS microplate, flat bottom	Greiner bio-one	655098
96F untreated black micro well SH plate	Nunc	237108
96-Well tissue culture plate, flat bottom with lid, sterile	Sarstedt	831835
CASYton solution	Schärfe System	43001
Conical test tube PP 15 mL, sterile	Nerbeplus GmbH	25027001
Conical test tube PP 50 mL, sterile	Nerbeplus GmbH	25007001
Cryos PP with screw cap, sterile	Greiner bio-one	123263
Fetal bovine serum	Sigma	F 7524
Glas pasteur pipettes (230 mm)	VWR international	612-1702
Growth medium D-MEM 5671	Sigma	M 5650
Growth medium RPMI-1640	PAN Biotech GmbH	P 0416500
Membrane filter 0.2 μm , sterile	Whatman	10462200
Norm-Ject 10 mL syringe	Henke Sass Wolf	4100-000V0
Norm-Ject 20 mL syringe	Henke Sass Wolf	4200-000V0
Penicillin-Streptomycin solution	Sigma	P 0781
Serological pipette 10 mL, sterile	Sarstedt	86.1254.001
Serological pipette 25 mL, sterile	Sarstedt	86.1685.001
TipOne 0.1-10.0 μL natural pipette tips	Starlab	S 1111-3000
TipOne 1-200 μL yellow pipette tips	Starlab	S 1111-0006
TipOne 101-1000 μL natural pipette tips	Starlab	S 1111-2020

Table 3.2: Materials for assays and cell culture.

Material	Manufacturer	Article number
Tissue culture flasks, 25 cm ² , sterile, filter cap	Greiner bio-one	690175
Tissue culture flasks, 75 cm ² , sterile, filter cap	Greiner bio-one	658175
Tissue culture flasks, 175 cm ² , sterile, filter cap	Greiner bio-one	660175
Trypsin-EDTA solution	PAN Biotech GmgH	P100231SP

3.1.3 Instruments

Table 3.3: Instruments used in this work.

Instruments	Manufacturer	Serial number
Accu-Jet suction pump	Brand	441938
Avanti centrifuge J-25	Beckman	JHY97G35
Axiovert 25 microscope	Zeiss	660197
CASY1 model TT	Schärfe System	SC1 TT
CO ₂ cell	MMM Group	-
CO ₂ water jacket incubator	Forma scientific	-
FLUOstar Optima fluorescence plate reader	BMG Lab Technologies	4131164, 4132279
Laminar flow cabinet (model: Antares 48)	Steril S.p.A.	10155/1996
pH-Meter 744	Metrohm	20506
Pipette 0.1-2.5 μ L	Eppendorf	3638475
Pipette 2-20 μ L	Eppendorf	33407866
Pipette 20-200 μ L	Eppendorf	3534296
Pipette 100-1000 μ L	Eppendorf	4741196
Pipette 500-5000 μ L	Eppendorf	3615095
POLARstar Galaxy fluorescence plate reader	BMG Lab Technologies	4030639
RH basic magnetic stirrer	IKA Labortechnik	3061661
Vacuum pump BVC21	Brand	08E12592
Vortex stirrer Minishaker	Vacuubrand	2.88069E+13

Table 3.3: Instruments used in this work.

Instruments	Manufacturer	Serial number
Waterbad type 1083	GFL	11530203

3.2 Cell culture

In this work, cells were aseptically cultivated in T75 flasks (or T175 flasks in case of cultures used for a backup) and conserved in humidified atmosphere containing 5% CO₂ at 37 °C. The used culture medium is specific for each cell line. The composition of each used culture medium will be described in chapter 3.3. To prevent cell stress, the culture medium was preventively warmed up in warm water at 37 °C. The culture medium contained phenol red, a pH indicator. If the color of the medium (due to phenol red) changes from red to yellow, the culture medium has to be changed. If the culture medium had to be changed, the old medium was drawn, using a vacuum pump. After removing the old culture medium, 15-20 mL of fresh medium was added to a T75 flask and 25-30 mL to a T175 flask. When the cell culture reaches a confluence value of about 90 %, the cells must be harvested. If cell harvesting was required, first the old culture medium was removed. In order to detach the cells from the flask surface, 3 mL of trypsin-EDTA (0.05-% trypsin/0.02-% EDTA) were added to the culture. After 3-5 minutes of incubation at 37 °C (15 minutes at 37 °C for MDCK cells), 7 mL of fresh culture medium were added to the culture and the suspension was centrifuged (266 xg, 4 °C, 4 min). The supernatant was removed and the cell pellet was then resuspended in 4-5 mL of fresh culture medium.

3.2.1 Thawing of cells

The vial containing frozen cells was taken from the liquid nitrogen tank and put in warm water (37 °C) under moderate agitation, until ice was no longer visible. The suspension was transferred under sterile conditions into a T75 flask containing 15-20 mL culture medium and incubated in humidified atmosphere containing 5-% CO₂ at 37 °C. After 6 h the old medium was removed and new warm (37 °C) medium was added.

3.2.2 Freezing of cells

Cells were cultivated in a T175 flask until reaching a confluence of about 90 %. When the ideal confluence was reached, cells were harvested as described in chapter 3.2. The obtained cell pellet was suspended in 2.7 mL fresh medium. Each vial was then filled with about 900 μL of the obtained cell suspension and 100 μL of DMSO. The vials were firstly stored at a temperature of $-80\text{ }^{\circ}\text{C}$ and than in liquid nitrogen.

3.3 Cell lines

3.3.1 MCF-7 and MCF-7 MX cell lines

The human breast cancer cell line MCF-7 (ECACC No. 86012803) and the mitoxantrone resistant cell line MCF-7 MX, kindly provided by Dr. E. Schneider (Wadsworth Center, Albany, NY, USA), were cultivated in RPMI-1640 medium supplemented with 20-% fetal bovine serum, 50 $\mu\text{g}/\text{mL}$ streptomycin and 50 U/mL penicillin G. The cells were incubated in a humidified atmosphere containing 5-% CO_2 at $37\text{ }^{\circ}\text{C}$.

3.3.2 A2780 and A2780 Adr cell lines

The A2780 and A2780 Adr cell lines were purchased from the European collection of animal cell cultures (ECACC, Salisbury, Wiltshire, UK). The A2780 cell line (ECACC, No.93112519, UK) was obtained from a human epithelial ovarian carcinoma. The A2780 Adr cell line (ECACC, No.93112520, UK) was selected by treating the A2780 cell line with doxorubicine several times. The A2780 and A2780 Adr cell lines were cultivated in RPMI-1640 medium supplemented with 10-% fetal bovine serum, 50 $\mu\text{g}/\text{mL}$ streptomycin and 50 U/mL penicillin G. Cells were incubated in a humidified atmosphere containing 5-% CO_2 at $37\text{ }^{\circ}\text{C}$.

3.3.3 MDCK II and MDCK II-BCRP cell lines

The canine kidney cell line MDCK was purchased from the European collection of animal cell cultures (ECACC No. 84121903). The cell line MDCK II-BCRP was a kind gif from Dr. A. Schinkel (The Netherlands Cancer Institute, Amsterdam, The Netherlands). Both cell lines were cultivated in DMEM medium supplemented with 10 % fetal bovine serum and 50 $\mu\text{g}/\text{mL}$ streptomycin and 50 U/mL penicillin G.

3.4 Preparation of the buffer solutions

3.4.1 Preparation of a 5x concentrated Krebs HEPES buffer

Table 3.4: Composition of the Krebs HEPES buffer.

Chemical name	Chemical formula	Molecular weight (g/mol)	Taken weight (g)
Sodium chloride	NaCl	58.4	17.330
Potassium chloride	KCl	74.6	0.876
Potassium dihydrogen phosphate	KH ₂ PO ₄	136.1	0.408
Sodium hydrogen carbonate	NaHCO ₃	84.0	0.882
D-glucose monohydrate	C ₆ H ₁₂ O ₆ · H ₂ O	198.2	5.796
HEPES (free acid)	C ₈ H ₁₈ N ₂ O ₄ S	238.3	5.958

The 5x stock Krebs HEPES buffer was prepared by mixing the substances indicated in table 3.4 into a 500 mL volumetric flask. The used mass of each substance is reported on the last column of the table. Distilled water was added to reach a volume of 450 mL and the pH of the solution was adjusted to 7.4 using 1 M sodium hydroxide. Finally, distilled water was added until the final volume of 500 mL was reached. The 5x stock Krebs HERPES buffer solution was stored at a temperature of -20 °C. The 1x Krebs HEPES buffer was prepared using 100 mL of the 5x stock solution. This solution was diluted to a volume of 400 mL with distilled water. 650 μ L of 1 M calcium chloride solution and 600 μ L of 1 M magnesium sulfate solution were added to the buffer solution. The final volume of 500 mL was reached adding distilled water. The 1x Krebs HERPES buffer solution was stored at a temperature of -20 °C.

3.4.2 Preparation a the PBS buffer

Table 3.5: Composition of the phosphate buffered saline (PBS) buffer.

Chemical name	Chemical formula	Molecular weight (g/mol)	Taken weight (g)
Sodium chloride	NaCl	58.4	8.0

Table 3.5: Composition of the phosphate buffered saline (PBS) buffer.

Chemical name	Chemical formula	Molecular weight (g/mol)	Taken weight (g)
Potassium chloride	KCl	74.6	0.2
Potassium dihydrogen phosphate	KH ₂ PO ₄	136.1	0.2
Disodium hydrogen phosphate	Na ₂ HPO ₄	141.9	1.44

The phosphate buffered saline (PBS) buffer was prepared by mixing in a flask the substances indicated in table 3.5 with about 900 mL distilled water. Afterwards, the pH was adjusted to 7.4. Finally, the volume of 1000 mL was reached, adding distilled water. This solution was sterilized by autoclaving (20 min, 121 °C, liquid cycle) and stored at a temperature of 4 °C.

3.5 Cell counting using the CASY[®] model TT

In order to count the cell number in a cell suspension and to test their viability, 20 μ L of the suspension of cells and culture medium were taken and put into 10 mL of sterile CASY[®]ton solution, an isotonic and iso-osmotic electrolyte (Schaerfe System Ltd.). The obtained diluted suspension was measured using a CASY[®] 1 Modell TT. This instrument uses two methods: the Electrical Current Exclusion (ECE) and the Pulse Area Analysis. The Electrical Current Exclusion method is based on the principle that viable and dead cells suspended in CASY[®]ton have different permeability to the electrical current: Viable cells have an intact membrane that is an electrical insulator, whereas dead cells possess pores in the membrane that allow the current passage. The cells in the suspension are drawn one by one through a capillary with defined diameter (150 μ m). The suspension is then exposed to a low and cycled electrical current. A decreasing of the electrical signal is correlated with the passage of a cell and is proportional to the cell volume. Each cell is measured several hundred times by the machine. The results related to each cell were analyzed by Pulse Area Analysis, a digital pulse processing technique developed by Schaerfe System Ltd [137]. In the present work, the dimension limits for viable cells were set between 8 or 10 μ m and 40 μ m. Particles with diameter bigger than 40 μ m were associated with cell agglomerates.

3.6 Accumulation Assays

3.6.1 Hoechst 33342 assay

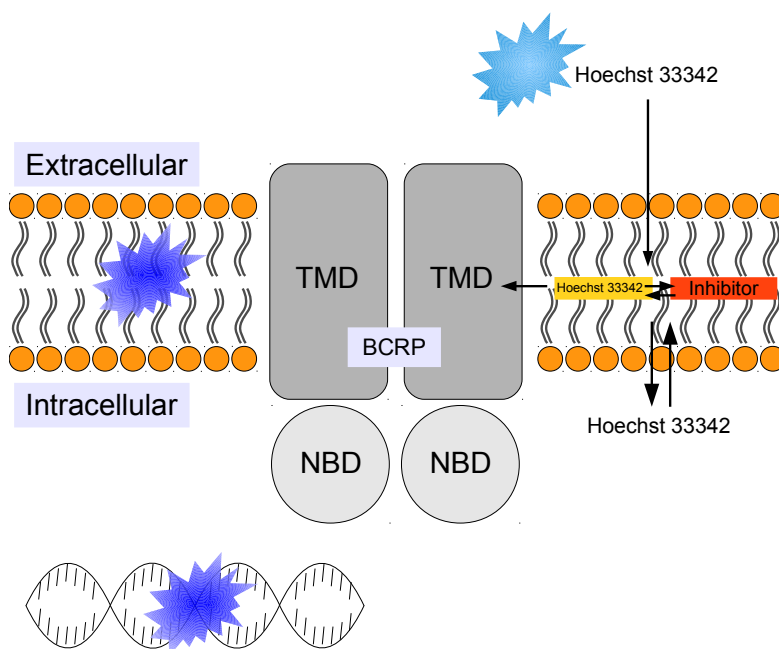


Figure 3.1: Hoechst Assay

Principles of the method

The Hoechst 33342 assay is based on two main aspects: First, the bisbenzimidazole Hoechst 33342, used in this assay, is a well-known BCRP and P-gp substrate [55]. Second, Hoechst 33342 forms strong and non-covalent complexes with DNA, binding to the minor groove of DNA [138]. Both, binding to DNA and a lipophilic environment, like the cell membrane, lead to a several hundred fold increase of the intrinsic fluorescence of the dye. [139, 140]. In cells not expressing BCRP or P-gp, Hoechst 33342 accumulation gives increasing fluorescence, while in cells expressing one of these two transporters the Hoechst 33342 accumulation is decreased, due to active efflux of the dye. In cells expressing BCRP or P-gp, accumulation of Hoechst 33342 is increased if a modulator is added. In this case, the accumulation of Hoechst 33342 depends to the concentration of the modulator and its potency.

Assay execution

Cells were harvested after reaching a confluence of 80-90-% with trypsin-EDTA (0.05-% trypsin/0.02-% EDTA), centrifuged (266 g, 4 °C, 4 min) and then resuspended in fresh culture medium. The cell density was measured using the Casy I Modell TT cell

counter device. Later, cells were again centrifuged and the medium eliminated. The pellet was then resuspended in Krebs-Hepes buffer. 90 μL of this final suspension, containing 27,000 cells, were seeded into each well of a black 96-well plate and 10 μL of different inhibitor concentrations was added. Finally, the prepared plates were incubated under 5-% CO_2 atmosphere and 37 $^\circ\text{C}$ for 30 minutes. 20 μL of a 6 μM Hoechst 33342 solution was added to each well. The fluorescence of each well was measured at constant intervals of 60 seconds, at excitation wavelength of 355 nm and an emission wavelength of 460 nm, using a BMG POLARstar microplate reader tempered at 37 $^\circ\text{C}$. The data was analyzed considering the fluorescence kinetic as not linear. With the use of the program Graph Pad Prism (version 5.1)[®] it was possible to fit the kinetic curve as a one-phase exponential association function. For each modulator concentration, the top of this function was used to build the dose response curve.

3.6.2 Calcein AM assay

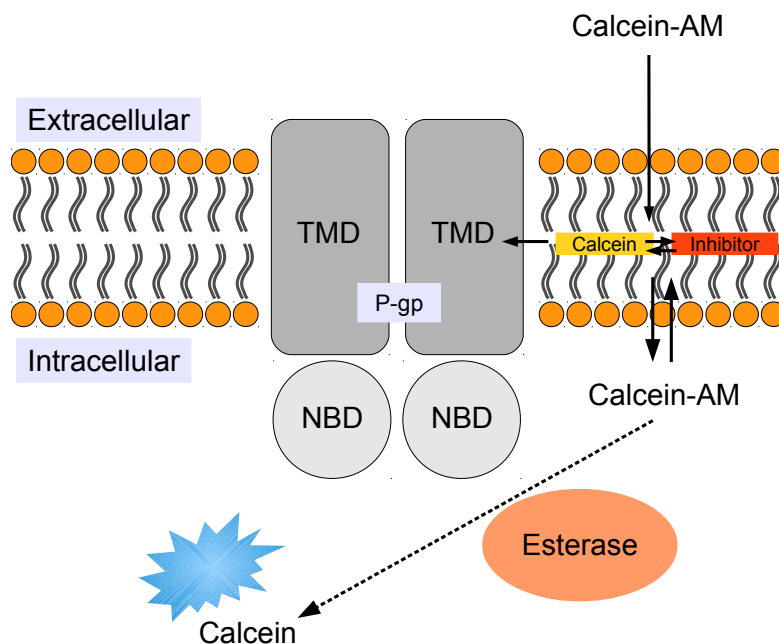


Figure 3.2: Calcein Assay

Principles of the method

Calcein acetoxymethyl ester (AM) is a non-fluorescent lipophilic P-gp substrate. Calcein AM is rapidly hydrolyzed by intracellular esterases to the free acid calcein, which is not a P-gp substrate. Calcein is a fluorescent non-cytotoxic compound with a high molar emission coefficient. The fluorescence excitation maximum of calcein is 496 nm and its emission maximum is at 517 nm. If calcein AM is administered to P-gp

3 Materials and Methods (Biological testing)

expressing cells, the intracellular concentration of calcein AM remains low and consequently also the total amount of calcein, resulting in low fluorescence. In presence of P-gp modulators the transport of calcein AM to outside the cells is inhibited, causing accumulation of calcein in the cells and increasing fluorescence.[141, 142]

Assay execution

The doxorubicin resistant cell line A2780 Adr and the parental cell line A2780 were grown in RPMI-1640 medium supplemented with 10-% fetal bovine serum, 50 $\mu\text{g}/\text{mL}$ streptomycin, and 50 U/mL penicillin G, in a 5-% CO_2 atmosphere at 37 °C. After reaching a confluence of about 90-%, the cells were harvested using trypsin-EDTA solution (0.05-% trypsin/0.02-% EDTA). The cells were centrifuged (266 g, 4 °C, 4 min) and then resuspended in fresh culture medium. The cell density was measured with the Casy I Modell TT cell counter device. Cells were again centrifuged and the medium eliminated. The pellet was than resuspended in Krebs-Hepes buffer. 90 μL of this final suspension, containing 30,000 cells, were seeded into each well of a black 96-well plate and 10 μL of different inhibitor concentrations were added. The prepared plates were than incubated under 5-% CO_2 atmosphere and 37 °C for 30 minutes. Finally, 33 μL of a 1.25 μM calcein-AM solution were added to each well. The fluorescence of each well was detected at constant intervals of 60 seconds up to 60 min using an excitation wavelength of 485 nm and an emission wavelength of 520 nm at a temperature of 37 °C in a BMG POLARstar microplate reader. The slope of the initial linear part of each fluorescence-time curve was calculated by linear regression. Concentration-response curves were generated from the slopes by nonlinear regression using the 3- or 4-parameter logistic equation (with or without variable Hill slope) implemented in GraphPad Prism[®].

3.7 MTT assay

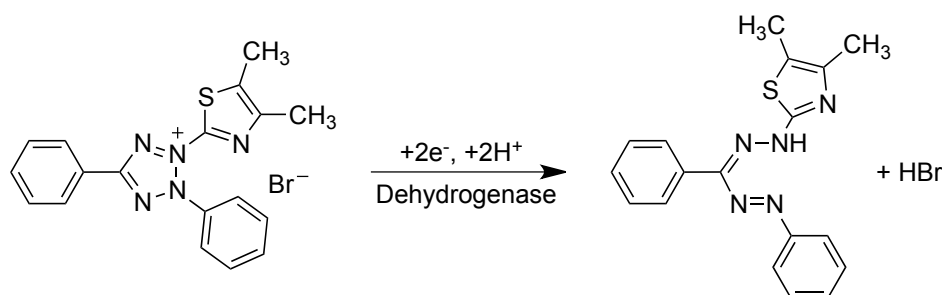


Figure 3.3: Reduction of the yellow compound MTT to the violet compound Formazan

Principles of the method

The toxicity of cytotoxic drugs or new substances against several cell lines can be measured using MTT assay, as described for the first time by Mosmann et. al. [143]. The method is based on the reduction of the yellow substance 3-(4,5-dimethylthiazole-2-yl)-2,5-diphenyl tetrazolium bromide, MTT, (Thiazolyblau) to violet-colored formazan precipitate. This reaction is catalyzed by mitochondrial dehydrogenases and it is possible only in living cells. The absorption of the formazan solution is linearly correlated with cell viability.

Shift of the dose-response curve of cytotoxic drugs

Cells were seeded into a 96-well tissue culture plates with a density of 10,000 cells per well (for MCF-7 and MCF-7 MX cells) or 1,250 cells per well (for MDCK cells) in a total volume of 160 μL and kept at 37 °C under 5-% CO_2 for 6 h. After the cells had attached to the well bottom, 20 μL of test compound and 20 μL of cytostatic solution were added. Cells were then incubated for 72 hours and then the MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) reagent (40 μL of a 5mg/mL solution) was added to each well. After 90 minutes of incubation with MTT, the supernatant was removed and cells were lysed with 100 μL DMSO per well. Viability of the cells was measured with spectrophotometry by absorbance at 544 nm and background corrected at 710 nm using a BMG POLARstar microplate reader.

Determination of cytotoxicity of the studied compounds

MCF-7 and MCF-7 MX cells were seeded into 96-well tissue culture plates with density of 10,000 cells per well in a total volume of 90 μL and kept at 37 °C under 5 % CO_2 for 6 h. After the cells had attached to the bottom of the well, 10 μL of test compound at four different concentrations was added. 10 μL of PBS and 10 μL of DMSO were used as negative and positive controls respectively. Cells were incubated for 72 h and then the MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) reagent (40 μL of a 5mg/mL solution) was added. After 90 minutes of incubation with MTT, the supernatant was removed and cells were lysed with 100 μL DMSO per well. Viability of the cells was measured with spectrophotometry by absorbance at 544 nm and background corrected at 710 nm using a BMG POLARstar microplate reader.

3.7.1 Analysis of the data using GraphPad Prism[®]

In this work, the potency of the studied compounds against BCRP and P-gp was reported as $\text{IC}_{50} \pm \text{SD}$ values. The fluorescence values obtained by Hoechst 33342 assay and calcein-AM assay were analyzed using the statistics software GraphPad

3 Materials and Methods (Biological testing)

Prism[®] (version 5.1). In order to calculate the IC₅₀ value of each single dose-response curve, both the 4-parametric logistic equation and the 3-parametric logistic equation (with Hill slope equal to 1) were calculated and compared. In case of not complete inhibition of the transporters at the highest used compound concentrations, the top values of the test compounds were fixed to the top value of the standard compound. All assays were performed at least three times and the reported IC₅₀ values are the reported as mean ± SD.

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

4.1 Optimization of the Hoechst 33342-assay

The fluorescent compound Hoechst 33342 has been recognized to be a P-gp [32] and BCRP [144, 145] substrate. Therefore, Hoechst 33342 can be used to investigate the inhibitory activity of small molecules against those two ABC-transporters. The Hoechst 33342-assay has been developed by Dr. Henrik Mueller and was used as HTS (High Throughput Screening) assay [146].

A goal of this work is the determination of the BCRP inhibitory potency of a new class of BCRP inhibitors. The determination of the potency of these compounds was determined using the Hoechst-assay in MCF-7 MX cells. In order to optimize the quality of the results, parameters of the Hoechst 33342-assay in MCF-7 MX cells, like the Hoechst 33342 concentration and the used cells number, were optimized.

4.1.1 Optimization of the used Hoechst 33342 concentration

To date, only a relatively small number of BCRP-inhibitors are known. One of these compounds is the potent P-gp inhibitor XR9577, used in this work as reference compound for the validation of the Hoechst-Assay. The standard protocol developed in our laboratory by Dr. Henrik Mueller used an Hoechst 33342 final concentration of 5 μ M and 27,000 cells in each vial. The relationship between fluorescence and Hoechst 33342 was investigated in MCF-7 cells, MCF-7 MX cells and MCF-7 MX cells to which the modulator XR9577 was added at a final concentration of 10 μ M. The fluorescence values reported are the difference of the fluorescence measured in the vials containing buffer, cells and fluorescent substrate minus the fluorescence of the vials containing only buffer and fluorescent substrate.

As shown in figure 4.1, the fluorescence of a 5 μ M Hoechst 33342 solution is not in the linear region of the concentration-fluorescence relationship and is already in a plateau region. This deviation from linearity is observed in MCF-7 cells and in MCF-7 MX cells with and without modulator. In all these three cases, the fluorescence increases

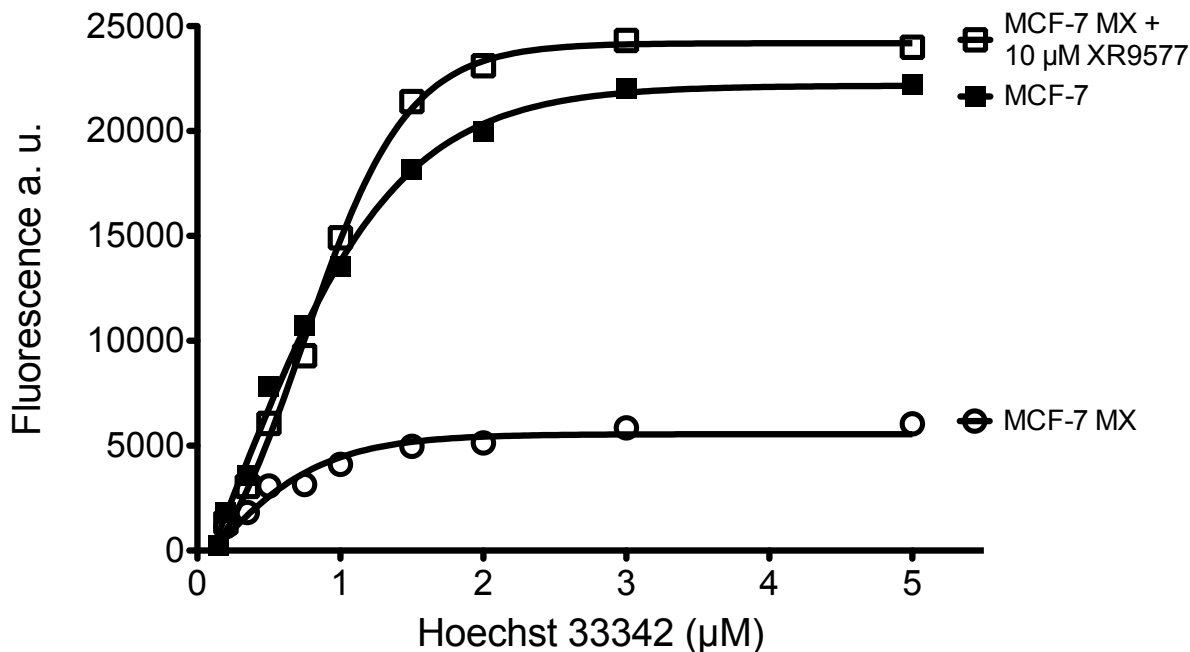


Figure 4.1: Fluorescence of Hoechst 33342 by increasing the substrate concentration (MCF-7 and MCF-7 MX cells).

with the substrate concentration only until a substrate concentration of about 2 μM . However, the relationship is linear only until a Hoechst 33342 concentration of about 1.5 μM .

The plateau in the fluorescence of Hoechst 33342 could be due to saturation of the binding of Hoechst 33342 to the membrane or to the DNA. Another explanation could be that the relatively high concentration of fluorescent substrate causes quenching effects. In figure 4.2, the concentration-fluorescence relationship for lower concentration of Hoechst 33342 up to 1 μM is shown. In this interval of concentrations, the fluorescence of the cells increases linearly with the concentration of Hoechst 33342.

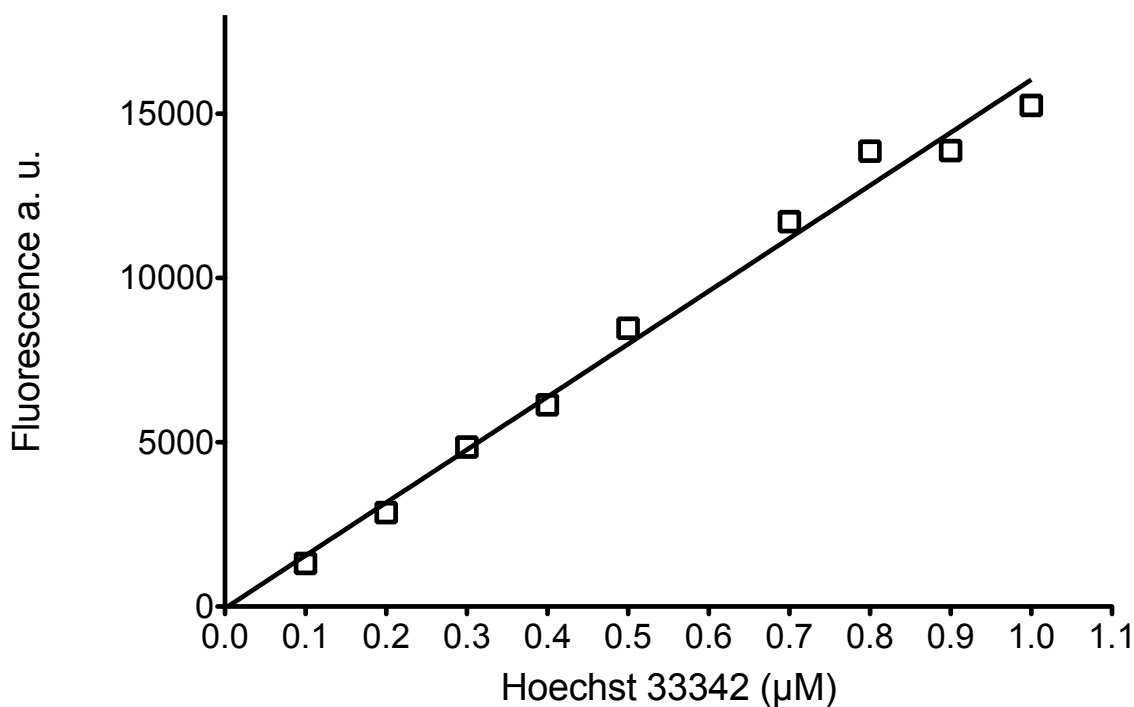
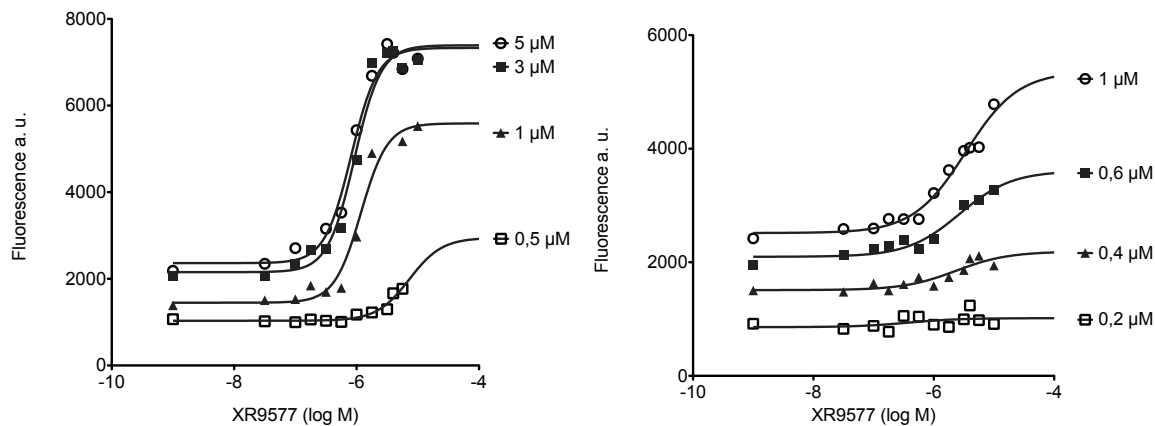


Figure 4.2: Fluorescence of Hoechst 33342 versus substrate concentration. The concentrations of Hoechst 33342 were between 0.1 and 1.0 μM (MCF-7 cells at 27,000 cells/well.).

In order to investigate which concentration of Hoechst 33342 should be used to obtain good quality dose response curves and, several dose response curves for the modulator XR9577 at different concentrations of Hoechst 33342 were determined. In figure 4.3a is shown that for concentrations of Hoechst 33342 between 1 and 5 μM , there is no difference in the span of the obtained dose response curves. The curves obtained with concentrations of Hoechst 33342 of 3 μM and 5 μM also show very similar bottom and top values, in agreement with the findings presented in fig 4.1. The dose response curves obtained using an Hoechst 33342 concentration of 1 μM has a similar span value as the other two curves, but has lower bottom and top values. The dose response curves obtained using an Hoechst 33342 concentration of 0.5 μM showed a further decrease of its bottom and top values and also a considerable reduction of the span.

As shown in figure 4.3b, all the dose response curves of the modulator XR9577 obtained with Hoechst 33342 concentration between 0.2 and 1 μM gradually present lower values of bottom, top and span. Those results are consistent with the experiment on the linearity of the relationship between fluorescence and concentration of Hoechst 33342. Considering the results of the experiments, it was decided to set the Hoechst 33342 concentration for the Hoechst 33342 assay to the value of 1 μM .

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.



(a) Dose-response curves obtained using concentrations of XR9577 between 0.5 and 5 μM .

(b) Dose-response curves obtained using concentrations of XR9577 between 0.2 and 1 μM .

Figure 4.3: Dependence of the Dose response curve of the modulator XR9577 on the concentration of Hoechst 33342 .

4.1.2 Optimization of the cell number.

In order to find the optimal number of cells for each well, the influence of the cell number on the fluorescence was investigated. The experiment was performed on the same 96-well plate, using MCF-7 cells from the same culture and different concentration of Hoechst 33342. The results reported in figure 4.4 show that for concentrations lower than 1 μM the correlation between the Hoechst 33342 concentration and the fluorescence remains linear for all used cell numbers. The quality of the linear correlation also increases with the number of cells, reaching a linear correlation for values higher than 20,000 cells/well. Also in figure 4.4 it is shown that for values of a cell number up to 40,000 the fluorescence increases for each Hoechst 33342 concentration with the cell number. Increasing the cell number in the well to 50,000 cells/well does not increase the fluorescence further. In figure 4.5 the correlation between the fluorescence and the number of MCF-7 and MCF-7 MX cells is represented. The relationship was investigated using cell suspensions with different concentrations of cells and with 10 μM final concentration of the modulator XR9755 and MCF-7 MX cells without modulator. In this experiment, the fluorescence of the buffer solution containing Hoechst 33342 at the same concentration was subtracted from the fluorescence in presence of cells. The results show that the fluorescence increases linearly for cells with and without modulator.

4.1 Optimization of the Hoechst 33342-assay

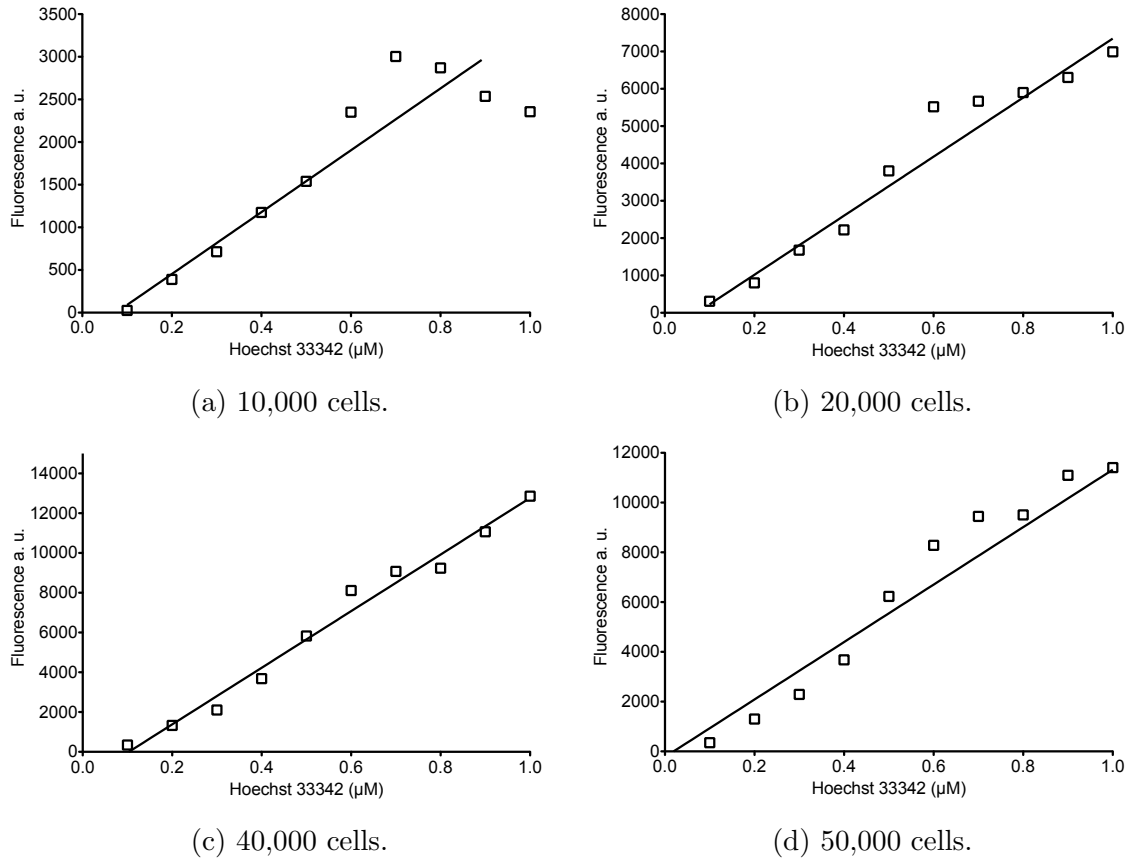


Figure 4.4: Linearity of the fluorescence for Hoechst 33342 in relation with the number of cells in each vial (MCF-7 cells).

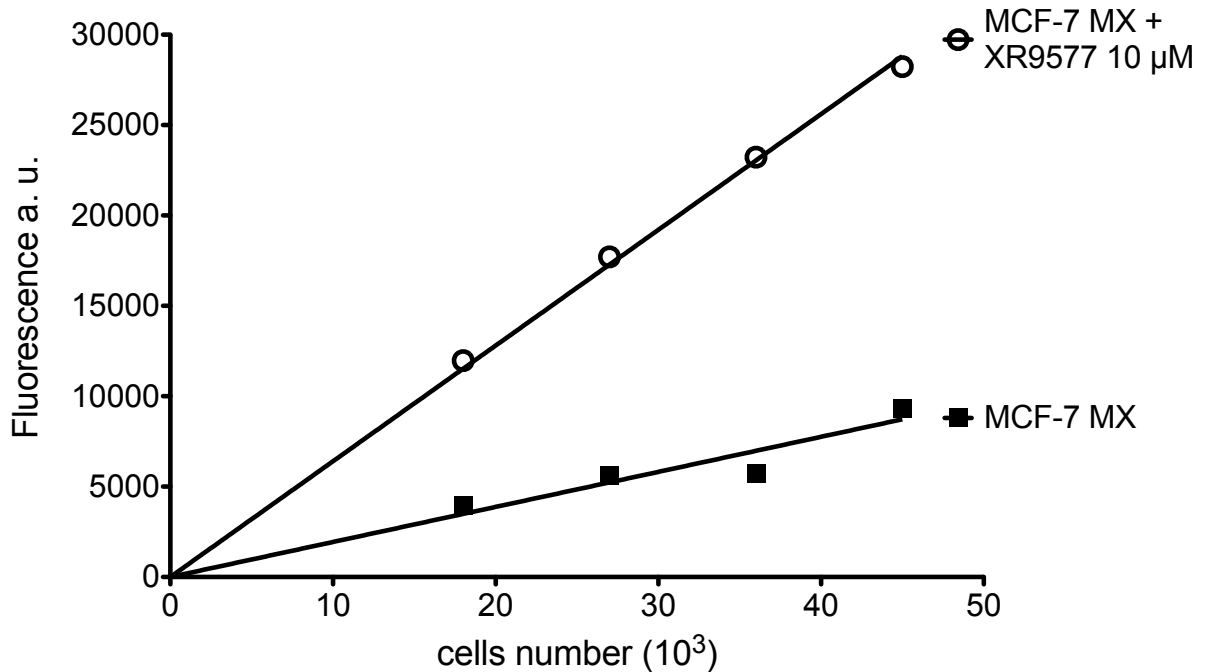


Figure 4.5: Dependence of cell number on the fluorescence in presence or absence of the modulator XR9577 at a concentration of 10 μ M (MCF-7 MX cells). The Hoechst 33342 concentration is equal to 1 μ M.

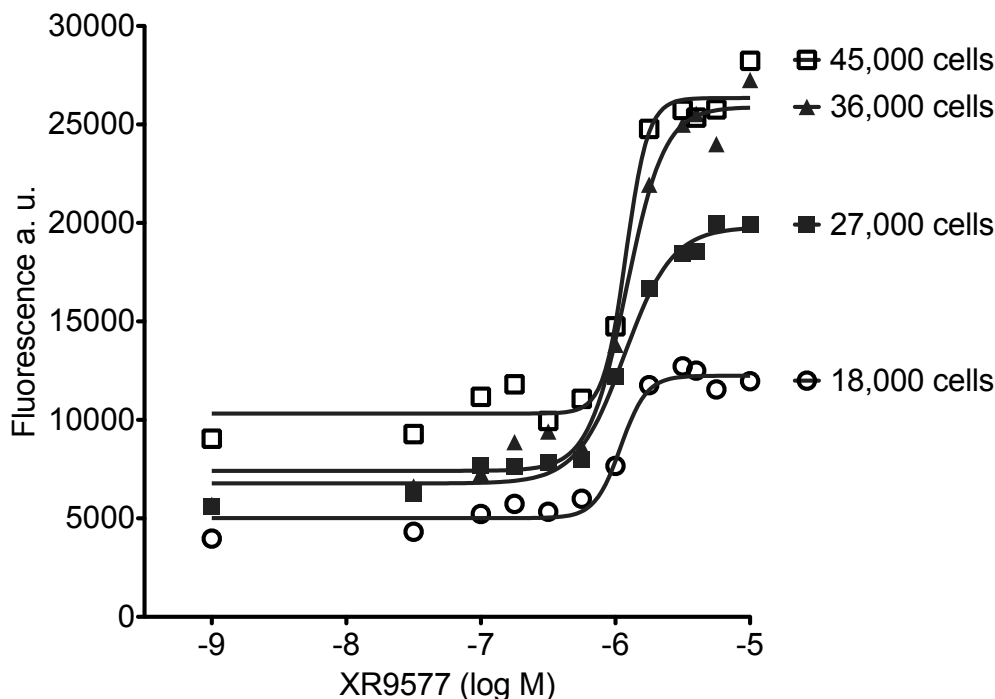


Figure 4.6: Dependence of the dose-response curve of the modulator XR9577 on the number of cells (MCF-7 MX cells). The used Hoechst 33342 concentration is equal to $1 \mu\text{M}$.

For cells numbers higher than 36,000 cells/well the fluorescence of the MCF-7 MX cells treated with Hoechst 33342 at a final concentration of $10 \mu\text{M}$ does not increase anymore with the cell number, but reaches a plateau. In figure 4.6 is shown how the cell number influences the dose response curve parameters. If the concentration of cells is higher than 36,000 cells/well, the top of the dose response curve does not change anymore with the cell number. A notable decrease of the top value of the dose response curve with respect to the curves obtained using 45,000 and 36,000 cells/well is seen for 27,000 cells/well. A further decrease of the span of the dose response curve is observed if only 18,000 cells/well are used. In conclusion, the use of a cell concentration higher than 36,000 cells/well leads to the saturation of the fluorescence values of the curve. In figure 4.5 is also shown that the quality of the curve obtained using 27,000 cells/well is better than that obtained using higher cells numbers. The use of cells concentrations lower than 36,000 cells/well is also to prefer considering the cost that the use of high cells concentrations implies in terms of laboratory materials.

On the other hand, cell concentrations lower than 20,000 cells/well lead to dose response curves with lower span, what complicates the use of this assay. For those two reasons, it was decided to use for the next assays a cell concentrations equal to 27,000 cells/well.

4.1.3 Analysis of the raw data obtained from the Hoechst 33342 assay

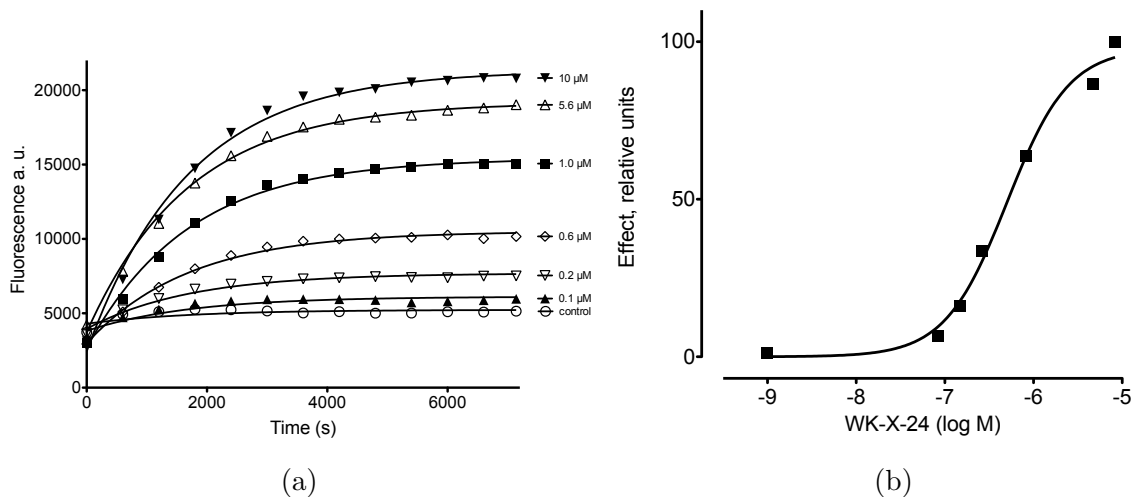


Figure 4.7: (a) Fluorescence against time curves for different concentrations of the modulator Hoechst 33342, generated with the Hoechst 33342 assay in MCF-7 MX cells. The fluorescence-time curves were analyzed using the one-phase association model. (b) Dose-response curve of XR9577 (WK-X-24) in MCF-7 MX cells obtained from the plateau values of the one-phase association model ($pIC_{50} = 6.30$, Hill Slope = 1.24).

In order to determine if XR9577 (WK-X-24) is able to increase the intracellular fluorescence of Hoechst 33342, different concentrations of this modulator were added to MCF-7 MX cells at a concentration of 27,000 cells/well. Afterwards, Hoechst 33342 was added to obtain a final concentration of 1 μM in each well. The fluorescence-time diagram was registered and analyzed. In figure 4.7a a typical fluorescence-time curves are shown, obtained with XR9577 as modulator and MCF-7 MX cells. Increasing concentrations of the modulator XR9577 also increase the fluorescence and the fluorescence reaches a maximum after 5000 seconds. The fluorescence-time curves were approximated by one-phase association functions. Their plateau values were plotted against the corresponding logarithmic concentrations in order to obtain the dose-response curve of the modulator. Alternatively, the last fluorescence values of the fluorescence-time curve (7200 s) can be used and plotted against the logarithmic concentration of the modulator. In figure 4.7b the obtained sigmoidal dose-response curve is reported.

4.2 Determination of the activity of a new class of BCRP-inhibitors.

4.2.1 Tariquidar analogues with modified anthranilic acid partial structure

Tariquidar (XR9576) is a potent third generation P-gp inhibitor and also a BCRP-inhibitor [147, 148]. Several Tariquidar analogues were synthesized in our laboratory by Dr. Werner Klinkhammer [149]. The structure of Tariquidar can be divided in six substructures, as illustrated in figure 4.8:

- The tetrahydroisoquinoline residue
- The first aromatic ring
- The first linker
- The second aromatic ring
- The second linker
- The third aromatic ring

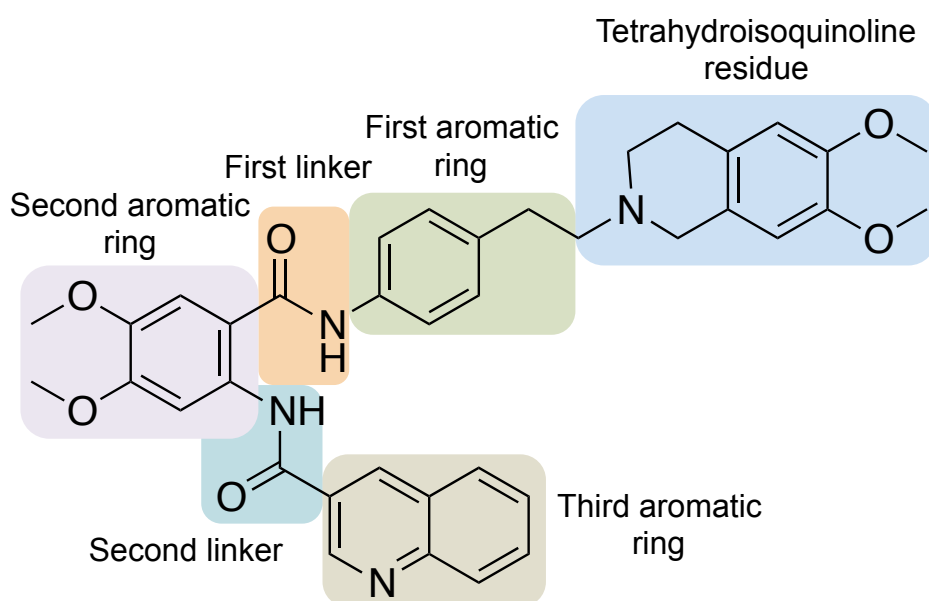


Figure 4.8: Elements of the Tariquidar structure.

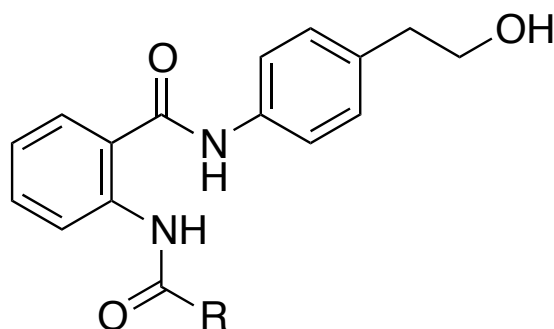


Figure 4.9: General structure of the selective BCRP-inhibitors synthesized by Werner Klinkhammer. R = phenyl (WK-X-7), R = 3-chinolinylyl (WK-X-8), R = 2-nitrophenyl (WK-X-9), R = 2-aminophenyl (WK-X-20), R = 4-nitrophenyl (WK-X-27, also known as KS-176), R = 4-methylphenyl (WK-X-28).

Dr. Werner Klinkhammer synthesized six compounds without the tetrahydroisoquinolin residue, that was replaced by a hydroxyl group. The biological results showed that these compounds were BCRP-inhibitors but not P-gp or MRP inhibitors [150]. The loss of P-gp inhibitory activity is probably due to the absence of the basic nitrogen of the tetrahydroisoquinoline moiety in these compounds. That seems to be important for the P-gp inhibitory activity of different classes of compounds [151]. At a later stage, Dr. Kerstin Steggeman synthesized more compounds of this new class of anthranile amide derivatives, with modifications of the above named substructures into which the structure of Tariquidar can be divided.

A goal of this work was the characterization of the BCRP-inhibitory activity of these compounds, using the Hoechst 33342 assay. The inhibitory activity against P-gp was also measured using the calcein-AM assay, in order to determine the selectivity of the compounds for BCRP. For selected compounds the capacity to reverse the resistance against cytostatic compounds of cells expressing BCRP was evaluated using the MTT assay. The MTT assay was also used to determine the cytotoxicity of selected compounds.

4.2.2 Characterization of the inhibitory activity of the compound WK-X-27

Between the selective inhibitors synthesized by Werner Klinkhammer, the most active is the compound WK-X-27 (KS-176). It shows an IC_{50} value of $1.58 \mu\text{M}$ and does not have an inhibitory effect on P-gp.

The inhibitory activity against BCRP of compound KS-176 was tested using the Hoechst 33342 assay, while the inhibitory activity for P-gp was tested using the calcein AM assay. As shown in figure 4.10, KS-176 is able to increase the accumulation of Hoechst 33342 in MCF-7 MX cells with increasing concentrations.

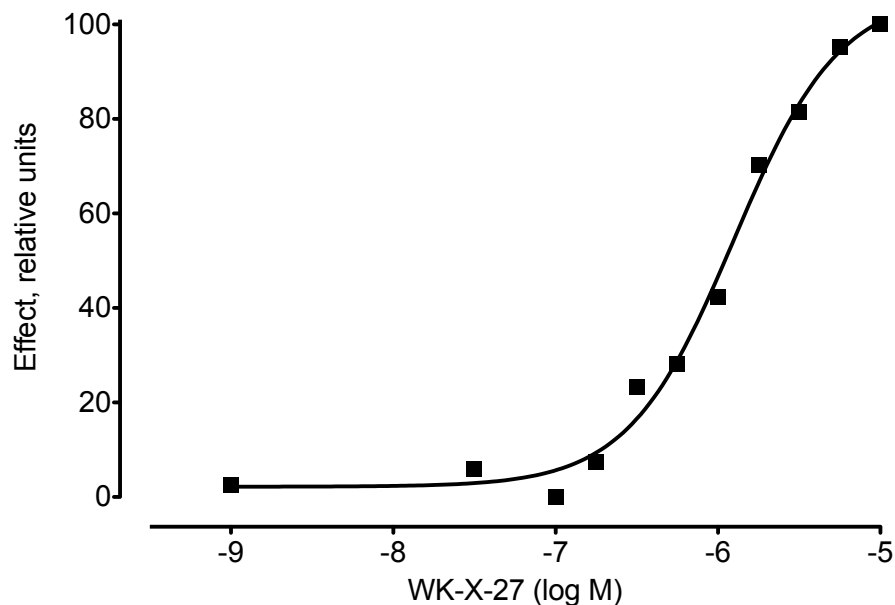


Figure 4.10: Dose-response curve of WK-X-27 (KS-176). The BCRP inhibitory activity was tested using the Hoechst 33342 assay and MCF-7 MX cells.

To investigate the reversal of BCRP-mediated resistance by KS-176, the toxicity of mitoxantrone and Hoechst 33342 were determined in the presence and absence of this compound. Cell viability was determined by the MTT assay. The cytotoxicity of mitoxantrone was determined in parental MCF-7 and resistant MCF-7 MX cells, while the cytotoxicity of Hoechst 33342 was determined in parental MDCK and resistant MDCK BCRP cells. MDCK BCRP cells were included because in this transfected cell line the resistance can be solely attributed to the presence of BCRP. The IC_{50} values of mitoxantrone were determined for parental MCF-7 cells and resistant MCF-7 MX cells without the addition of inhibitor and in the presence of two different inhibitor concentrations. Figure 4.11a shows the effect of compound KS-176 on the EC_{50} value of mitoxantrone in MCF-7 cells. In the presence of an inhibitor concentration of $5 \mu\text{M}$, a significant decrease in resistance is observed, and in presence of compound KS-176 at $10 \mu\text{M}$ the resistance is fully reversed and the cytotoxicity of mitoxantrone is restored, leading to a sensitivity of the MCF-7 MX cell line almost identical to that of the parental MCF-7 cells. This result shows that this compound is able to reverse the resistance of MCF-7 MX cells against mitoxantrone completely. Figure 4.11b shows the dose-response curve of Hoechst 33342 in the presence of compound KS-176. The shift in the dose-response curves of MDCK BCRP cells caused by progressively higher inhibitor concentrations of compound KS-176 indicates that this compound also dose-dependently inhibits BCRP-mediated resistance to Hoechst 33342. The results prove the conclusions that the inhibitory effect of compound KS-176 is not substrate dependent.

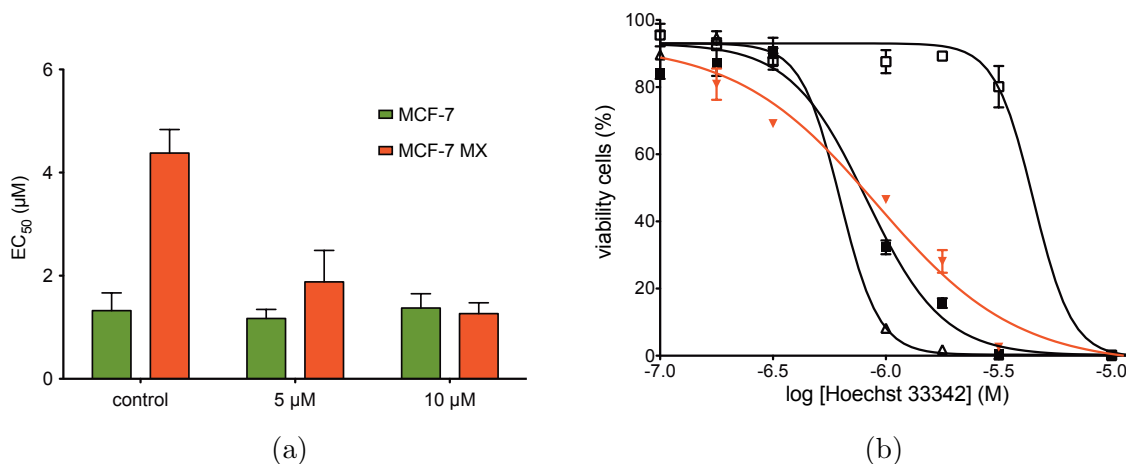


Figure 4.11: (a) The effect of compound KS-176 on the EC₅₀ value of mitoxantrone in MCF-7 cells (green) and MCF-7 MX cells (red) investigated at a concentration of 5 and 10 µM. Cells in absence of the inhibitor were used as control. Data is the average ±SD from three independent experiments. (b) Shift of the dose-response curve of Hoechst 33342 caused by increasing inhibitor concentration. MDCK BCRP cells (empty square), MDCK BCRP cells + 5 µM KS-176 (black square), MDCK BCRP cells + 10 µM KS-176 (empty triangle), parental MDCK cells (red down triangle)

4.2.3 Variations of the first linker

The first linker of the structure of this new class of BCRP inhibitors was modified in order to understand and clarify the structure activity relationship for this class of compounds. Compound KS-176 is used as reference. As reported in table 4.1, the variations of the first linker generally lead to low active or inactive compounds and only in one case enhancement of activity is observed. All compounds having modifications of the first linker, with exception of the compound KS-311, have also exclusion or substitution of the carboxylic group in the anthranilic acid substructure or insertion of one or more atoms between the carboxylic group and the aromatic ring. The direct binding of carbonyl group of the amide to the aromatic ring seems to be important for the activity of these compounds, probably because of the hydrogen bridge that forms between the nitrogen of the anthranile amide substructure and the oxygen of the carbonyl group. This intramolecular interaction could be important to maintain the correct conformation of the molecule in the binding site of the protein. The results showed also that for the modified compounds, the presence of rotatable linkers leads to active compounds, in respect to the not rotatable homologue. For example, compound KS-360, with the insertion of an ethenyl group between the second aromatic ring and the carboxylic group does not show any inhibitory activity at a concentration up to 10 µM. KS-366 is structurally similar to KS-360, but it possesses an ethyl group instead

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

of the ethenyl group. The cellular testing revealed that compound KS-366 is a BCRP inhibitor with an IC_{50} value of $7.58 \mu\text{M}$. An other example involves the compounds KS-364 and KS-357, which differ only by the presence of an amide nitrogen in the linker (compound KS-357). The presence of this amide nitrogen leads to loss of activity, with respect to compound KS-364 ($IC_{50} = 9.69 \mu\text{M}$). This amide nitrogen probably confers a higher rigidity to the linker as compared to compound KS-364 due to the mesomeric effect of the urea linker. The length of the linker also influences the activity of the compounds. Comparing compounds KS-366 and KS-364, which have an alkyl linker between the central aromatic ring and the carbonyl group of the amide, the inhibitory activity against BCRP increases with the length of the linker ($IC_{50} = 7.58 \mu\text{M}$ for KS-366 and $IC_{50} = 9.69 \mu\text{M}$ for KS-364). The presence of heteroatoms directly connected to the second aromatic ring, as for example in compounds KS-351 and KS-392 leads to a loss of activity. The best compound with a modified first linker is compound KS-311, with an ethyl group between the amide and the first aromatic group. This compound shows an IC_{50} value of $0.64 \mu\text{M}$, also lower than compound KS-176. In conclusion, the carbonyl group of the first linker is important for the activity of this class of compounds and must be directly connected to the second aromatic ring. The inhibitory activity against BCRP is also increased by insertion of an alkyl linker between the amide linker and the first aromatic ring, as in compound KS-311.

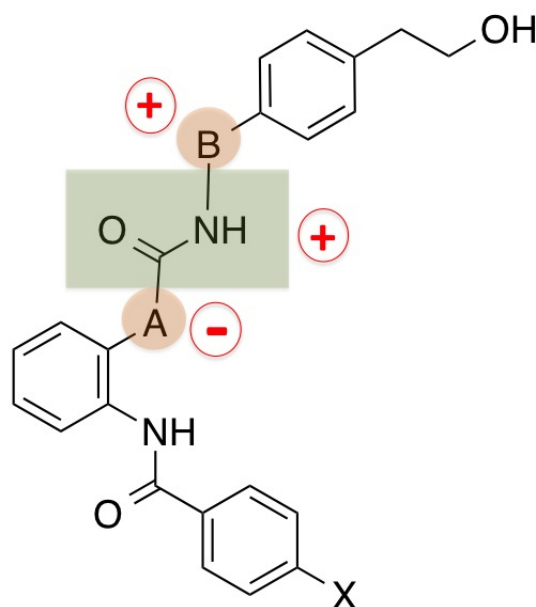


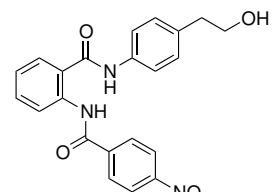
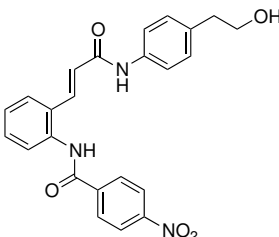
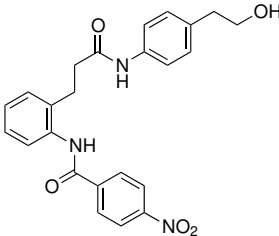
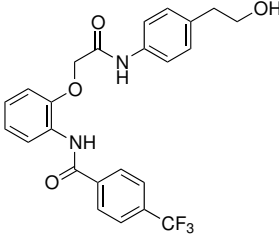
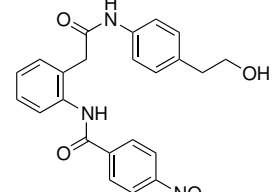
Figure 4.12: Summary of structural features of the investigated compounds influencing BCRP inhibitory activity. The symbols + and - are used to indicate if the considered feature plays a favorable or an unfavorable role for the activity of the compounds.

Additionally, the inhibitory activity against P-gp was investigated for those com-

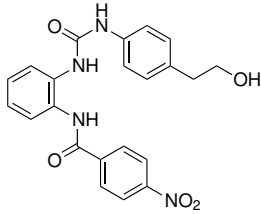
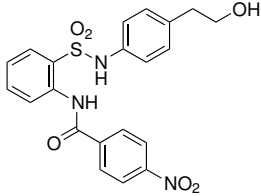
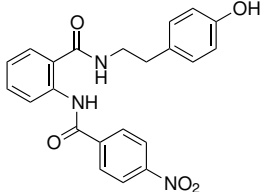
4.2 Determination of the activity of a new class of BCRP-inhibitors.

pounds. The results showed that the insertion of the alkyl group after (compound KS-311) or before (compound KS-366) the amide group of the linker leads to a weak P-gp inhibition. In case of compound KS-366, the IC_{50} is $14.5 \mu M$, slightly lower than of compound KS-311, that has an IC_{50} value of $25.4 \mu M$. Also compound KS-392, with sulfonamide group as first linker is a weak P-gp inhibitor, with an IC_{50} value of $20.8 \mu M$.

Table 4.1: Inhibitory activity of compounds with variations of the first linker. Activities were measured using Hoechst 33342 assay and MCF-7 MX cells for BCRP (reference substance WK-X-24: $IC_{50} = 0.90 \pm 0.07 \mu M$) and calcein AM assay and A2780 Adr cells for P-gp (reference substance WK-X-24: $IC_{50} = 0.41 \pm 0.08 \mu M$). n.a. = no activity for concentration up to $10 \mu M$.

Formula	Compound	IC_{50} BCRP $\mu M \pm$ SD	IC_{50} P-gp $\mu M \pm$ SD
	KS-176	1.58 ± 0.36	n.a.
	KS-360	n.a.	n.a.
	KS-366	7.58 ± 1.20	14.5 ± 3.9
	KS-351	n.a.	n.a.
	KS-364	9.69 ± 2.65	n.a.

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

Formula	Compounds	IC ₅₀ BCRP $\mu\text{M} \pm$ SD	IC ₅₀ P-gp $\mu\text{M} \pm$ SD
	KS-357	n.a.	n.a.
	KS-392	n.a.	20.8 \pm 5.9
	KS-311	0.644 \pm 0.149	25.4 \pm 7.7

4.2.4 Variations of the second linker

The modification of the second linker of this class of compounds does not strongly decrease their activity, as observed with the modification of the first linker. As shown in table 4.2, the insertion of an alkyl group between the amide and the third aromatic ring leads to loss of activity (compound KS-181). The structurally related compound KS-094, with an urea linker instead of the amide group, is a BCRP inhibitor with a IC_{50} value equal of $4.04 \mu M$. Compound KS-367 differs from compound KS-094 only by the presence of a trifluoromethyl group instead of the nitro group on the third aromatic ring. Its BCRP inhibitory activity is slightly lower than of compound KS-094 ($IC_{50} = 7.21 \mu M$). The difference of the activity between compounds KS-181 and KS-094 could be explained with the different orientation which the nitrogen of the urea linker confers to the third aromatic group.

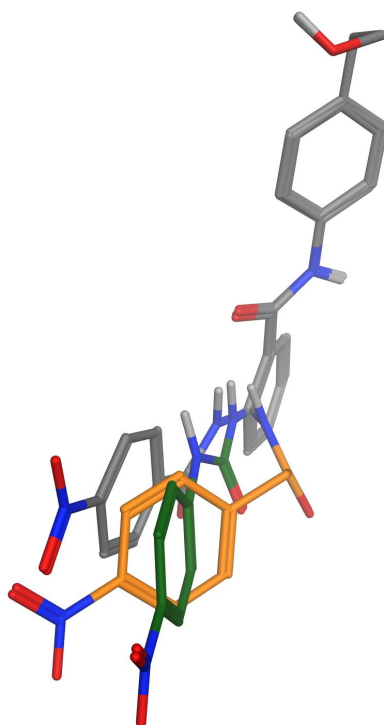


Figure 4.13: Superposition of the compound KS-176 (reference compound) with compounds KS-181 and KS-094. The orange substructure belongs to compound KS-181, while the green substructure belongs to compound KS-094.

To explain the remarkable difference in the activity of these three compounds, the three molecular structures of compounds KS-176, KS-181 and KS-094 were aligned first. Then, the common parts of the three compounds (the first aromatic ring, the first linker and the second aromatic ring) were fixed. Next, the unfixed parts of the structures (the second linker and the third aromatic ring) were minimized using PM3 (MOPAC) calculation. The flexible alignment, and the structure minimizations were performed using the program MOE [97]. The results of this calculation showed that

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

the compounds KS-176 and KS-094 have the same orientation of the third aromatic ring, while the orientation of the third aromatic ring for compound KS-181 is rotated by about 45° with respect to the axis of the other two compounds. This difference of orientation is due to the tetrahedral geometry of the alkyl linker, while the geometry of the amide and urea linkers is quite planar. This different ring orientation for compound KS-181 could cause loss of important non-covalent interactions of the compound with the binding site and consequently loss of the inhibitory activity. In support of this assumption, the introduction of an oxygen between the alkyl linker and the third aromatic ring, as in case of compound KS-187, restores the inhibitory potency against BCRP and the IC₅₀ value of this compound (7.69 μM) is comparable to that of compound KS-094. The BCRP inhibitory activity of compound KS-187 could be explained by the change of the orientation of the third aromatic ring with respect to compound KS-181. This new orientation is similar to that of compound KS-094. This hypothesis was confirmed by the superposition of the minimized structures of compound KS-187 and compound KS-094.

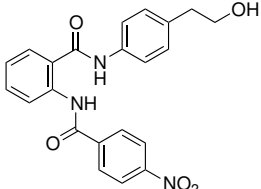
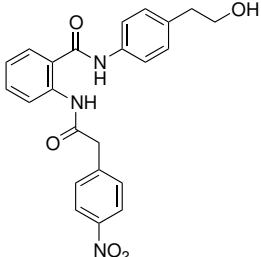
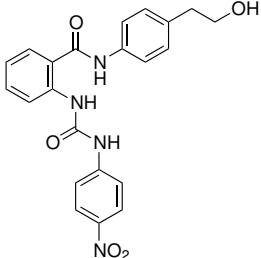
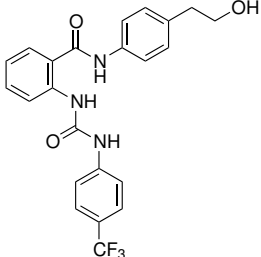
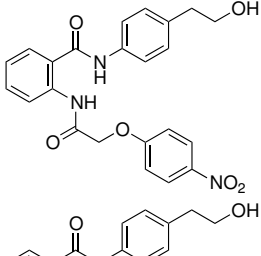
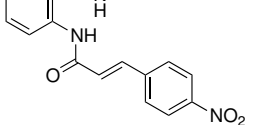
The three compounds with an ethenyl group between the amide and the third aromatic ring were found to be BCRP inhibitors. Compound KS-186, with a nitro group at *para* position on the third aromatic ring, is a good BCRP inhibitor, with an IC₅₀ value equal of 2.93 μM. Compound KS-257, with a chlorine as substituent at *para* position on the third aromatic ring is slightly more active than compound KS-186, with an IC₅₀ value of 1.85 μM, comparable to that of the reference compound KS-176. The last compound of this series with an ethenyl group in the second linker, compound KS-246, possesses an IC₅₀ value of 0.52 μM and is the most active compound in this group. This high inhibitory activity could be explained by the electronic effect of the two chlorine substituents at *ortho* positions on the third aromatic ring. This effect could modify the electronic density of the aromatic ring and the interactions of this ring with the binding site of BCRP.

The last compound of this series is KS-290, has an amino group as linker between the second and the third aromatic ring. This compounds is not a BCRP inhibitor, probably due to the disadvantageous orientation of the third aromatic ring or to the too low distance between the two aromatic rings.

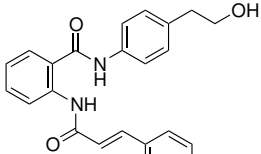
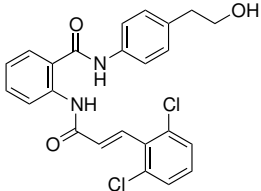
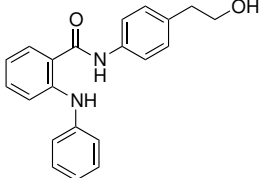
Regarding the inhibitory activity of these compounds against P-gp, only two of them have a detectable capacity to inhibit this protein. Both of these compounds have an ethenyl group in the second linker, between the amide and the third aromatic ring. The best P-gp inhibitor in this series is compound KS-186, with an IC₅₀ value equal of 6.62 μM. Compound KS-246 is also able to inhibit P-gp, but with an IC₅₀ value of only 11.0 μM. The third compound with an ethenyl group in the linker does not inhibit the transport mediated by P-gp.

4.2 Determination of the activity of a new class of BCRP-inhibitors.

Table 4.2: Activity of compounds with variations in the first linker. Activities were measured using Hoechst 33342 assay and MCF-7 MX cells for BCRP (reference substance WK-X-24: $IC_{50} = 0.90 \pm 0.07 \mu\text{M}$) and calcein AM assay and A2780 Adr cells for P-gp (reference substance WK-X-24: $IC_{50} = 0.41 \pm 0.08 \mu\text{M}$). n.a. = no activity for concentration up to $10 \mu\text{M}$.

Formula	Compound	IC_{50} BCRP $\mu\text{M} \pm$ SD	IC_{50} P-gp $\mu\text{M} \pm$ SD
	KS-176	1.58 ± 0.36	n.a.
	KS-181	n.a.	n.a.
	KS-094	4.04 ± 0.48	n.a.
	KS-367	7.21 ± 4.88	n.a.
	KS-187	7.69 ± 0.55	n.a.
	KS-186	2.93 ± 0.59	6.62 ± 0.98

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

Formula	Compounds	IC ₅₀ BCRP $\mu\text{M} \pm$ SD	IC ₅₀ P-gp $\mu\text{M} \pm$ SD
	KS-257	1.85 \pm 0.75	n.a.
	KS-246	0.52 \pm 0.10	11.03 \pm 4.24
	KS-290	n.a.	n.a.

4.2.5 Reduction of the scaffold size

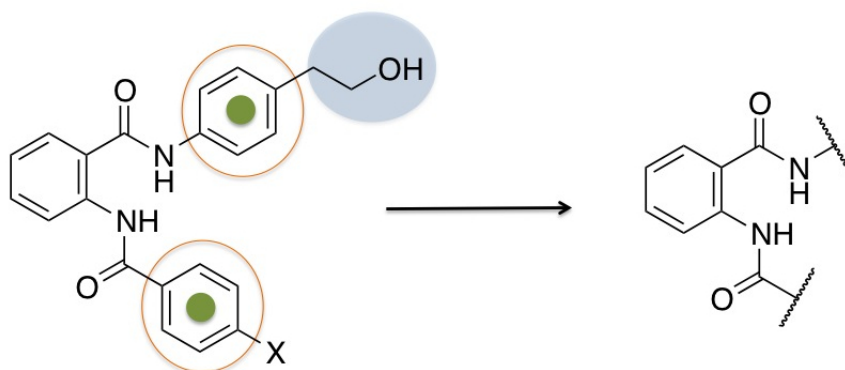


Figure 4.14: Diagram of the scaffold reduction for the antranile amide derivatives.

In order to verify if it is possible to further reduce the size of the scaffold of this class of BCRP inhibitors, a series of compounds without the first or third aromatic ring or with other substituents instead of the aromatic rings were synthesized by Dr. Steggemann.

The reduction process could be summarized in three phases and it is shown in figure 4.14. In the first phase the substituent on the first aromatic ring (highlighted in blue) was replaced by a hydroxy group directly bound to the aromatic ring. In the second phase the first and third aromatic rings (the green points in the figure) were replaced by aliphatic rings, in order to test if aromaticity at this position is really necessary for activity. In the third phase the first or third aromatic ring was deleted (highlighted in orange).

As shown in table 4.3, the cellular testing results showed that the replacement of hydroxyethyl group by a phenolic hydroxy group (compound KS-251) leads to high BCRP inhibitory activity, with an IC_{50} value for this compound equal of $1.76 \mu\text{M}$. This result suggests that the hydroxyethyl group is not necessary for the BCRP inhibitory activity of these compounds and could be simplified.

The deletion of the third aromatic ring or its replacement by an aliphatic cyclohexyl ring leads to loss of the inhibitory activity against BCRP. In particular, the complete deletion of the second linker with deletion of the third aromatic ring, as for compound KS-092, leads to loss of the BCRP inhibitory activity. The replacement of the third aromatic ring by a methyl group (compound KS-293) as well as by a cyclohexyl ring (compound KS-170) leads also to loss of the inhibitory activity for BCRP. In the light of these results, the third aromatic ring seems to be essential for the BCRP inhibitory activity of this class of compounds.

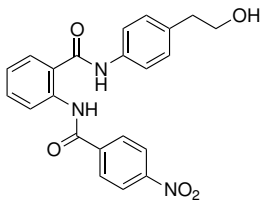
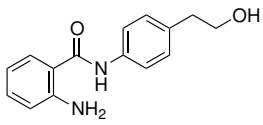
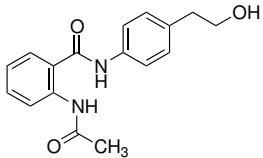
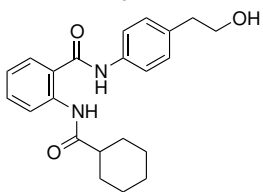
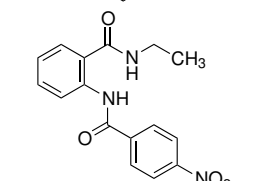
When the first aromatic ring is replaced by a 4-hydroxycyclohexyl group, the BCRP inhibitory activity decreases but does not disappear. This is the case for compounds

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

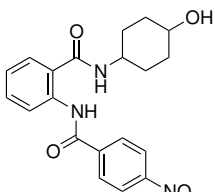
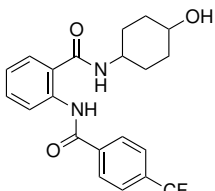
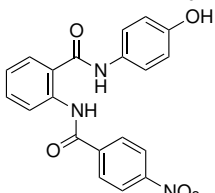
KS-305 and KS-304, with IC_{50} values of $6.11 \mu\text{M}$ and $3.35 \mu\text{M}$ respectively. Otherwise, the complete deletion and its replacement by an ethyl group, as for compound KS-328, leads to a loss of BCRP inhibitory activity. These results indicate that the first aromatic ring increases the BCRP inhibitory activity of this class of compounds but is not inherently essential for the activity of these inhibitors as it can be changed to a cyclic aliphatic group.

The calcein-AM assay results show that the last three compounds KS-305, KS-304 and KS-251 are also weak inhibitors of P-gp with IC_{50} values higher than $20 \mu\text{M}$. This result suggests that probably the hydroxyl group directly bound to the ring system leads to an increase of the P-gp inhibitory activity.

Table 4.3: Activity of compounds with reduction of the scaffold. Activities were measured using Hoechst 33342 assay and MCF-7 MX cells for BCRP (reference substance WK-X-24: $IC_{50} = 0.90 \pm 0.07 \mu\text{M}$) and calcein AM assay and A2780 Adr cells for P-gp (reference substance WK-X-24: $IC_{50} = 0.41 \pm 0.08 \mu\text{M}$). n.a. = no activity for concentration up to $10 \mu\text{M}$.

Formula	Compound	IC_{50} BCRP $\mu\text{M} \pm$ SD	IC_{50} P-gp $\mu\text{M} \pm$ SD
	KS-176	1.58 ± 0.36	n.a.
	KS-092	n.a.	n.a.
	KS-293	n.a.	n.a.
	KS-170	n.a.	n.a.
	KS-328	n.a.	n.a.

4.2 Determination of the activity of a new class of BCRP-inhibitors.

Formula	Compounds	IC ₅₀ BCRP $\mu\text{M} \pm$ SD	IC ₅₀ P-gp $\mu\text{M} \pm$ SD
	KS-305	6.11 \pm 2.58	51.8 \pm 12.1
	KS-304	3.35 \pm 1.06	20.8 \pm 4.4
	KS-251	1.76 \pm 0.18	27.5 \pm 1.9

4.2.6 Variations on the second aromatic ring

The variations on the second aromatic ring involved the positions 4 (R^1) and 5 (R^2) of the ring, with insertion of substituents such as chlorine, methyl and methoxyl groups on those positions. The third aromatic ring is substituted with 4-nitro, 4-trifluoromethyl or chlorine groups. Compound KS-274 has a naphthyl group instead.

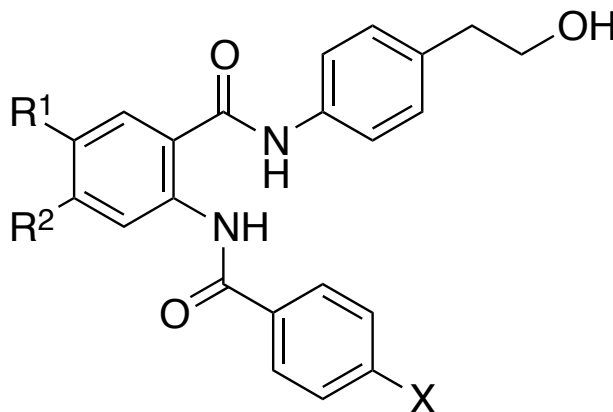


Figure 4.15: Positions of the substituents on the second aromatic ring.

As reported in table 4.4, in compounds with a 4-nitro substituent on the third aromatic ring, the presence of any substituent on the second ring decreases the inhibitory potency of these compounds relative to that of the reference compound, KS-176. The substitution at position R^2 causes a slight decrease in activity (compounds KS-191, KS-401, and KS-282), whereas substitution at R^1 leads to a more remarkable decrease in activity (compounds KS-206, KS-399, and KS-249).

The presence of a methoxy group at position R^2 (compound KS-191) results in a slight decrease in activity ($IC_{50} = 3.93 \mu\text{M}$), but at position R^1 the presence of the same substituent (in KS-206) decreases the inhibitory activity by about 5 fold ($IC_{50} = 8.43 \mu\text{M}$).

The same trend is observed with methyl substituents: A methyl group at position R^2 (compound KS-401) decreases the activity (with respect to compound KS-176) to an IC_{50} value of $2.69 \mu\text{M}$, whereas the same substitution at position R^1 (compound KS-399) gives an even less active compound, with an IC_{50} of $4.08 \mu\text{M}$. Compounds with chlorine substituents at positions R^1 (KS-249) or R^2 (KS-282) have similar inhibitory effects against BCRP (IC_{50} values of 3.09 and $2.59 \mu\text{M}$, respectively).

As shown in figure 4.16 the activity of the substituted compounds decreases linearly with increasing van der Waals volume of the substituted aromatic ring, as calculated with the software MOE [97]. This unfavorable steric effect is more evident at position R^1 than at position R^2 .

Compound KS-196, which has methyl groups at both positions and compound KS-274 with a naphthol group instead are slightly more active than expected. In case of

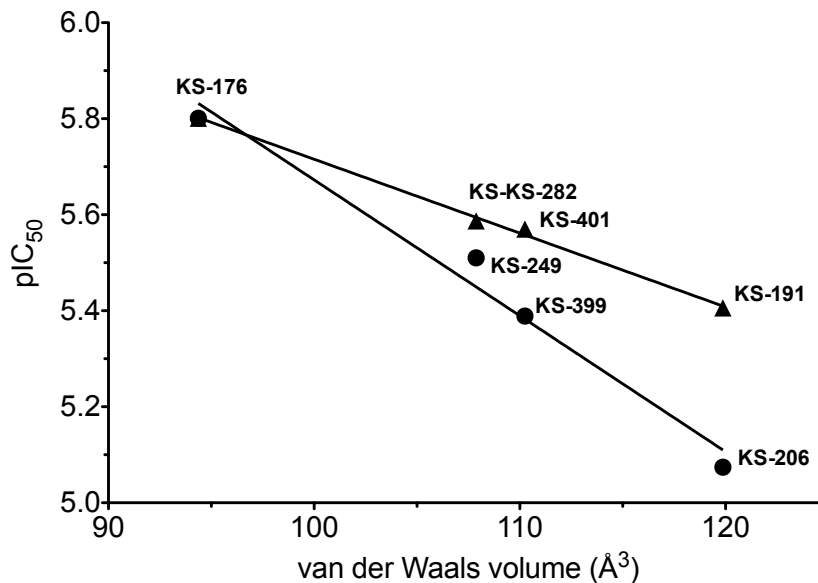


Figure 4.16: Plot of the van der Waals volumes of the substituted second aromatic ring versus pIC_{50} values for compounds with a substituent on the second aromatic ring at position R^1 (circles) and with substitution at position R^2 (triangles).

compound KS-196, this effect could be explained by the predominance of the steric effect at position R^1 over the effect at position R^2 or by additional interaction of the compound with the protein. The activity of compound KS-274 could be explained with an extra interaction of the additional aromatic ring with the binding site. Compounds KS-308, and KS-294 and KS-255 possess IC_{50} values lower than expected and they do not follow the QSAR observed for the nitro series of compounds. These exceptions could be explained from the plot of the three principal components of the 2D and 3D molecular descriptors calculated for the compounds containing substituents on the second aromatic ring. It is seen that the compounds included in the QSAR are grouped into a cluster, while the molecules that do not follow the QSAR, including compounds KS-196 and KS-274, are outside the cluster.

Regarding the principal component analysis, the first component is high correlated with the van der Waals surface area, while the second component is high correlated with the lipophilicity. For the other compounds, the relevant modification of those two parameters in respect to the reference compound could lead to a different kind of interaction with the transporter or with the membrane and resulting to a consequential effect on the IC_{50} values.

All together, the SAR of the second aromatic ring may be explained by a disadvantageous steric interaction of the substituents with the binding site of the protein. This explanation of the activity differences is possible only for compounds with similar global dimension and lipophilicity values.

The only two compounds of this group possessing an inhibitory activity against P-

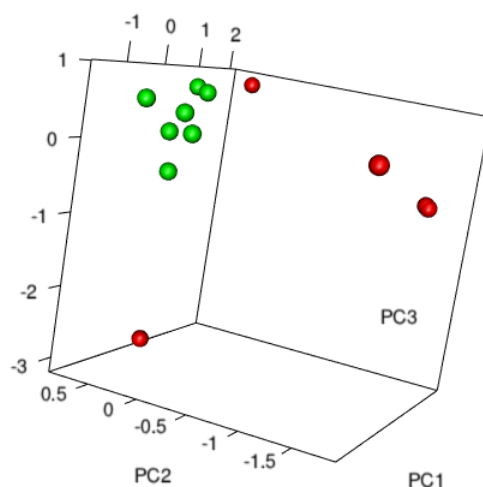
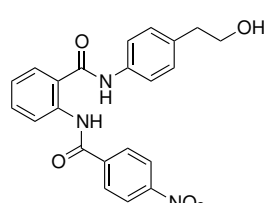
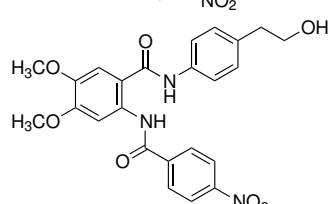


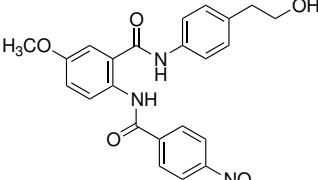
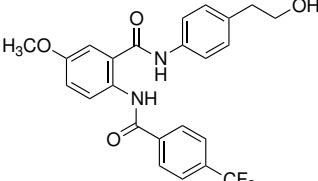
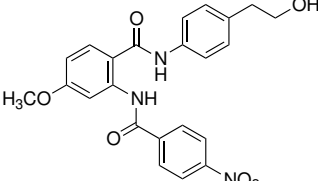
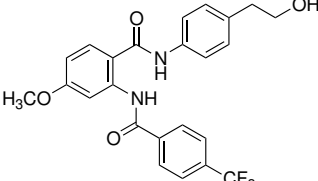
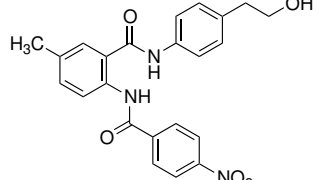
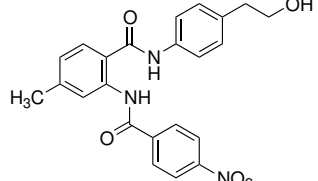
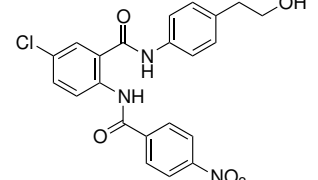
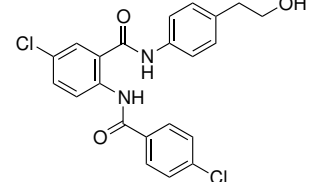
Figure 4.17: The chemical space of the compounds with a substituent on the second aromatic ring, obtained by plotting the three principal components of the 2D and 3D molecular descriptors calculated using MOE. The green spheres are associated with the compounds included in the QSAR, while the red spheres are associated with the compounds not included in the QSAR.

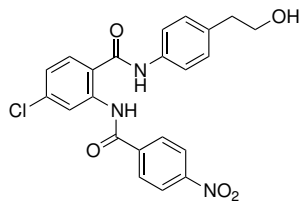
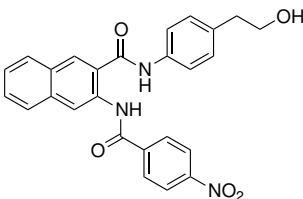
gp are the compounds KS-294 and KS-255, with IC_{50} values of $18.1 \mu\text{M}$ and $2.54 \mu\text{M}$ respectively. This effect could be related to the high lipophilicity of the compounds.

Table 4.4: Activity of compounds with variations on the second aromatic ring. Activities were measured using Hoechst 33342 assay and MCF-7 MX cells for BCRP (reference substance WK-X-24: $IC_{50} = 0.90 \pm 0.07 \mu\text{M}$) and calcein AM assay and A2780 Adr cells for P-gp (reference substance WK-X-24: $IC_{50} = 0.41 \pm 0.08 \mu\text{M}$). n.a. = no activity for concentration up to $10 \mu\text{M}$.

Formula	Compound	IC_{50} BCRP $\mu\text{M} \pm$ SD	IC_{50} P-gp $\mu\text{M} \pm$ SD
	KS-176	1.58 ± 0.36	n.a.
	KS-196	6.96 ± 2.20	n.a.

4.2 Determination of the activity of a new class of BCRP-inhibitors.

Formula	Compounds	IC ₅₀ BCRP $\mu\text{M} \pm$ SD	IC ₅₀ P-gp $\mu\text{M} \pm$ SD
	KS-206	8.43 \pm 1.63	n.a.
	KS-308	0.928 \pm 0.308	n.a.
	KS-191	3.93 \pm 0.11	n.a.
	KS-294	4.02 \pm 1.48	18.1 \pm 3.1
	KS-399	4.08 \pm 0.48	n.a.
	KS-401	2.69 \pm 0.86	n.a.
	KS-249	3.09 \pm 0.22	n.a.
	KS-255	1.54 \pm 0.22	2.54 \pm 0.29

Formula	Compounds	IC ₅₀ BCRP $\mu\text{M} \pm$ SD	IC ₅₀ P-gp $\mu\text{M} \pm$ SD
	KS-282	2.59 ± 0.87	n.a.
	KS-274	4.58 ± 0.35	n.a.

4.2.7 Variations on the third aromatic ring

The activity of the compounds with modifications on the third aromatic ring (reported in table 4.5) is correlated with the electronic effect of the substituents present on this ring and with the lipophilicity of the aromatic system. The best correlation uses the σ -Hammett parameter and the hydrophobic surface area of the substituted ring (equation 4.1). The σ values were selected with the amide group as reference position. $V_{\text{surf_S}}$ is the interaction surface area of the substituted aromatic ring calculated with MOE [97]. The relative importance of the two descriptors is 1 for σ and 0.963 for $V_{\text{surf_S}}$. Figure 4.18 shows the plot of observed versus calculated pIC_{50} values.

$$\text{pIC}_{50} = 0.503 \times \sigma_{p+m} + 0.00932 \times V_{\text{surf_S}} + 3.010 \quad (4.1)$$

$$n = 13, r^2 = 0.81, s = 0.156, F = 45.9$$

A possible explanation for this trend could be given by the interaction of the aromatic ring system at position R² with aromatic residues of the protein binding site and the consequential π - π stacking. Furthermore, the presence of bulky substituents with a large contact surface area on this ring increases the affinity of the ligands and hence the inhibitory activity. However, two compounds were excluded from the correlation. The first of these excluded compounds, KS-405, is less active than calculated from the equation. A hypothesis suggested by the similar activity values of compounds KS-405 and KS-406 is the possible hydrolysis of the ester group to a phenolic hydroxy group under the assay conditions. To test this hypothesis, TLC was performed with compound KS-405 after various incubation times in assay buffer at 37°C. After 10 minutes

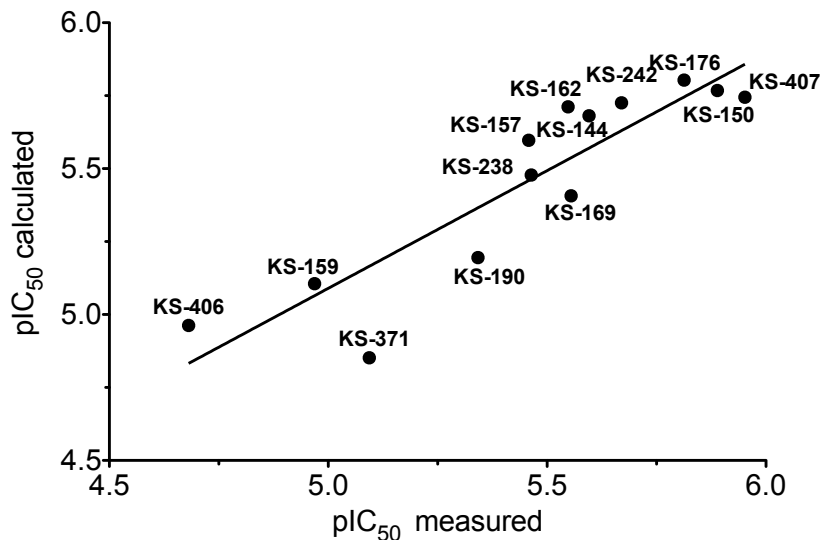


Figure 4.18: Plot of measured versus calculated potencies for inhibition of BCRP in MCF-7 MX cells according to equation 6.1 for compounds with variation on the third aromatic ring.

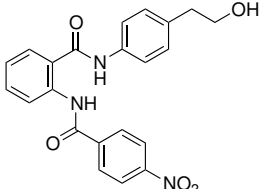
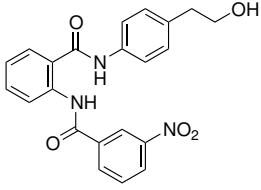
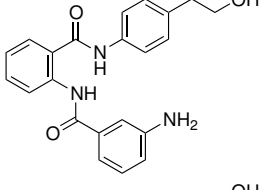
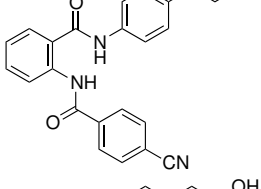
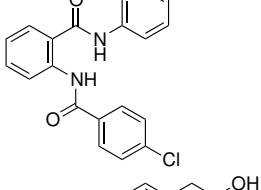
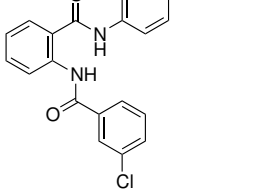
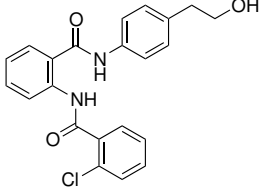
incubation, the original spot disappeared and a new spot was observed with the same R_f value as that of compound KS-406. Thus it can be concluded that compound KS-405 is not stable under the assay conditions. The second outlier from the correlation is compound KS-237, with a chlorine group in *ortho* position of the third aromatic ring. This compound was found to be not active against BCRP. A possible explanation could be an interaction (steric and electrostatic) of the chlorine group with the oxygen of the amide of the second linker. This interaction would change the orientation of the aromatic ring.

Regarding the selectivity of these compounds, most of them were not able to inhibit P-gp. However, some exceptions were observed. The highest inhibitory activity against P-gp was measured for compound KS-157, which contains a dimethoxy substituent. This is in agreement with P-gp inhibitors from different structural classes, for which this substitution pattern was found to be advantageous [152].

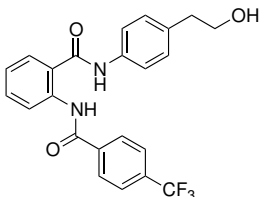
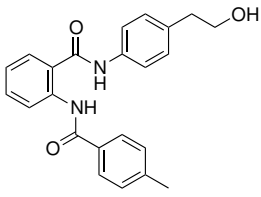
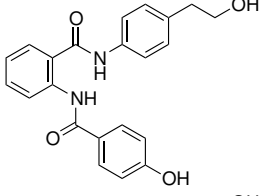
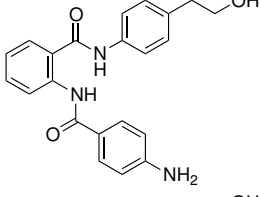
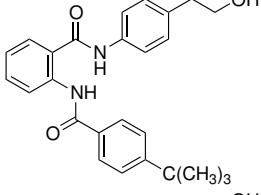
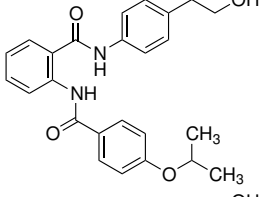
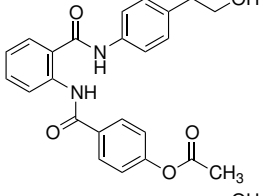
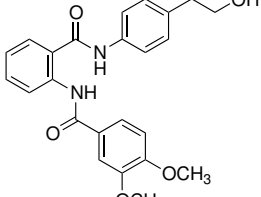
The presence of a voluminous lipophilic group on the third aromatic ring, as in compounds KS-162 and KS-407, results in a weak inhibitory activity against P-gp. The amino derivatives KS-159 and KS-371 also show some inhibition of P-gp, while from the corresponding nitro compounds only the 3-nitro derivative showed some, albeit low inhibitory activity against P-gp. Very low inhibitory activity against P-gp was also observed for compound KS-237 ($IC_{50} = 18.8$).

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

Table 4.5: Activity of compounds with variations on the third aromatic ring. Activities were measured using Hoechst 33342 assay and MCF-7 MX cells for BCRP (reference substance WK-X-24: $IC_{50} = 0.90 \pm 0.07 \mu\text{M}$) and calcein AM assay and A2780 Adr cells for P-gp (reference substance WK-X-24: $IC_{50} = 0.41 \pm 0.08 \mu\text{M}$). n.a. = no activity for concentration up to $10 \mu\text{M}$.

Formula	Compound	IC_{50} BCRP $\mu\text{M} \pm$ SD	IC_{50} P-gp $\mu\text{M} \pm$ SD
	KS-176	1.58 ± 0.36	n.a.
	KS-150	1.31 ± 0.25	21.28 ± 0.29
	KS-159	8.65 ± 2.37	14.0 ± 4.2
	KS-144	2.54 ± 0.20	n.a.
	KS-169	3.09 ± 1.32	n.a.
	KS-238	3.48 ± 0.54	n.a.
	KS-237	n.a.	18.8 ± 1.3

4.2 Determination of the activity of a new class of BCRP-inhibitors.

Formula	Compounds	IC ₅₀ BCRP $\mu\text{M} \pm$ SD	IC ₅₀ P-gp $\mu\text{M} \pm$ SD
	KS-242	2.20 \pm 0.57	n.a.
	KS-190	4.59 \pm 0.57	n.a.
	KS-406	21.2 \pm 3.7	n.a.
	KS-371	8.12 \pm 1.00	15.5 \pm 2.1
	KS-162	2.95 \pm 0.83	11.7 \pm 3.9
	KS-407	1.16 \pm 0.29	8.42 \pm 2.45
	KS-405	12.8 \pm 2.1	n.a.
	KS-157	3.61 \pm 0.96	7.75 \pm 1.19

4.2.8 Variations on the first aromatic ring

The ethyl alcohol on the first aromatic ring was modified especially for compounds possessing 4-nitro, 4-trifluoromethyl and chlorine groups on the third aromatic ring. The results are reported in table 4.6. The cellular test results showed that the substitution of the hydroxyethyl group with a phenolic hydroxy group does not significantly change the activity of the compound with respect to the reference compound KS-176 (compound KS-251, $IC_{50} = 1.76 \mu M$). The same effect was observed for the pair of compounds containing a 4-trifluoromethyl substituent on the third aromatic ring. The inhibitory activity of compound KS-342 ($IC_{50} = 1.80 \mu M$), with a hydroxy group on the first aromatic ring is very similar to that of compound KS-242 ($IC_{50} = 2.20 \mu M$), its homologue with the hydroxyethyl group at the same position. The same trend is also observed for compounds possessing a chlorine group at position 4 on the third aromatic ring: The inhibitory activity of the phenolic compound KS-252 ($IC_{50} = 2.91 \mu M$) is very similar to that of its homologue KS-169, with the hydroxyethyl group ($IC_{50} = 3.09 \mu M$). Exceptions from this trend are compounds with 3,4-dimethoxy substituents on the third aromatic ring. For this subclass of compounds, the exchange of the hydroxyethyl group by a phenolic hydroxy group leads to an increase of the inhibitory activity against BCRP, as observed for compound KS-265 ($IC_{50} = 1.16 \mu M$, three fold more potent than compound KS-157).

The methylation of the phenolic hydroxy group in *para* position on the first aromatic ring reduces the BCRP inhibitory activity to an IC_{50} value of $3.56 \mu M$ (compound KS-267). For compounds with 3,4-dimethoxy substitution on the third aromatic ring, the introduction of a methoxy group also on the first aromatic ring does not change the inhibitory activity against BCRP ($IC_{50} = 3.34 \mu M$ for compound KS-266 against $IC_{50} = 3.61 \mu M$ for compound KS-157. The structure and activity of compound KS-157 are reported in table 4.1.). If the third aromatic ring is substituted with a 4-chlorine group, the introduction of a methoxyl group on the first aromatic ring leads to loss of activity.

The methylation of the hydroxyethyl group leads to opposite results for compounds possessing 4-nitro and 4-trifluoromethyl groups. Indeed, compound KS-292 has an IC_{50} value of $1.23 \mu M$, a bit lower than the reference compound, while compound KS-322 has an IC_{50} value of $5.11 \mu M$, two fold higher than that the homologue KS-242.

The acetylation of the hydroxyethyl group was investigated for the compound with a 4-nitro group on the third aromatic ring (compound KS-194). This compound has an IC_{50} value of $1.32 \mu M$, slightly lower than the reference compound KS-176.

The increase of inhibitory activity was obtained with the replacement of the hydroxyethyl group by a benzoyl group. The two compounds with this substitution on the first aromatic ring, KS-174 and KS-184, contain a 4-nitro and a methyl group

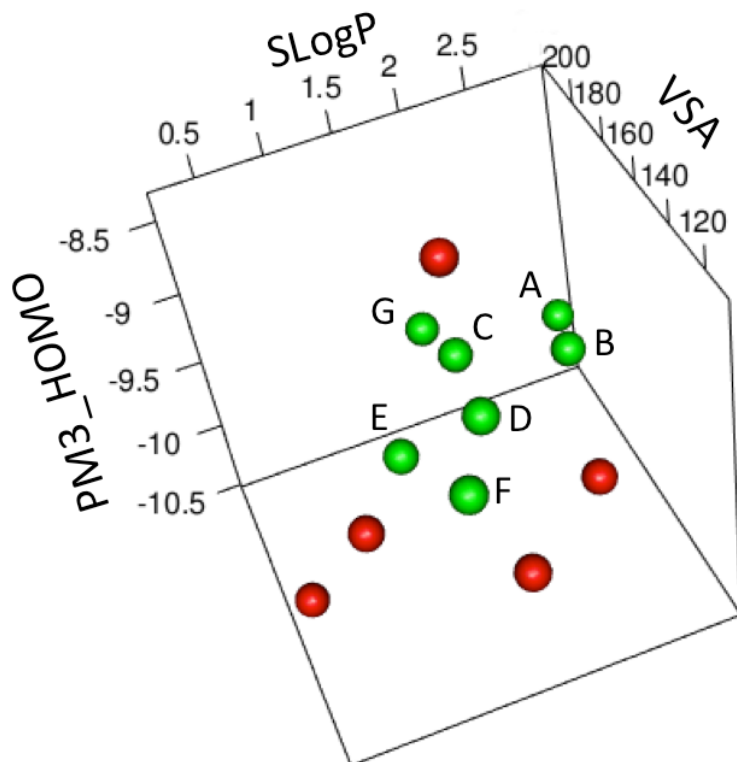


Figure 4.19: 3D plot of the SlogP, PM3_HOMO and VSA descriptors of the substituted first aromatic ring. The red points are associated to substituted aromatic rings present in compounds without inhibitory activity against BCRP, if the substituent on the third aromatic ring is a 4-nitro group. The green points are related to substituted aromatic rings of active BCRP inhibitors, if the third aromatic ring has a 4-nitro group, and correspond to the compounds: (A) KS-174, (B) KS-385, (C) KS-292, (D), KS-267, (E) KS-176, (F) KS-251, (G) KS-194.

respectively on the third aromatic ring. The IC_{50} value for compound KS-174 is equal to $1.63 \mu\text{M}$, slightly lower than the reference compound KS-176, while the IC_{50} value for compound KS-184 is equal to $0.56 \mu\text{M}$, about 8-folds lower than the corresponding compound with the hydroxyethyl group at the same position (KS-190, $IC_{50} = 4.59 \mu\text{M}$).

The absence of substituents on the first aromatic ring leads almost always to not active compounds or to compounds with very low inhibitory activity. The exceptions are the compound KS-166, with two methoxy groups on the third aromatic ring compound KS-173, with an acetamide at position 4 of the third aromatic ring. The IC_{50} -values of these compounds against BCRP are $1.76 \mu\text{M}$ and to $2.09 \mu\text{M}$, respectively.

Further compounds with a 4-nitro group on the third aromatic ring and bromine, sulfonamide, dimethylamino and amide groups on the first aromatic ring were synthesized by Dr. Steggemann. For these compounds no inhibitory activity against BCRP was found. The structure-activity relationship of the first aromatic ring can be explained by the 3D plot of SlogP, PM3_HOMO and VSA descriptors (figure 4.19).

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

These three molecular descriptors are not correlated with each other.

The most active compounds have groups, on the first aromatic ring, characterized by high values of SlogP. The increase of the lipophilicity must be also accompanied by an increase of the van der Waals surface area. As illustrated in figure 4.19, these compounds must also have values of PM3_HOMO (a descriptor high correlated with σ -Hammett) for the first aromatic ring between values of -9 eV and -10 eV. The two compounds (KS-383 and KS-374) out of this range are not active against BCRP. It can be concluded that for obtaining compounds with high inhibitory activity against BCRP, the first aromatic ring must be substituted by a lipophilic group with high van der Waals surface area. This conclusion is valid for compounds with 4-nitro substituent on the third aromatic group and can be extended also to compounds possessing other substituents on this position as trifluoromethyl and chloro.

As already mentioned, other compounds with the third aromatic ring substituted by a 3,4-dimethoxy group or by acetamide are an exception to this trend. A possible explanation of the exceptional activity of these compounds can be found observing the van der Waals surface (VSA) of the third aromatic ring for compounds with no substitution or with only a methyl group on the first aromatic ring. Compounds with IC₅₀ values higher than 10 μ M or no activity against BCRP have low values of van der Waals surface areas of between 126.79 (the aniline of compound KS-172) and 142.19 (the trifluoromethyl benzene of compound KS-358). The van der Waal surface area value of the third aromatic ring of compound KS-173 is equal to 167.09. This compound is a BCRP inhibitor with an IC₅₀ value of 2.09 μ M. The value of van der Waals surface area for compound KS-166 is equal to 171.26 and its IC₅₀ value is equal to 1.76 μ M. The importance of the surface area of the third aromatic ring for compounds without substituents on the first aromatic ring can be explained by the increase of the importance of the interaction in which the third aromatic ring is involved after disappearance of the interaction of the substituent on the first aromatic ring with the binding site. Indeed, it was already demonstrated, that a bulky substituent on the third aromatic ring contributes to an increase in the inhibitory activity against BCRP for this class of inhibitors.

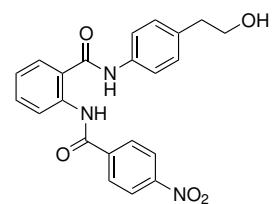
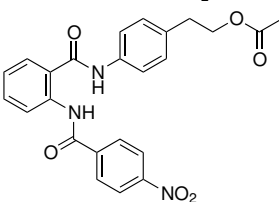
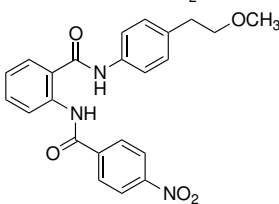
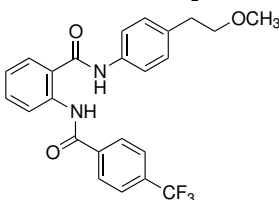
Regarding the inhibitory activity against P-gp, the absence of substituents on the first aromatic ring leads to compound KS-166, with an IC₅₀ value equal to 3.98 μ M. The introduction of an hydroxy or methoxy substituent on the first aromatic ring does not really change the inhibitory activity against P-gp with respect to the compounds without substituents at the same position. Indeed, the IC₅₀ value of compound KS-266, with a methoxy group on the first aromatic ring and the IC₅₀ of compound KS-265, with hydroxy group at the same position are respectively equal to 4.09 μ M and 4.47 μ M, respectively.

A severe reduction of the inhibitory activity against P-gp was measured for com-

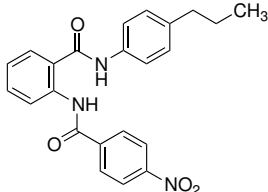
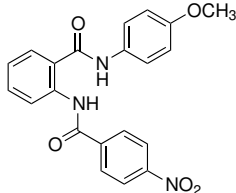
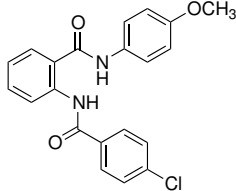
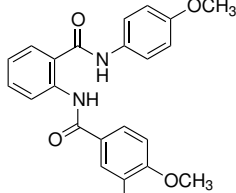
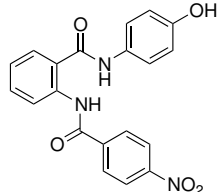
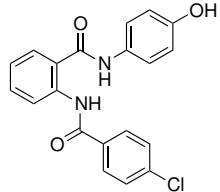
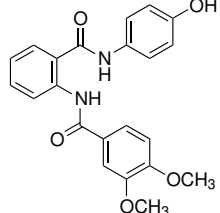
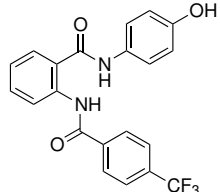
4.2 Determination of the activity of a new class of BCRP-inhibitors.

compound KS-277, with a methyl group on the first aromatic ring. This compound has an IC_{50} value of $13.5 \mu\text{M}$. The presence of a chlorine group at the same position (compound KS-272) leads to a loss of activity. For compounds with a nitro group or with a trifluoromethyl group on the third aromatic ring, the only active compounds are those with a hydroxy group on the first aromatic ring. These compounds possess low inhibitory activity against P-gp, with IC_{50} values of $27.5 \mu\text{M}$ and $14.0 \mu\text{M}$ for compounds KS-251 and KS-342 respectively. The presence of a benzoyl group on the first aromatic ring leads to compounds with moderate inhibitory activity against P-gp: Indeed, the IC_{50} values of compounds KS-174 and KS-184 are respectively equal to 5.18 and $6.07 \mu\text{M}$, respectively. This result was expected, according to what was already observed by Chiba et. al. [153].

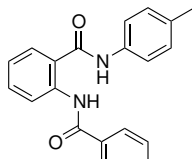
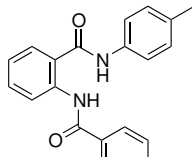
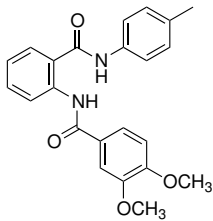
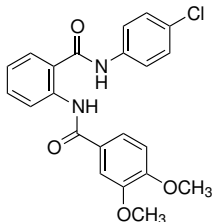
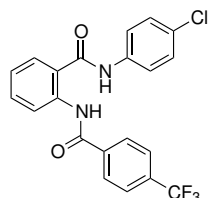
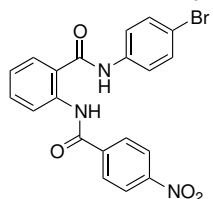
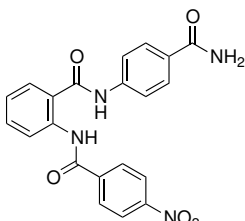
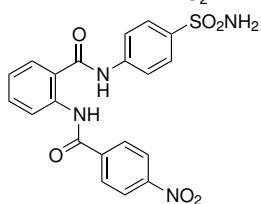
Table 4.6: Activity of compounds with variations on the first aromatic ring. Activities were measured using Hoechst 33342 assay and MCF-7 MX cells for BCRP (reference substance WK-X-24: $IC_{50} = 0.90 \pm 0.07 \mu\text{M}$) and calcein AM assay and A2780 Adr cells for P-gp (reference substance WK-X-24: $IC_{50} = 0.41 \pm 0.08 \mu\text{M}$). n.a. = no activity for concentration up to $10 \mu\text{M}$.

Formula	Compound	IC_{50} BCRP $\mu\text{M} \pm$ SD	IC_{50} P-gp $\mu\text{M} \pm$ SD
	KS-176	1.58 ± 0.36	n.a.
	KS-194	1.32 ± 0.28	n.a.
	KS-292	1.23 ± 0.40	n.a.
	KS-322	5.11 ± 0.87	n.a.

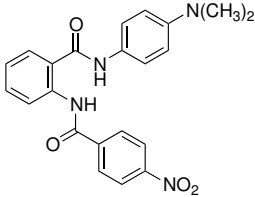
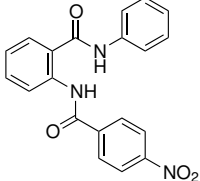
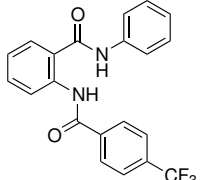
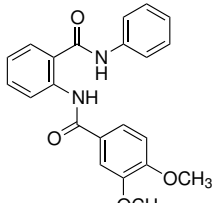
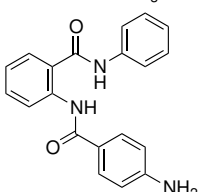
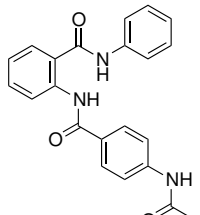
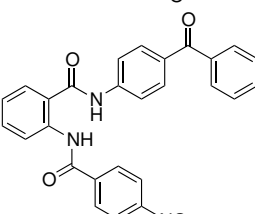
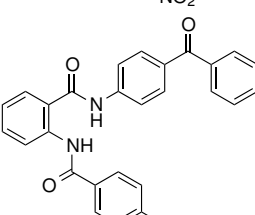
4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

Formula	Compounds	IC ₅₀ BCRP $\mu\text{M} \pm$ SD	IC ₅₀ P-gp $\mu\text{M} \pm$ SD
	KS-385	0.94 \pm 0.33	n.a.
	KS-267	3.56 \pm 0.84	n.a.
	KS-268	n.a.	n.a.
	KS-266	3.34 \pm 0.25	4.09 \pm 0.50
	KS-251	1.76 \pm 0.18	27.5 \pm 1.9
	KS-252	2.91 \pm 0.88	n.a.
	KS-265	1.16 \pm 0.21	4.47 \pm 1.00
	KS-342	1.80 \pm 0.43	14.0 \pm 3.5

4.2 Determination of the activity of a new class of BCRP-inhibitors.

Formula	Compounds	IC ₅₀ BCRP $\mu\text{M} \pm$ SD	IC ₅₀ P-gp $\mu\text{M} \pm$ SD
	KS-231	n.a.	n.a.
	KS-228	n.a.	n.a.
	KS-277	22.8 \pm 2.8	13.5 \pm 2.6
	KS-272	8.29 \pm 3.10	n.a.
	KS-375	n.a.	n.a.
	KS-382	n.a.	n.a.
	KS-381	n.a.	n.a.
	KS-383	n.a.	n.a.

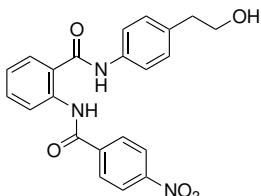
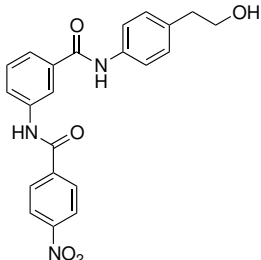
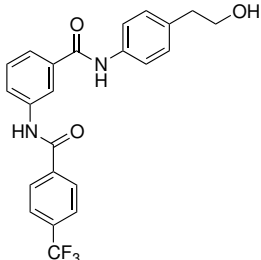
4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

Formula	Compounds	IC ₅₀ BCRP $\mu\text{M} \pm$ SD	IC ₅₀ P-gp $\mu\text{M} \pm$ SD
	KS-374	n.a.	n.a.
	KS-168	13.4 \pm 4.2	n.a.
	KS-358	n.a.	n.a.
	KS-166	1.76 \pm 0.64	3.98 \pm 0.84
	KS-172	23.3 \pm 8.9	n.a.
	KS-173	2.09 \pm 0.20	n.a.
	KS-174	1.63 \pm 0.33	5.18 \pm 0.28
	KS-184	0.56 \pm 0.24	6.07 \pm 1.48

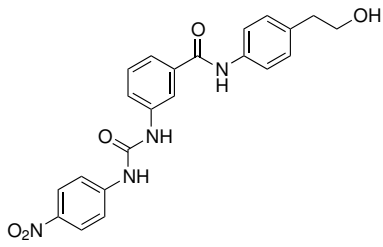
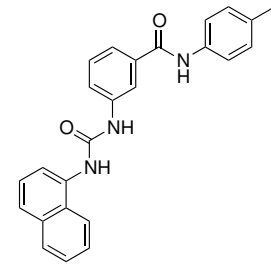
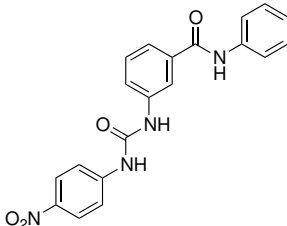
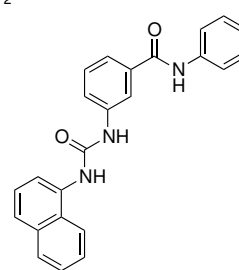
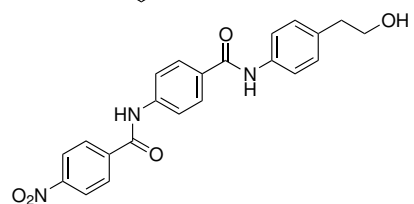
4.2.9 Variations of the linker position

None the compounds with *meta* or *para* substitutions on the second ring show any activity against BCRP in the Hoechst 33342 assay, as shown in table 4.7. Therefore, the relative position of the two amide groups seems to be crucial. Only the *ortho*-substituted compounds can establish a hydrogen bond between the NH and the C=O of the amide groups on the second ring. As already presumed analyzing the activity of compounds with modifications in the first linker, this interaction may be important for the active conformation of this class of compounds.

Table 4.7: Activity of compounds with variations of the linker position. Activities were measured using Hoechst 33342 assay and MCF-7 MX cells for BCRP (reference substance WK-X-24: $IC_{50} = 0.90 \pm 0.07 \mu\text{M}$) and calcein AM assay and A2780 Adr cells for P-gp (reference substance WK-X-24: $IC_{50} = 0.41 \pm 0.08 \mu\text{M}$). n.a. = no activity for concentration up to $10 \mu\text{M}$.

Formula	Compound	IC_{50} BCRP $\mu\text{M} \pm$ SD	IC_{50} P-gp $\mu\text{M} \pm$ SD
	KS-176	1.58 ± 0.36	n.a.
	KS-279	n.a.	n.a.
	KS-280	n.a.	n.a.

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

Formula	Compounds	IC ₅₀ BCRP $\mu\text{M} \pm$ SD	IC ₅₀ P-gp $\mu\text{M} \pm$ SD
	KS-090	n.a.	n.a.
	KS-129	n.a.	n.a.
	KS-108	n.a.	n.a.
	KS-109	n.a.	n.a.
	KS-091	n.a.	n.a.

4.2.10 Compounds containing a tetrahydroisoquinoline moiety

The compounds reported in table 4.8 contain a tetrahydroisoquinoline group in their structure. The first compound reported in this table is compound WK-X-24 (XR9577) that was used as reference compound in this work. As reported in the table below, the substitution of the amide in the second linker with an urea group leads to compounds with a high inhibitory potency against BCRP, as compound KS-105 ($\text{IC}_{50} = 0.35 \mu\text{M}$). The substitution of the nitro group with a methyl group leads to a decreasing

4.2 Determination of the activity of a new class of BCRP-inhibitors.

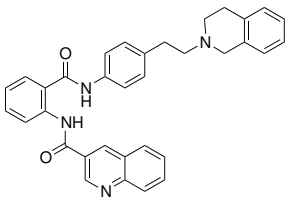
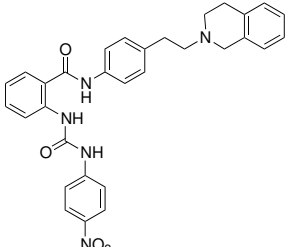
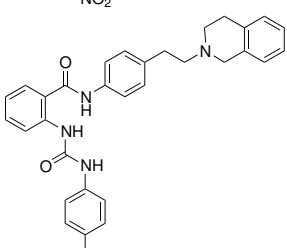
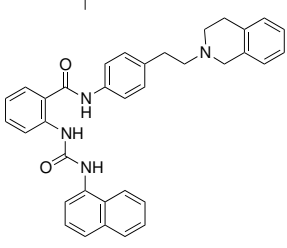
BCRP inhibitory activity by about 3-folds (compound KS-216). A further decreasing of the inhibitory activity is caused by substitution of the third aromatic ring with a naphthalene (compound KS-220, $IC_{50} = 5.17 \mu\text{M}$). This result suggests that a slightly hydrophobic and strong electron acceptor group as the nitro group increases the activity, as already observed for compounds without the tetrahydroisoquinoline group. The introduction of two methoxy groups at position 4 and 5 of the tetrahydroisoquinoline group leads to a decrease of the BCRP inhibitory activity by about 9-folds with respect to its homologue compound KS-216 (compound KS-226, $IC_{50} = 9.30 \mu\text{M}$). This result is in accordance with the Free-Wilson analysis proposed by Pick et. al. [148]. The substitution of the amide group in the first linker with an urea group (KS-136) increases the IC_{50} to a value of $10.35 \mu\text{M}$, that is about 30 folds higher. Also the inhibitory activity of six compounds with a tetrahydroisoquinoline group but without the third aromatic ring was tested. Between them, the only two compounds that resulted to be BCRP inhibitors are the compounds KS-221 and KS-122. As shown on table 4.8, the first one ($IC_{50} = 3.47 \mu\text{M}$) has as first linker an urea group, two methoxyl groups on the tetrahydroisoquinoline rest and a bromine at *para* position on the aromatic ring. Normally, the presence of methoxyl groups on the tetrahydroisoquinoline group decreases the inhibitory activity of compounds against BCRP, therefore the bromine group could be responsible of the inhibitory activity of this compound. Unfortunately, no compounds with similar structure but without the bromine group were synthesized to confirm this hypothesis. The other active BCRP inhibitor, KS-122, has a moderate inhibitory activity against BCRP, with an IC_{50} value of $6.70 \mu\text{M}$. Its structure has an amide group as first linker and the terminal phenyl group is substituted at *para* position with a cyano group. The analogue compound KS-210, with a methyl group in place of the cyano group has no effects against BCRP at concentrations up to $10 \mu\text{M}$. Four compounds with the tetrahydroisoquinoline group in their structure have an *ortho* substitution on the second aromatic ring. Only one of them (KS-075, $IC_{50} = 4.68 \mu\text{M}$) is able to inhibit BCRP. This compound differs from the other three by the presence of two methoxyl groups on the tetrahydroisoquinoline rest and by the presence of an urea as first linker.

Regarding the inhibitory activity of these compounds against P-gp, the comparison between the results of the calcein-AM assay of compounds KS-105 and KS-216 with the results of compounds WK-X-36 and WK-X-29 [154] showed that the introduction of an urea as second linker decreases the inhibitory activity against P-gp. A slight improvement of the inhibitory activity against P-gp is obtained by the introduction of two methoxy groups on the tetrahydroisoquinoline group. The introduction of a second urea group as first linker leads to loss of inhibitory activity against P-gp (KS-136). Between compounds without the third aromatic group, only the compounds KS-221 and KS-122 showed to be able to inhibit P-gp, with IC_{50} values of 3.47 and

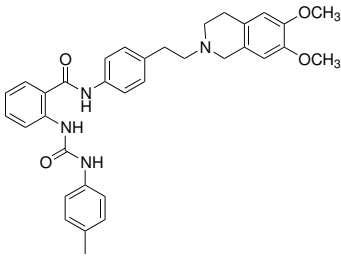
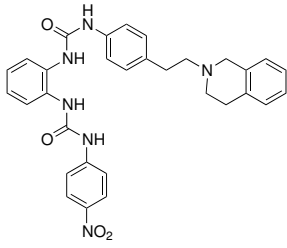
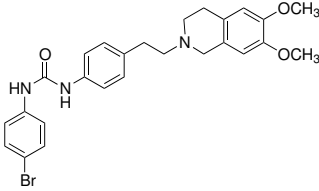
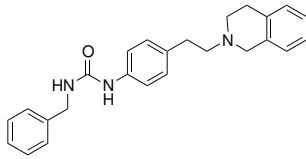
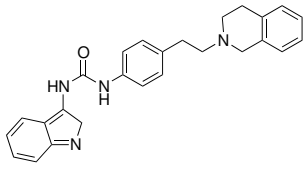
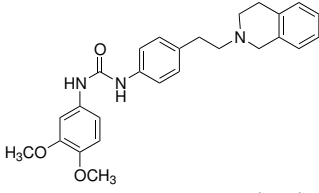
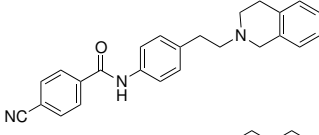
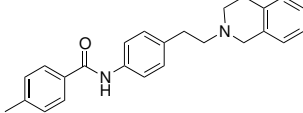
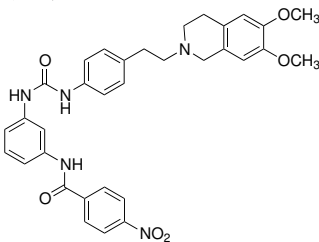
4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

6.70 μM , respectively. The compounds KS-075 and KS-098, with the two linkers on *meta* position, are weak inhibitors of P-gp, with IC_{50} values of 21.75 and 20.50 μM , respectively. The analogue compound KS-101 does not inhibit P-gp. It is interesting that the presence of the urea group on the first or on the second linker does not significantly change the inhibitory activity against P-gp. Finally, the compound KS-132, with two amides as linkers, has an IC_{50} value of 4.87 μM only about 3-folds higher than of its *ortho* analogue WK-X-36 ($\text{IC}_{50} = 1.45 \mu\text{M}$) [154].

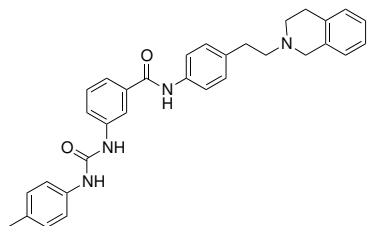
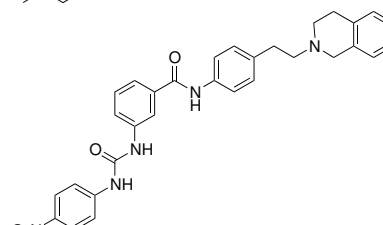
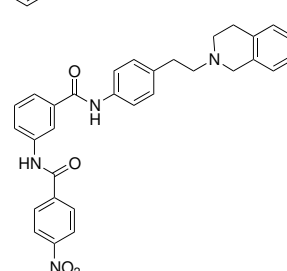
Table 4.8: Activity of compounds presenting a tetrahydroquinoline group. Activities were measured using Hoechst 33342 assay and MCF-7 MX cells for BCRP and calcein AM assay and A2780 Adr cells for P-gp. n.a. = no activity for concentration up to 10 μM .

Formula	Compound	IC_{50} BCRP $\mu\text{M} \pm$ SD	IC_{50} P-gp $\mu\text{M} \pm$ SD
	WK-X-24	0.901 ± 0.066	0.415 ± 0.080
	KS-105	0.35 ± 0.05	2.55 ± 0.71
	KS-216	1.10 ± 0.09	3.49 ± 0.36
	KS-220	5.17 ± 1.33	3.68 ± 0.36

4.2 Determination of the activity of a new class of BCRP-inhibitors.

Formula	Compounds	IC ₅₀ BCRP $\mu\text{M} \pm$ SD	IC ₅₀ P-gp $\mu\text{M} \pm$ SD
	KS-226	9.30 \pm 1.84	2.01 \pm 0.40
	KS-136	10.4 \pm 4.44	n.a.
	KS-221	3.47 \pm 0.76	6.17 \pm 0.84
	KS-224	n.a.	9.12 \pm 3.85
	KS-117	n.a.	n.a.
	KS-115	n.a.	n.a.
	KS-122	6.70 \pm 1.82	17.6 \pm 1.5
	KS-210	n.a.	10.6 \pm 5.7
	KS-075	4.68 \pm 1.55	21.8 \pm 2.7

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

Formula	Compounds	IC ₅₀ BCRP $\mu\text{M} \pm$ SD	IC ₅₀ P-gp $\mu\text{M} \pm$ SD
	KS-101	n.a.	n.a.
	KS-098	n.a.	20.5 \pm 5.4
	KS-132	n.a.	4.87 \pm 0.58

4.2.11 The effect of selected compounds on the cytotoxicity of mitoxantrone, SN-38 and Hoechst 33342

In order to investigate the influence of this new class of BCRP inhibitors on the anti-proliferative effect of cytotoxic drugs in presence or absence of BCRP, the cytotoxicity of SN-38 and Hoechst 33342 was tested in MDCK BCRP cells. The results of these tests further confirm the validity of the activity data collected with the Hoechst 33342 assay. In the experiments presented in this section, different BCRP inhibitors were added at concentrations of 5 and 10 μM to MDCK BCRP cells that were stuck on the bottom of a Greiner 96-well plate. As control, a PBS buffer solution administered to MDCK BCRP cells and to MDCK cells was used. The cytotoxic agent was then added at different concentrations.

The choice of using MDCK and the MDCK BCRP cells lines instead of the MCF-7 and MCF-7 MX cells lines was taken considering the low quality of the curves obtained using these second cells lines. In figure 4.20 is shown the effect gives by increasing concentrations of the inhibitor KS-407 on the dose-response curves of the cytotoxic compound SN-38. The curves on the left side of the figure (4.20a) are obtained using MCF-7 and MCF-7 MX cell lines, while the curves on the right side (4.20b) are obtained using MDCK and MDCK BCRP cell lines. The MTT assay in which MCF-7 and MCF-7 MX cell lines were used, showed a lower span between the curve corresponding to the resistance cells and that corresponding to the parental cell lines (pEC_{50} of 7.19 and 6.55, respectively) in respect with the MTT assay using MDCK and MDCK BCRP cell lines (pEC_{50} of 6.87 and 5.93, respectively).

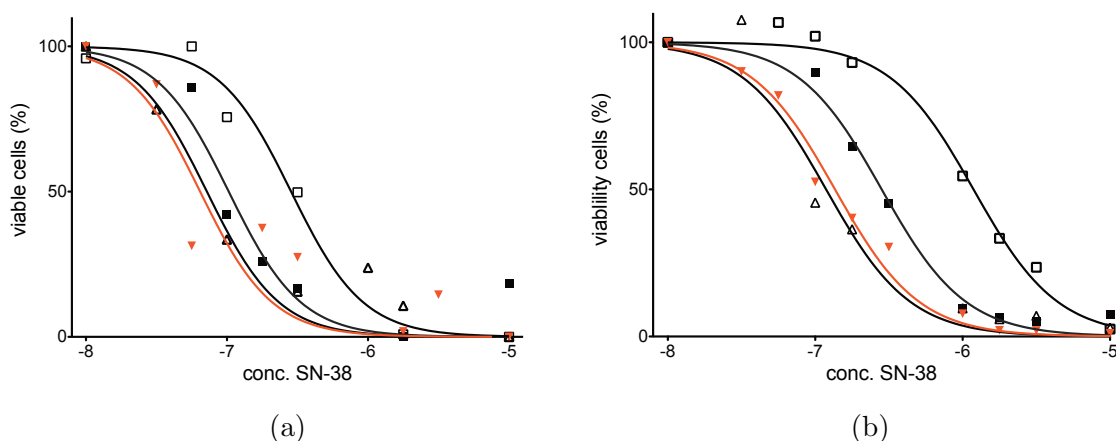


Figure 4.20: The shifts of the dose-response curves caused by increasing inhibitor concentrations using the compound KS-407 as inhibitor in (a) MCF-7 and MCF-7 MX cells or (b) MDCK and MDCK BCRP cells. Resistant cells (empty square), resistant cells + 5 μM inhibitor (black square), resistant cells + 10 μM inhibitor (empty triangle), parental cells (red down triangle).

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

Furthermore, the goodness of fit is higher for the curves obtained using MDCK and MDCK BCRP cell lines in respect with the curves obtained using MCF-7 and MCF7-MX cell lines.

In order to exclude possible artefacts, the cytotoxic effects of selected compounds at concentrations of 5 and 10 μM on the MDCK BCRP cells also were analyzed. The results reported in figure 4.21 showed that at the studied concentrations, no cytotoxic effects in MDCK BCRP cells were observed.

In the light of the above observations and considerations, it was decided to use the MDCK and MDCK BCRP cell lines for the next assays.

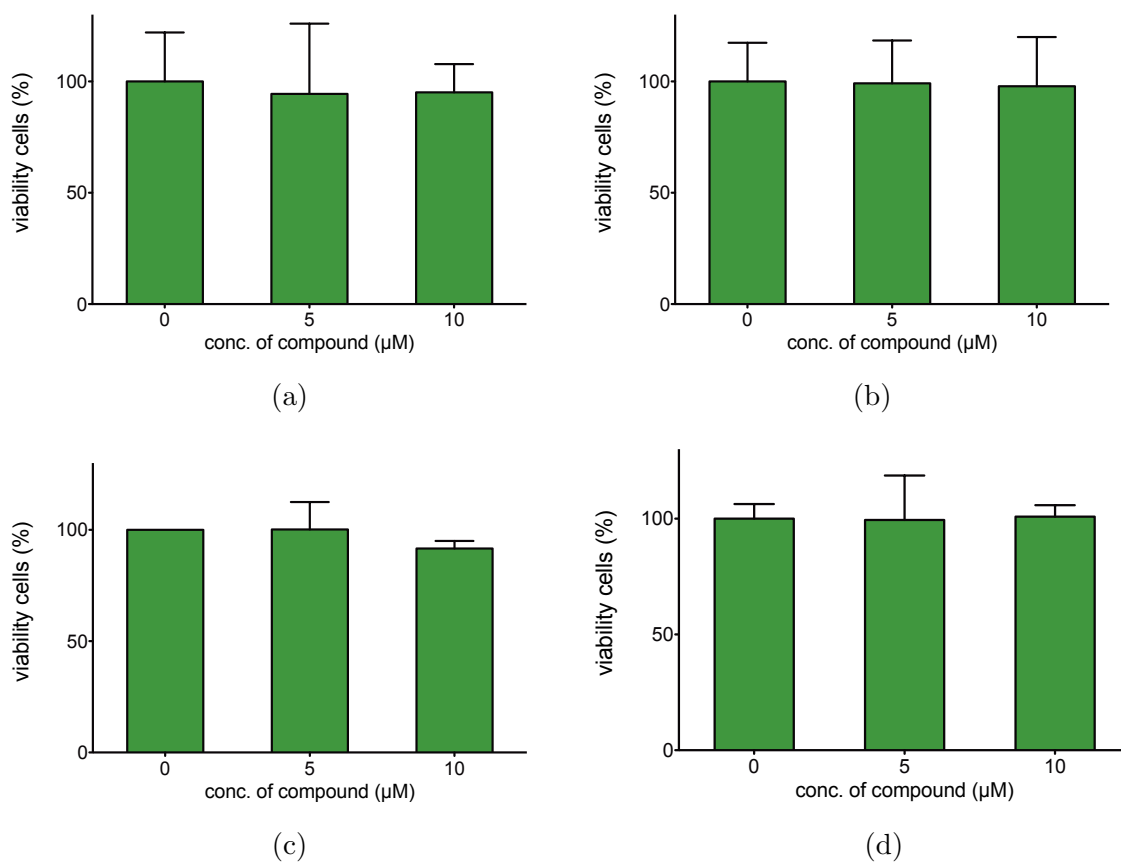


Figure 4.21: Cytotoxic effects of compounds (a) KS-407, (b) KS-251, (c) KS-166, and (d) KS-174 in MDCK BCRP cells at concentrations of 5 and 10 μM .

The importance of the ortho connection on the second aromatic ring

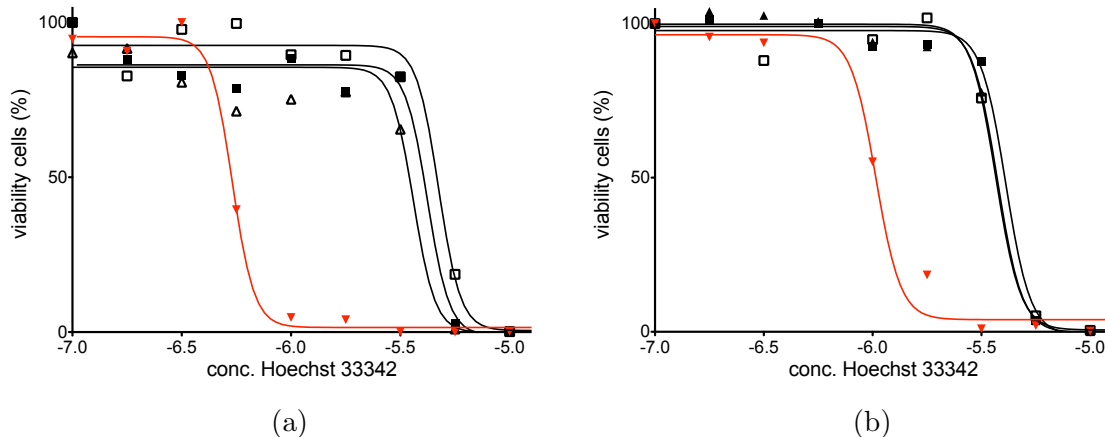


Figure 4.22: The shift of the dose-response curve of Hoechst 33342 caused by increasing inhibitor concentrations in MDCK cells. The shifts of the dose-response curves due (a) to compound KS-279 and (b) to compound KS-091. MDCK BCRP cells (empty square), MDCK BCRP cells + 5 μM inhibitor (black square), MDCK BCRP cells + 10 μM inhibitor (empty triangle), parental MDCK cells (red down triangle).

The results of the Hoechst 33342 assay showed that only compounds with the two linkers in *ortho* position on the second aromatic ring are able to effectively inhibit BCRP. This observation was confirmed by the MTT assay. As seen in figure 4.22, compound KS-279, that contains the two linkers in *meta* position and compound KS-091, that has the two linkers in *para* position are both not able to resistance towards Hoechst 33342 of MDCK BCRP cells.

Substitutions on the first aromatic ring

Figure 4.23 shows the effect on the dose-response curves of Hoechst 33342 caused by selected BCRP inhibitors with modifications on the first aromatic ring. The results of the MTT assays confirm that the hydroxyethyl group on the first aromatic ring is not essential for the activity and can be replaced by other groups. In particular, the substitutions of the hydroxyethyl group by benzoyl (compound KS-174), propyl (compound KS-385), hydroxyl (compound KS-251) and methoxy ethyl group (compound KS-292) lead to compounds that are able to completely reverse the resistance of MDCK BCRP cells against Hoechst 33342.

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

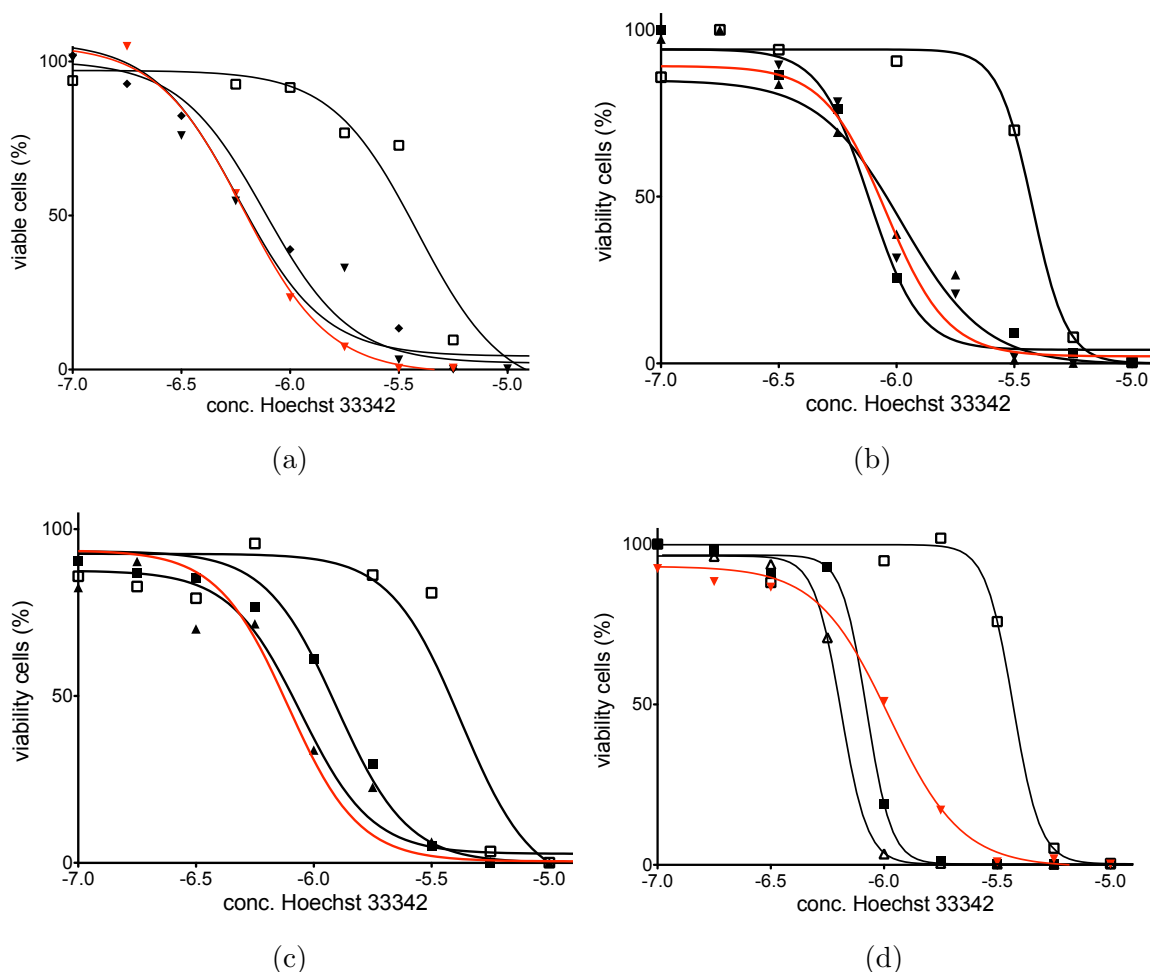


Figure 4.23: The shifts of the dose-response curves of Hoechst 33342 caused by increasing inhibitor concentrations in MCDK cells. The used inhibitors are (a) KS-174, (b) KS-385, (c) KS-251 and (d) KS-292. MDCK BCRP cells (empty square), MDCK BCRP cells + 5 μM inhibitor (black square), MDCK BCRP cells + 10 μM inhibitor (empty triangle), parental MDCK cells (red down triangle)

The influence on the activity of methoxy groups on the third aromatic ring on the activity of compounds without substituents on the first aromatic ring

The results of the Hoechst 33342 assay showed important differences of activity between compounds without any substituent on the first aromatic ring and with different substituents on the third aromatic ring. Mostly compounds do not showed any inhibitory activity or only low inhibitory activity against BCRP, with the exceptions of compounds KS-166 ($\text{IC}_{50} = 1.16 \mu\text{M}$) and KS-173 ($\text{IC}_{50} = 2.09 \mu\text{M}$). In order to confirm the results obtained with Hoechst 33342 assay, the MTT assay was also performed using low active compound KS-168 ($\text{IC}_{50} = 13.4 \mu\text{M}$) and the most active compound without substituents on the first aromatic ring, KS-166. The aim of this test was to determine if the large difference in activity between these two compounds, differing

4.2 Determination of the activity of a new class of BCRP-inhibitors.

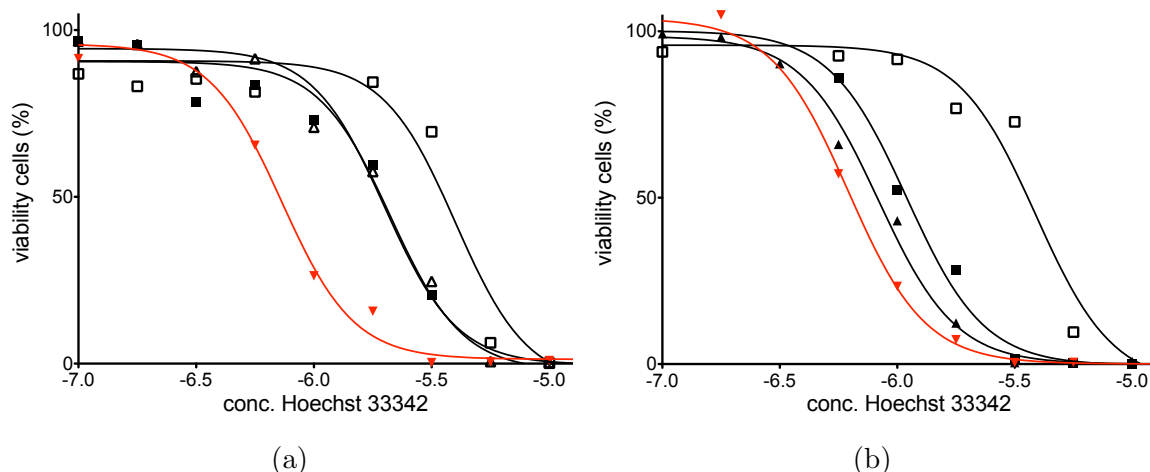


Figure 4.24: The shift of the dose-response curve of Hoechst 33342 caused by increasing inhibitor concentrations in MDCK cells. (a) The shift of the dose-response curve caused by compound KS-168. (b) The shift of the dose-response curve caused by compound KS-166. MDCK BCRP cells (empty square), MDCK BCRP cells + 5 μM inhibitor (black square), MDCK BCRP cells + 10 μM inhibitor (empty triangle), parental MDCK cells (red down triangle)

by only one chemical group (3,4-dimethoxy on the third aromatic ring in compound KS-166 instead of 4-nitro in compound KS-168), is also seen with this assay. In figure 4.24, the shift of the dose-response curve of Hoechst 33342 caused by compound KS-168 is reported on the left, and the shift of the dose-response curve of Hoechst 33342 caused by compound KS-166 is reported on the right. It is possible to see that the shift of the dose-response curve caused by compound KS-168 is only partial and also for an inhibitor concentration equal to 10 μM the resistance of the MDCK BCRP cells against Hoechst 33342 is not completely reversed. In contrast, the presence of compound KS-166 at concentration equal to 10 μM is able to completely reverse the BCRP mediated resistance against Hoechst 33342, as shown in figure 4.24b. In conclusion, the results of the MTT assays for compounds without substituents on the first aromatic ring confirmed the data obtained from the Hoechst 33342 assay.

The influence of the substituents on the third aromatic ring

The results of the Hoechst 33342 assay for compounds with modifications on the third aromatic ring have shown that the introduction of an isopropoxy group at position *para* of the ring (compound KS-407) increases the inhibitory activity of the molecule with respect to the reference compound KS-176. In order to confirm the result of the Hoechst 33342 assay, the MTT assay was performed for two different cytotoxic compounds, namely Hoechst 33342 and SN-38. The results, illustrated in figure 4.25, show that compound KS-407 is able to reverse the resistance of MDCK

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

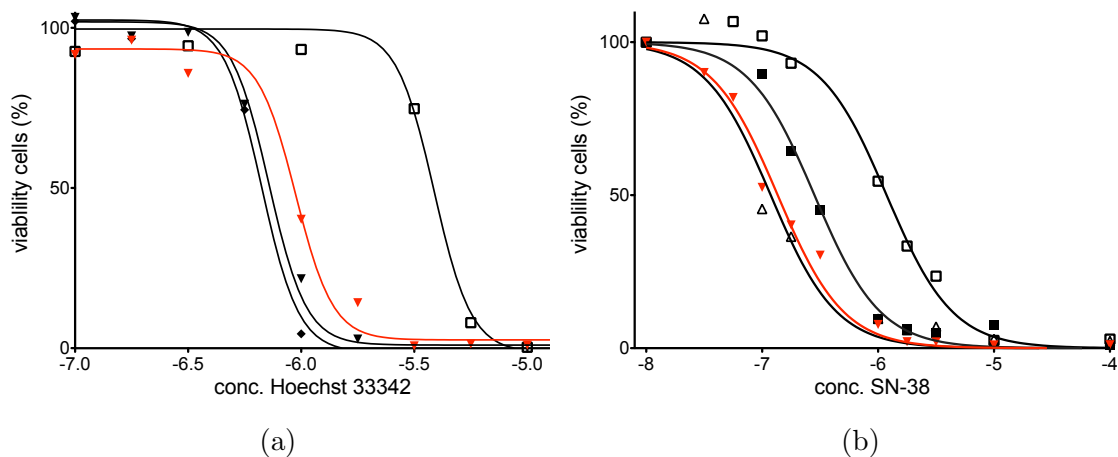


Figure 4.25: The shift of the dose-response curve of cytotoxic agents caused by increasing inhibitor concentration in MDCK cells. (a) The shift of the dose-response curve of Hoechst 33342 caused by compound KS-407. (b) The shift of the dose-response curve of SN-38 caused by compound KS-407. MDCK BCRP cells (empty square), MDCK BCRP cells + 5 μM inhibitor (black square), MDCK BCRP cells + 10 μM inhibitor (empty triangle), parental MDCK cells (red down triangle)

BCRP cells against both cytotoxic compounds. This result confirms the data obtained from the Hoechst 33342 assay and allows to consider compound KS-407 as a potent BCRP inhibitor.

4.2.12 Determination of the cytotoxicity of selected BCRP inhibitors

In order to measure the toxicity of the studied anthranile amide derivatives, MTT viability assays with selected compounds were performed. In order to determine the toxicity in the same organisms that were also used for the determination of the BCRP inhibitory activity, the MCF-7 and the resistant MCF-7 MX cell lines were chosen for the assays.

Cells were seeded in sterile 96-well plates with density of 10,000 cells per well. After 6 h incubation, the studied compounds were added at 4 different concentrations ranging between 1 and 31,6 μM . After 72 h incubation, the MTT solution, prepared following the standard procedure, was added to each vial. Finally, the suspension was removed from each vial and the cells were lysed using DMSO. The absorbance of each vial was measured using a BMG POLARstar microplate reader. Two vials were added with DMSO and other two vials were added only with PBS. The means of the absorbance values of those samples were used as positive and negative references and results were expressed as mean \pm SD of 3 independent experiments.

Toxicity of selected compounds with modifications on the third aromatic ring

The cytotoxicity of seven selected compounds with modifications on the third aromatic ring was measured. The results are illustrated in figure 4.26. The results of the cytotoxicity assays show that generally the parental cell line MCF-7 is more resistant against the cytotoxic effect of the studied compounds in respect to the MCF-7 MX cell line.

The results of these experiments show that compound KS-176 does not show cytotoxicity in parental MCF-7 cells at concentration lower than 10 μM , while its cytotoxicity in MCF-7 MX cells at concentration up to 10 μM . At a compound concentration equal to 31.6 μM , the decrease of the viability in both cell lines was significant. The decrease of the viability for compound KS-176 was also observed in other compounds. A possible explanation can be offered by the low solubility of these compounds in PBS: They could give precipitates, at concentrations up to 10 μM , with consequent reduction of the compound concentration in the solution. Similar results were observed also for compound KS-150, that differs from compound KS-176 only in the position of the nitro group on the third aromatic ring (at position *meta* instead of *para* in respect to the second linker). The same trend was also observed for compound KS-144.

Compounds containing halogens on the aromatic ring, like KS-169 (4-chlorine) and KS-242 (4-trifluoromethyl) are not cytotoxic for the parental cell line MCF-7 but show already at low concentrations high cytotoxicity in the MCF-7 MX cell line. On the other hand, compounds KS-407 and KS-157 do not show relevant cytotoxicity in

4 Investigation of the inhibitory activity of a new class of tariquidar analogues.

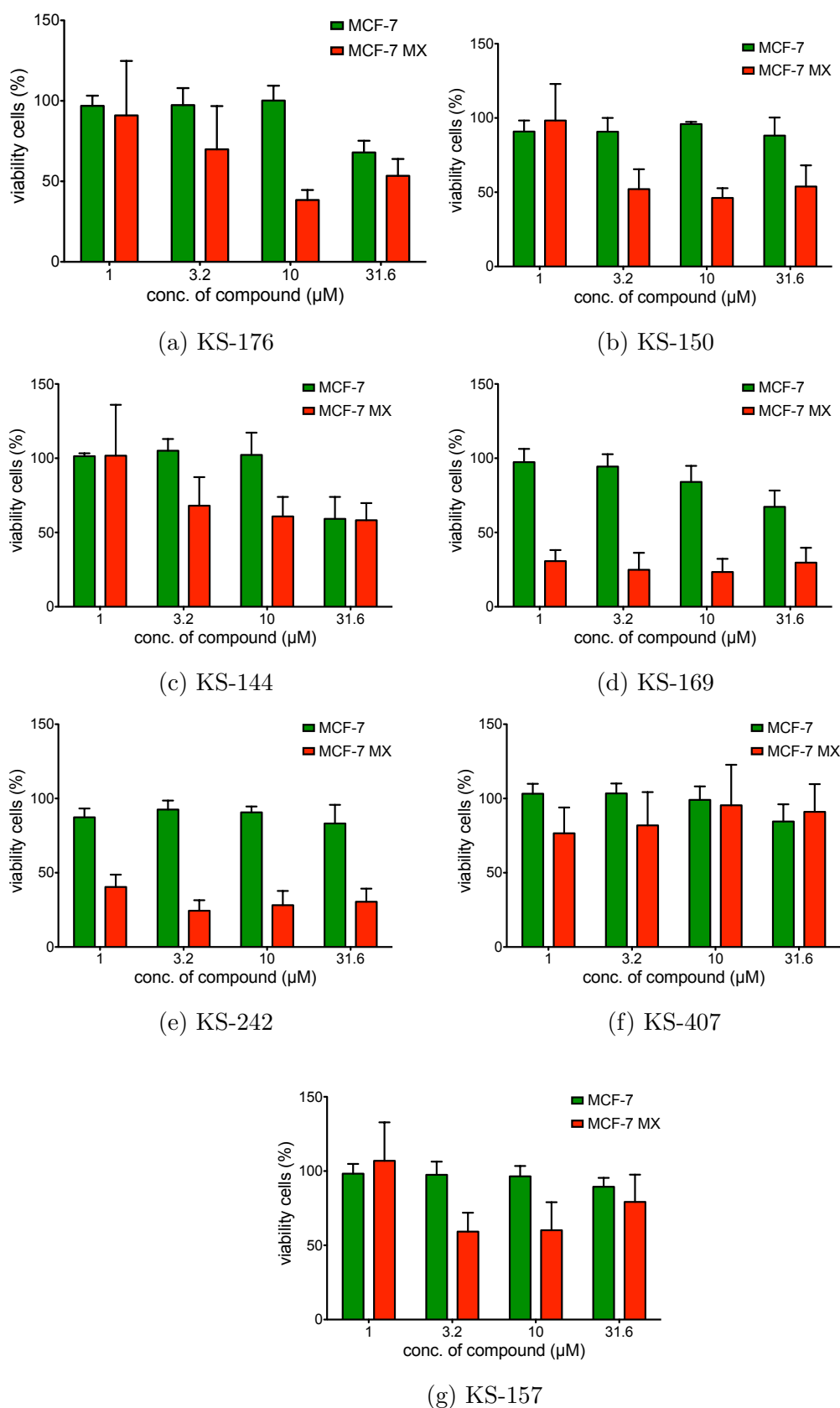


Figure 4.26: Cytotoxicity of compounds with substitutions on the third aromatic ring.

4.2 Determination of the activity of a new class of BCRP-inhibitors.

both cells lines also at concentrations up to 10 μM . The low value of IC_{50} of compound KS-407 and its low cytotoxicity make this compound a good candidate for further pharmacological studies.

Toxicity of selected compounds with modifications on the second aromatic ring

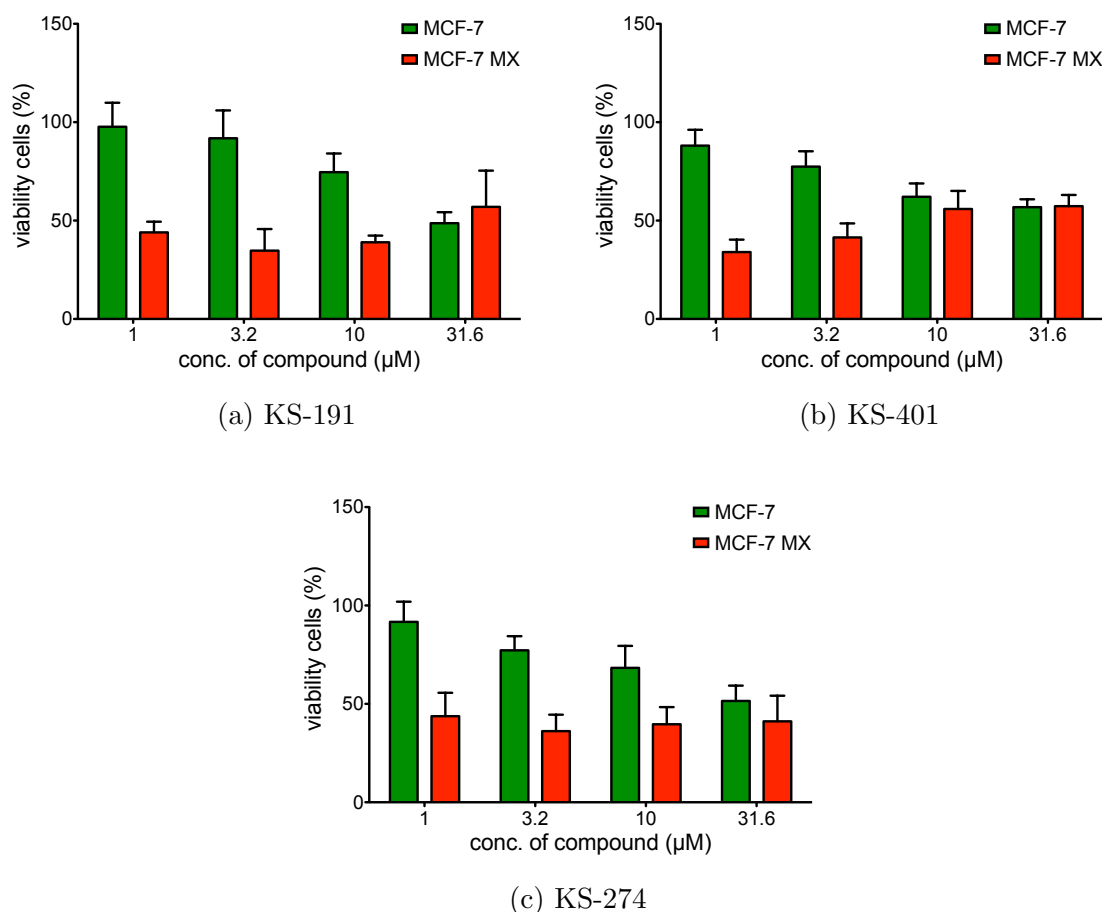


Figure 4.27: Cytotoxicity of compounds with substitutions on the second aromatic ring.

The data for the cytotoxicity of compounds with modifications on the second aromatic ring reported in figure 4.27 show that these compounds are cytotoxic for both MCF-7 and MCF-7 MX cell lines. In particular, also the viability of the MCF-7 cell line is reduced by increasing concentrations of the three investigated compounds. Regarding the resistant cell line MCF-7 MX, at all used compound concentrations, the cell viability was compromised, remaining at values of about 50 % with respect to the control. Considering the low inhibitory activity of these compounds combined with their high cytotoxicity, it can be concluded that these compounds are not good candidates for further investigations.

Toxicity of selected compounds with modifications on the first aromatic ring

Six compounds with modifications of the substituent group at *para* position on the first aromatic ring were selected to investigate their cytotoxicity. The result of the cytotoxicity assay of compound KS-176 is also reported in the figure 4.28 for comparison. It can be seen that the methylation of the hydroxyethyl group leads to an increase of the cytotoxicity of compound KS-292 for the MCF-7 MX cell line, while this compound is not cytotoxic for the MCF-7 cell line, also at high concentrations. The introduction of a methoxy group directly connected with the aromatic ring (compound KS-267) leads to high cytotoxicity for the MCF-7 MX cell line and a moderate cytotoxicity for the parental MCF-7 cell line. The presence of a hydroxy group directly connected with the ring, as in compound KS-251, leads to complete absence of cytotoxicity for the MCF-7 cell line and to reduction of the cell viability in MCF-7 MX cells only at high concentrations. The absence of substituents on the first aromatic ring (compound KS-166) leads to appearance of cytotoxic effects in MCF-7 cells only at concentrations higher than 10 μM , while the cell viability of the MCF-7 MX cell line is compromised already at low compound concentrations. An analogues effect was observed for compound KS-385, possessing a propyl group on the ring at position 4. Finally, compound KS-174, with a benzoyl group at *para* position on the ring, is not cytotoxic for the MCF-7 cell line at all the studied concentrations and shows only low cytotoxicity for the MCF-7 MX cell line at a concentration of 31.6 μM .

4.2 Determination of the activity of a new class of BCRP-inhibitors.

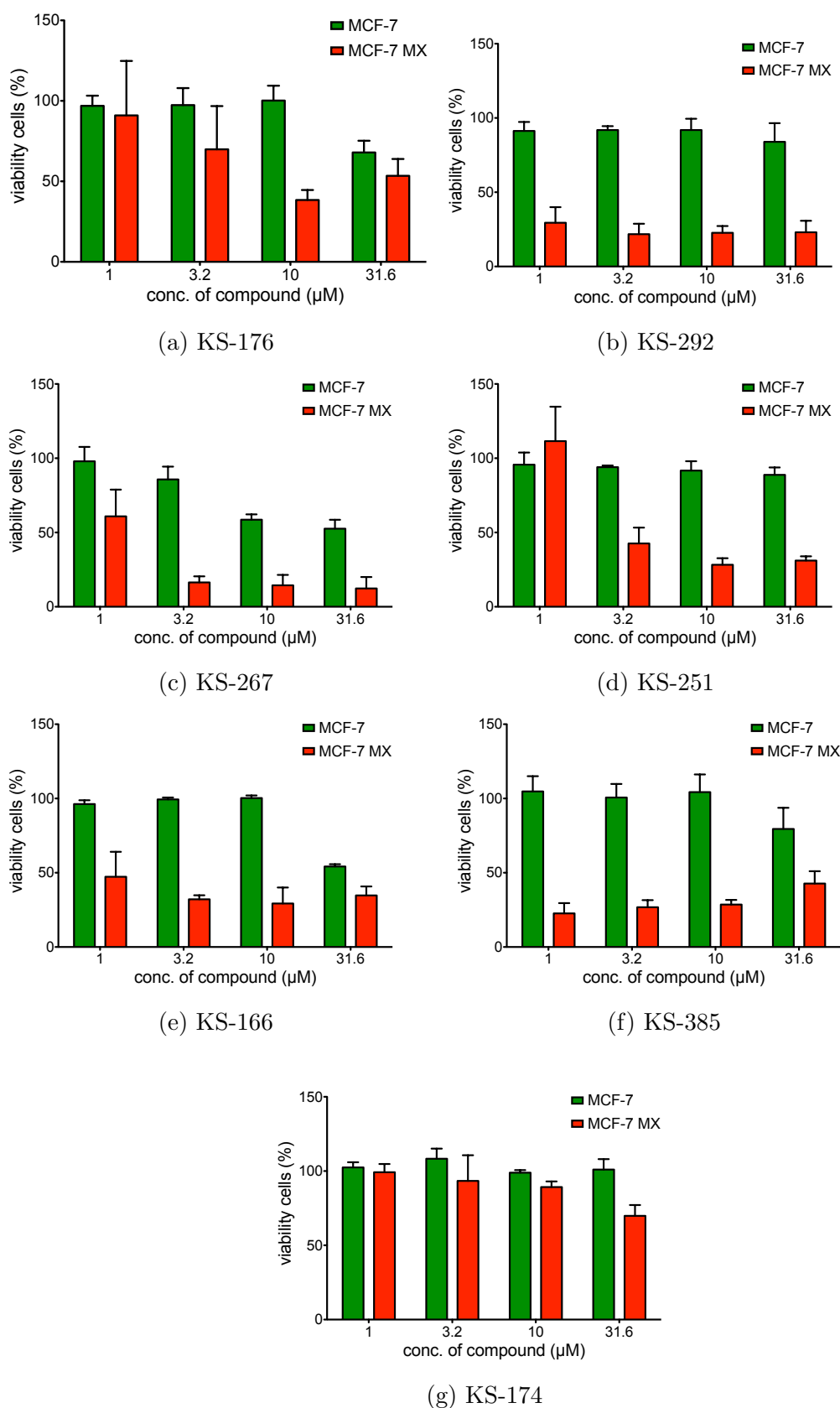


Figure 4.28: Cytotoxicity of compounds with substitutions on the first aromatic ring.

Toxicity of compounds presenting an hydroxyl group at position 4 on the first aromatic ring

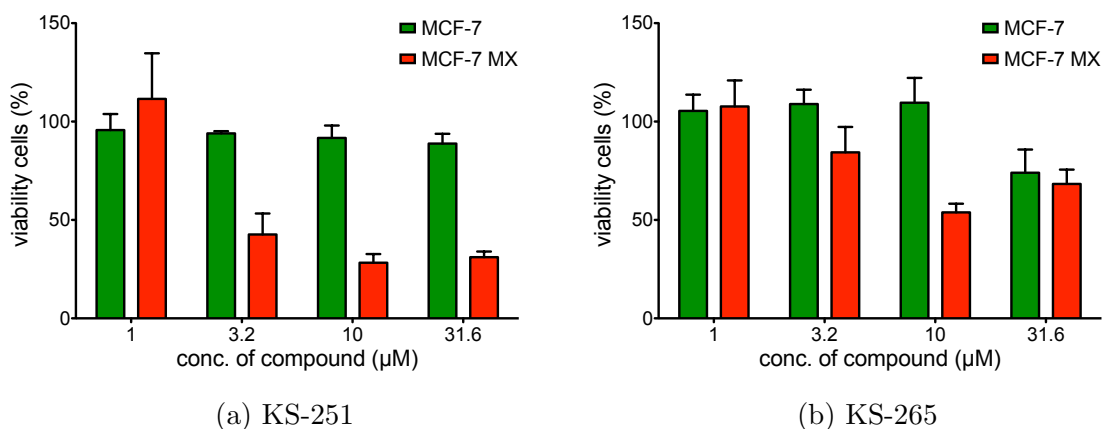


Figure 4.29: Cytotoxicity of compounds with 4-hydroxy group on the first aromatic ring and 4-nitro (KS-251) or 3,4-dimethoxy (KS-265) on the third aromatic ring.

The results from the Hoechst 33342 assay showed that compounds with an hydroxy group at *para* position on the first aromatic ring are potent BCRP inhibitors with similar IC_{50} values. Regarding the cytotoxicity of these compounds, the results of the MTT assays showed that they have only low cytotoxicity in the MCF-7 cell line at high concentrations (higher than 10 μM). In MCF-7 MX cells, the cytotoxicity is relevant, for compound KS-251 at concentrations higher than 3 μM . Compound KS-256, with two methoxy groups at position 3 and 4 of the third aromatic ring, reduces considerably the viability of the MCF-7 MX cells considerably only at concentrations of 10 μM and higher. This last observation confirms that the presence of two methoxy groups on the third aromatic ring generally reduces the cytotoxicity of the compounds, as already observed for compound KS-157.

5 Development of binary QSAR models for classification of BCRP inhibitors.

5.1 Aim of this work

The results of the Hoechst 33342 assay showed that a wide number of the tested compounds are weak inhibitors of BCRP or do not have any inhibitory activity at this transporter.

The presence of numerous inactive compounds in a dataset is a problem for two principal reasons: the first reason is that we are interested to obtain active inhibitors (if possible, highly active) for the studied protein target and the presence of many inactive compounds is an indication that drastic modification of the molecular scaffold is required or that some basic feature of the molecular structure must be modified. The second reason is that for these compounds is not possible to determine IC_{50} values, with the result that we lose information for the building of standard linear QSAR models.

In order to understand the reasons of such consistent presence of inactive compounds and to orient the synthesis of new BCRP inhibitors to active compounds, three machine learning algorithms were used to classify the training dataset constituted by the newly synthesized tariquidar analogues.

The three chosen methods are the Self-Organizing Maps (SOM), an Artificial Neural Network that uses unsupervised learning, the Support Vector Machine (SVM), a supervised learning algorithm, and the k-Nearest Neighbor (k-NN) algorithm, a lazy-learning algorithm.

The parameters of each used machine learning algorithm were optimized to increase its predictivity. Finally, an external dataset of compounds with a different scaffold was used as test set to estimate the predictivity of the best models.

The results show that the generated models could be used as filter for designing new compounds with increased chances to be active BCRP inhibitors.

5.2 Preparation of the molecular structures and descriptor calculation

5.2.1 Description of the used training dataset

The compounds used for the training dataset were synthesized by Dr. Kerstin Stegeman, a former PhD student of our research group in order to obtain a new class of selective BCRP-inhibitors [155].

The compounds included in the dataset are chemically derived from tariquidar and are structurally highly similar. In the dataset are included also structures containing the tetrahydroisoquinoline rest as present structure of tariquidar as well as compounds without this chemical group. All compounds of this dataset have been tested in the Hoechst 33342 assay, to evaluate their inhibitory activity against BCRP, and in the calcein AM assay to evaluate the P-gp inhibitory activity.

In conclusion, 99 compounds were tested and classified. The dataset was divided into two subgroups: 59 compounds with IC_{50} values for BCRP equal or lower than $10 \mu\text{M}$ were classified as BCRP inhibitors (class-code equal to 1). The remaining 40 compounds with an IC_{50} value higher than $10 \mu\text{M}$ were classified as non-inhibitors (class-code equal to 0). The compounds used for classification and their class-codes are reported in table 5.1.

5.2.2 Structure optimization

All the compounds mentioned in this work were drawn using the software MOE [97]. The molecular structures were optimized before calculating molecular descriptors, in order to find the optimal conformation and to calculate the partial charges of the atoms. The optimization consists of three different steps: the calculation of force field partial charges, the conformational search and the recalculation of partial charges using a quantum mechanic method.

In the first step the partial charges were calculated using the force field MMFF94x [156]. Afterwards, a stochastic conformational search was performed using a root mean square (RMS) gradient equal to 0.005 \AA and an iteration limit of 10000.

Due to high flexibility of the molecular structures and the calculations performed in vacuo, the conformation associated to the lowest potential energy was frequently an unnaturally bent structure. The problem was solved taking the conformation with lowest globularity instead of the conformation with lowest potential energy.

Finally, the charges of the chosen conformation were then recalculated using the PM3 quantum mechanic method included in MOPAC and the structure minimized to the local minimum energy.

Table 5.1: The compounds used in the Dataset and their activity class.

Compound	Class	Compound	Class	Compound	Class
WK-X-24	1	KS-255	1	KS-129	0
KS-072	1	KS-257	1	KS-132	0
KS-075	1	KS-265	1	KS-168	0
KS-094	1	KS-266	1	KS-170	0
KS-105	1	KS-267	1	KS-172	0
KS-122	1	KS-272	1	KS-181	0
KS-226	1	KS-274	1	KS-199	0
KS-136	1	KS-282	1	KS-210	0
KS-144	1	KS-292	1	KS-224	0
KS-150	1	KS-294	1	KS-228	0
KS-157	1	KS-304	1	KS-231	0
KS-159	1	KS-305	1	KS-237	0
KS-162	1	KS-308	1	KS-268	0
KS-166	1	KS-311	1	KS-277	0
KS-169	1	KS-322	1	KS-279	0
KS-174	1	KS-342	1	KS-280	0
KS-176	1	KS-364	1	KS-290	0
KS-184	1	KS-366	1	KS-293	0
KS-186	1	KS-367	1	KS-328	0
KS-187	1	KS-371	1	KS-348	0
KS-190	1	KS-385	1	KS-351	0
KS-191	1	KS-399	1	KS-357	0
KS-194	1	KS-401	1	KS-358	0
KS-196	1	KS-173	1	KS-360	0
KS-206	1	KS-407	1	KS-374	0
KS-216	1	KS-090	0	KS-375	0
KS-220	1	KS-091	0	KS-381	0
KS-221	1	KS-092	0	KS-382	0
KS-238	1	KS-098	0	KS-383	0
KS-242	1	KS-101	0	KS-405	0
KS-246	1	KS-108	0	KS-406	0
KS-249	1	KS-109	0		
KS-251	1	KS-115	0		
KS-252	1	KS-117	0		

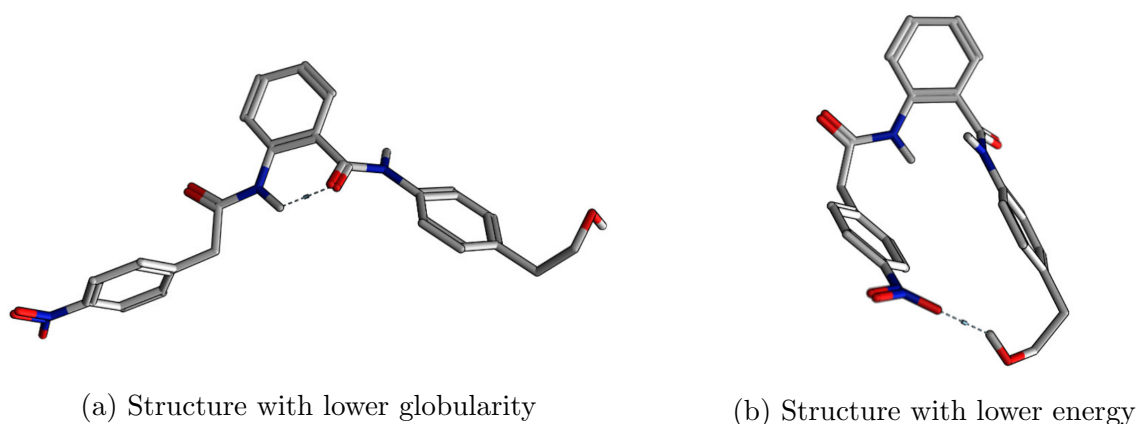


Figure 5.1: Comparison of two different conformation of compound KS-181. The first conformation shows the lowest globularity ($\text{glob} = 0.16$) and a potential energy value equal to 92.68 kcal/mol. The second structure has the lowest potential energy (87.48 kcal/mol) but an high globularity value (0.36).

5.2.3 Calculation of molecular descriptors

The molecular descriptors used in this work can be divided into zero-dimensional descriptors (0D), one-dimensional descriptors (1D), two-dimensional descriptors (2D), and three-dimensional descriptors.

0D descriptors are simply derived from the chemical formulas of compounds. 1D descriptors report the count of functional groups and substructures in the molecule. 2D descriptors are calculated using information deriving from the connectivity and distance matrix of structures. For calculation of these last descriptors no information about the molecular conformation are needed. Finally, 3D descriptors use atoms coordinates to be calculated.

The same class of descriptors, e.g. ATS descriptors, can be 2D or 3D descriptors according to the used kind of distance (topological or geometrical respectively). The calculation of molecular descriptors was performed using the software MOE [97] and the on-line version of the software DRAGON. For each molecule, 2164 descriptors were calculated and divided into 13 classes. Those 13 classes were used separately or in different combinations e.g. 2D and 3D RDF descriptors together or 0D, 1D, 2D, and 3D descriptors. Descriptors with only zero values were removed. The kind of descriptors used, the dimensionality and the number of descriptors calculated for each class are summarized in table 5.2.

Table 5.2: Classes and number of molecular descriptors calculated

Classes	Dimensionality	Number of descriptors calculated
0D, 1D and 2D MOE descriptors	0D, 1D, 2D	174
3D MOE descriptors	3D	124
Inductive descriptors MOE	3D	50
2D RDF descriptors	2D	75
3D RDF descriptors	3D	150
2D RECON descriptors	2D	143
3D RECON descriptors	3D	143
TAE RECON descriptors	3D	257
3dMORSE descriptors	3D	160
gATS descriptors	3D	260
tATS descriptors	2D	135
WHIM descriptors	3D	99
GETAWAY descriptors	3D	197
Total number		2164

5.3 Use of SOMs to discriminate between BCRP-inhibitors and non-inhibitors

5.3.1 Identification of the optimal parameters of the SOM

In order to increase the predictivity of the model based on SOMs, the parameters of the map were changed to find the best setting. The optimized parameters are the number of iterations, the topology of the SOM and the dimension of the SOM. The number of iterations refers to the number of times that the complete dataset is presented to the SOM. The number of iterations must be enough to guarantee reaching of convergence.

The optimal number of iterations was found using the largest descriptor set of this study (the combination of 2D, 3D and TAE RECON descriptors, in total 543 descriptors) in a 9 x 9 SOM. The results show that after 4000 iterations the mean distance between input vectors and the corresponding winning neuron appears to be constant (fig. 5.2).

Regarding the topology of SOMs, the major problem is to ensure that all neurons of the SOM have the same number of first-, second- and third- neighbors. This is not possible using a classical topology, as the rectangular and hexagonal topologies. Neurons at the edges of a rectangular or hexagonal SOM have incomplete neighborhood. This problem was solved using toroidal topologies.

The optimal dimension of the SOM was determined using a grid approach, for each used set of descriptors. Several SOMs were generated with increasing dimensions from 5 x 5 to 9 x 9 neurons. The dimension limit of 9 x 9 neurons is due to the number of

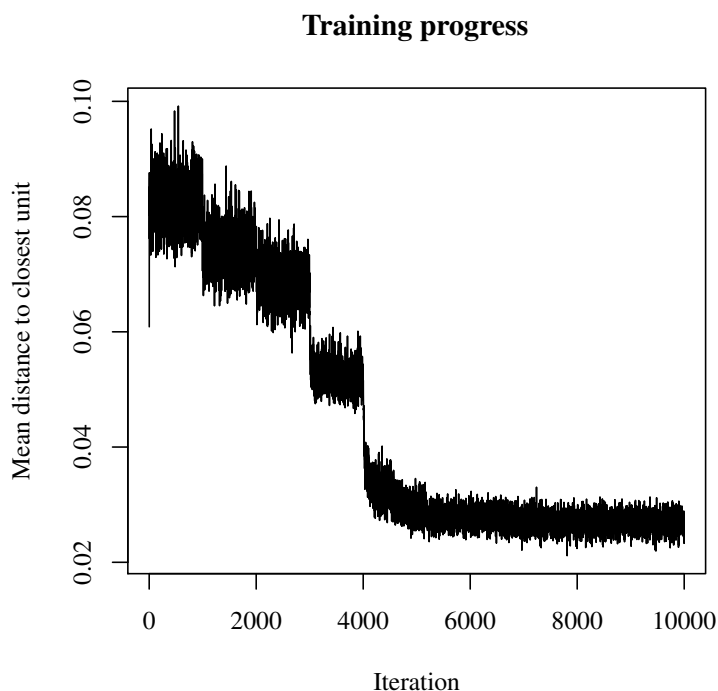


Figure 5.2: Example of training progress for a 9 x 9 SOM using 2D, 3D and TAE RECON descriptors.

molecules in the dataset that must be higher than the number of neurons of the SOM. The predictivity was evaluated for each generated SOM and the results are reported successively in this thesis.

5.3.2 Analysis of results

The package 'kohonen' implemented in R (version 2.15.0) [157] is used in this work to calculate self-organizing maps. Several models were generated using different sets of descriptors. For each set, molecular descriptors with variance equal to 0 were removed.

Before training of the SOM, each column of the input matrix was normalized, subtracting the mean of all values in the column and dividing by the standard deviation. Normalization of features is necessary in this work because the ranges of the used descriptors are widely different.

In order to obtain better prediction of the less represented class (weak active inhibitors), a down-sampling of the dataset was performed. In machine learning algorithms unbalanced datasets can cause problems when predicting the smaller class. An approach to solve this problem consists in taking, for training of the SOM, only as many elements from the larger class as elements in the smaller class, obtaining in this way a down-sampled dataset where the amount of elements of the two classes is 1:1. The elements taken from the larger class to obtain the down-sampled database were

5.3 Use of SOMs to discriminate between BCRP-inhibitors and non-inhibitors

randomly selected. For each training cycle, different elements were randomly selected for the down-sampled database. This down-sampled training set was used to build the SOM model. The SOM was trained using a learning rate decreasing linearly from 0.05 to 0.01. The initial weights were randomly generated.

The predictivity of the trained SOM was finally evaluated using the leave-one-out (LOO) cross validation (CV). The down-sampling, and the calculation of the SOM model were performed 100 times giving each time small difference in the accuracy values, due to different elements in the down-sampled dataset and to different initial weights of SOMs neurons. The predictivity of the best models was finally tested using an external dataset (test set). The overall process is shown in figure 5.3.

The accuracy of the prediction is expressed using the parameter "accuracy value", "true positive rate" (TP rate) and "true negative rate" (TN rate). These parameters were already explained in section 2.3.

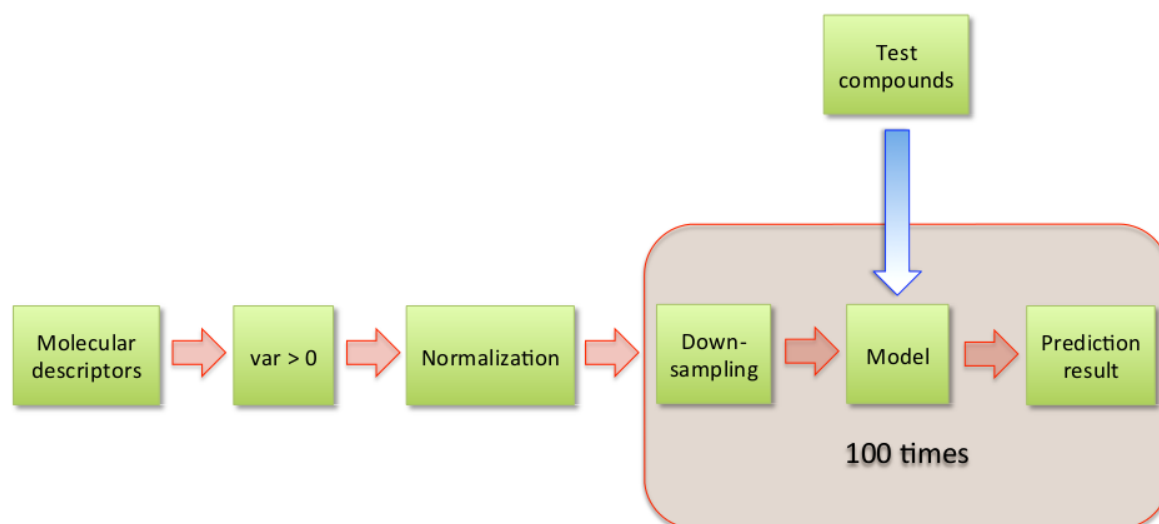
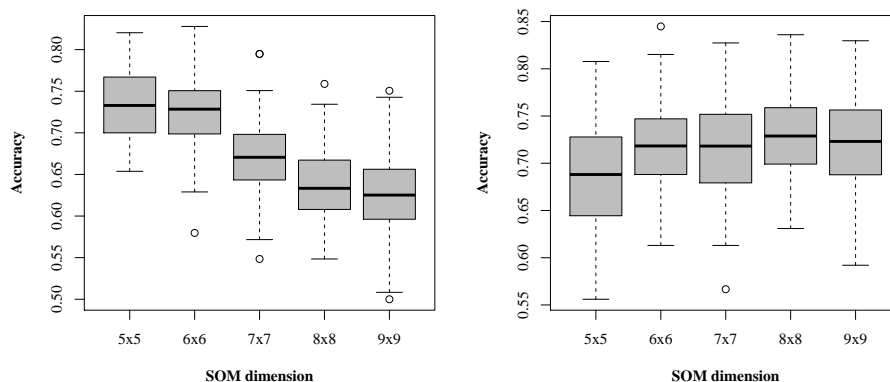


Figure 5.3: Framework for the generation of SOM models.

0D, 1D, 2D and 3D MOE descriptors

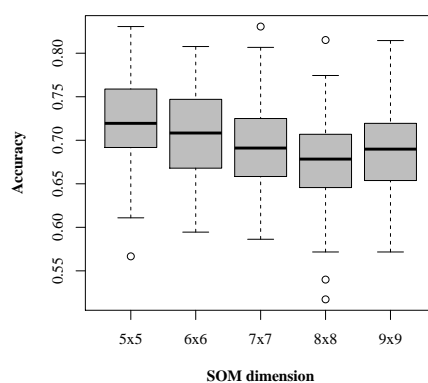
In his 2012 publication [158], Kohonen suggests to use self-organizing maps with a number of neurons between about 25 and 2000. Additionally, he asserts that it is not possible to know the exact dimension of the SOM array beforehand, but it has to be determined trying several sizes and evaluating the errors. The number of neurons in the SOM is also determined by the possible number of clusters in the dataset. Small SOMs are enough to map datasets with few clusters, whereas datasets with more than two clusters or with interesting fine structures need a higher resolution.

The 0D, 1D, 2D and 3D descriptors calculated with MOE were used to generate



(a) 0D, 1D and 2D MOE descriptors

(b) 3D MOE descriptors



(c) 0D, 1D, 2D and 3D MOE descriptors

Figure 5.4: The Box plots report the accuracies calculated for different SOM dimensions using combinations of 2D and 3D MOE descriptors.

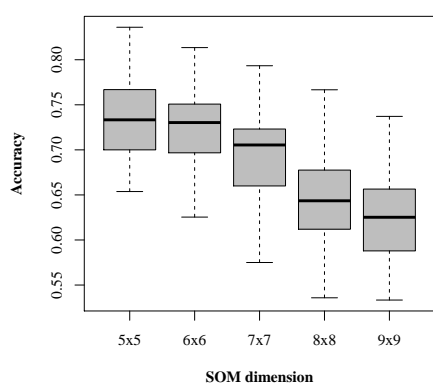
SOMs of the training dataset. As shown in figure 5.4, the accuracy for SOMs generated using 0D, 1D and 2D descriptors decreases with the size of the array. Indeed, the highest value of accuracy was obtained using SOM dimensions of 5x5 and 6x6 neurons (with accuracies equal to 0.73 and 0.72 respectively). The accuracy of SOMs generated using 3D descriptors remains almost constant with the increase of the size. The best accuracy value obtained using this set of descriptors is comparable with the best accuracy obtained with the non-3D descriptors and is associated with a SOM size of 8x8 neurons (accuracy equal to 0.73).

The use of both descriptors sets results in SOMs with accuracy decreasing with the size of the array, as already observed for SOMs calculated using non-3D descriptors. This similar trend might be due to the higher amount of non-3D descriptors with respect to the 3D (174 descriptors against 124 descriptors respectively). The best accuracy obtained using both the descriptors set together is equal to 0.71 for a SOM

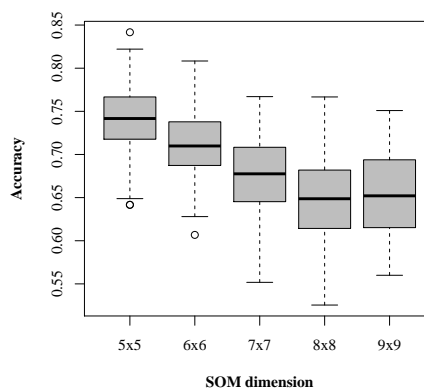
5.3 Use of SOMs to discriminate between BCRP-inhibitors and non-inhibitors

with size 5x5. This value is lower than the best values obtained using the separate datasets. This result suggests that a descriptors selection could play a role in increasing of accuracy.

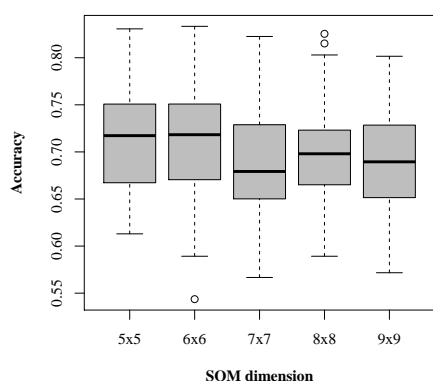
Inductive descriptors



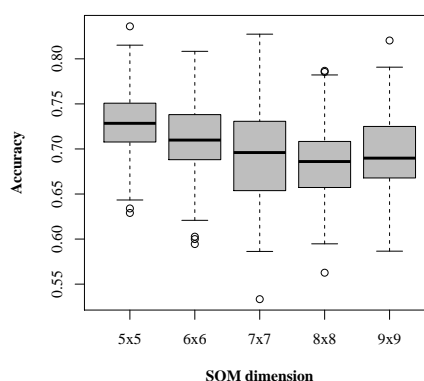
(a) Inductive descriptors.



(b) 2D MOE and inductive descriptors.



(c) 3D MOE and inductive descriptors.



(d) 2D, 3D and inductive descriptors.

Figure 5.5: Accuracies calculated for different SOM dimensions using combinations of inductive, 2D and 3D MOE descriptors.

The use of inductive descriptors leads to accuracy values strongly dependent on the size of the array. The best accuracy is obtained with an array size of 5x5 neurons (accuracy equal to 0.74). The combination of 0D, 1D, 2D and inductive descriptors leads also decreasing accuracy with the size of SOM. The best model obtained with this set of descriptors has also an accuracy of 0.74 for an array size of 5x5 neurons. The use of 3D and inductive descriptors together for the calculation of the SOM leads to models with accuracy values independent of the size of the array. The accuracy for

5 Development of binary QSAR models for classification of BCRP inhibitors.

these models is between 0.71 and 0.69. The best model (accuracy equal to 0.71) is obtained using an array size of 6x6 neurons. The model obtained by the combination of 0D, 1D, 2D, 3D and inductive descriptors is also independent of the size of the array and has the same trend as already observed for models generated using only 0D, 1D, 2D and 3D descriptors. This result suggests that the chemical information provided by the inductive descriptors is probably hidden by the other descriptors.

RDF descriptors

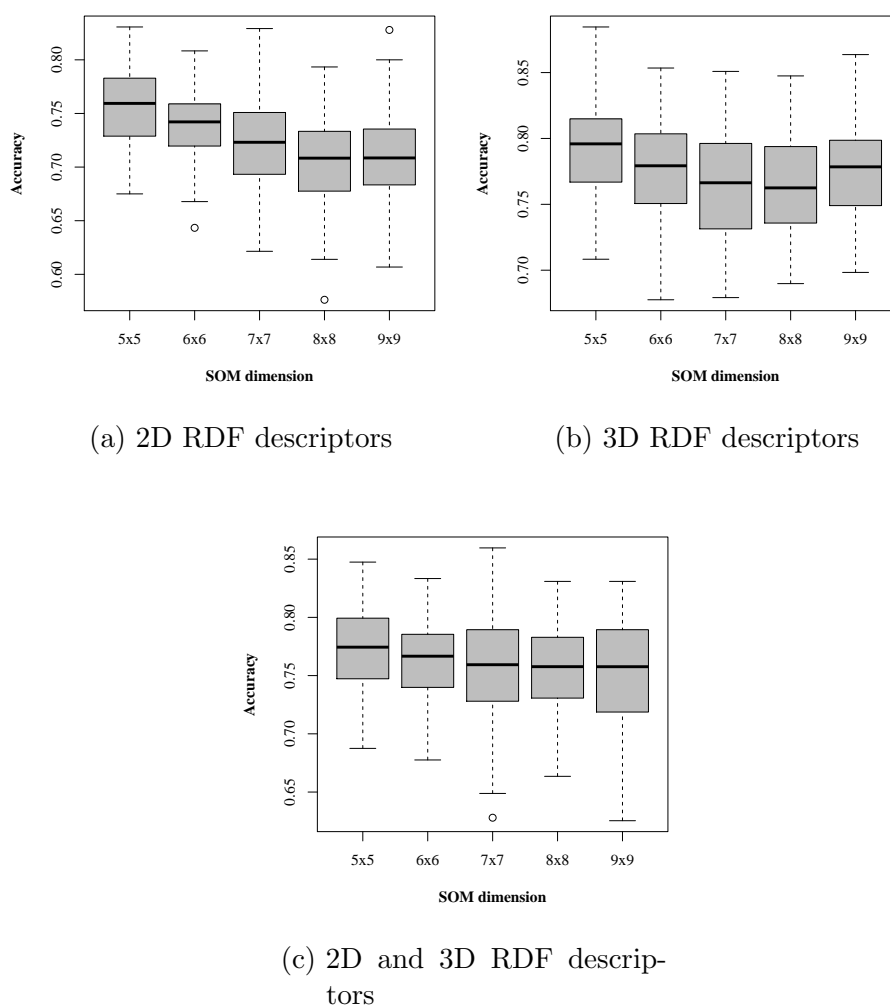


Figure 5.6: Accuracies calculated for different SOM dimensions using combinations of 2D and 3D RDF descriptors.

The accuracy of SOM models calculated using 2D RDF descriptors decreases with the size of the array. The highest accuracy is obtained by a SOM with a size of 5x5 neurons and it is equal to 0.76. The use of 3D RDF descriptors for SOMs calculation gives higher values of accuracy and they do not depend to the size of the array. The best model obtained using 3D RDF descriptors has an accuracy value equal to 0.79,

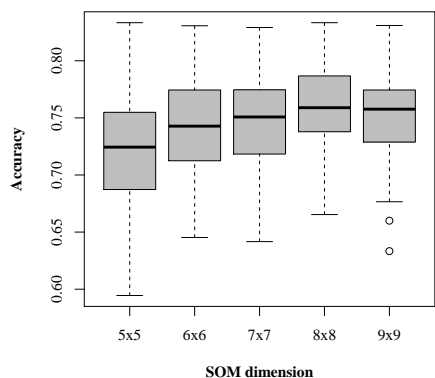
5.3 Use of SOMs to discriminate between BCRP-inhibitors and non-inhibitors

associated to a SOM with an array of 9x9 neurons. The use of 2D and 3D RDF descriptors together leads to models with accuracy that is not dependent on the array size but with lower accuracy with respect to models generated using only 3D RDF descriptors. This result could be explained considering the higher amount of 3D RDF descriptors (150 descriptors) respect to 2D RDF descriptors (only 75 descriptors).

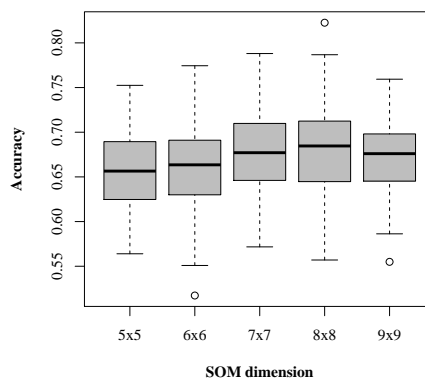
RECON descriptors

Generally, the use of RECON descriptors for calculation of SOMs leads to predictive models with low or no dependency on the size of the array. The model generated using 2D RECON descriptors showed the best accuracy values with respect to models generated using 3D RECON and TAE RECON descriptors. The accuracy of SOM models calculated using 2D RECON descriptors moderately increases with the size of the array from 0.72 for a 5x5 neurons SOM to a value of 0.76 for an 8x8 neuron SOM. The largest SOM generated using this set of descriptors (9x9 neurons) has a similar accuracy, equal to 0.75. The SOM models generated using 3D RECON descriptors, in contrast to what observed in the previous descriptor sets, has an accuracy lower than models calculated using 2D RECON descriptors. Indeed, all the models calculated using 3D RECON descriptors have an accuracy lower than 0.70, without dependency on the size of the array. The combination of 2D and 3D RECON descriptors leads to models with an accuracy slightly under what obtained using only 2D RECON descriptors. The worsening of the accuracy value could be due to the presence of descriptors uncorrelated with the biological activity of compounds, as already observed for other descriptors sets. The SOM models calculated using TAE RECON descriptors show low predictivity, with accuracy value of about 0.60 and no dependency on the size of the array. The use of TAE descriptors in combination with the other RECON descriptors also leads to models with lower predictivity with respect to the models calculated only with 2D and 3D RECON descriptors.

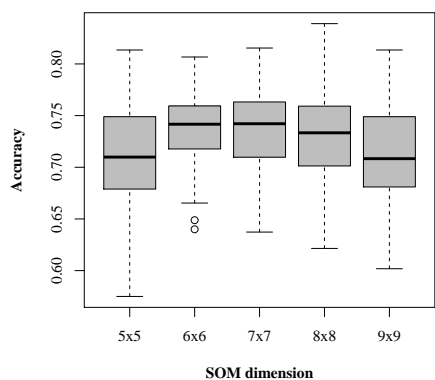
5 Development of binary QSAR models for classification of BCRP inhibitors.



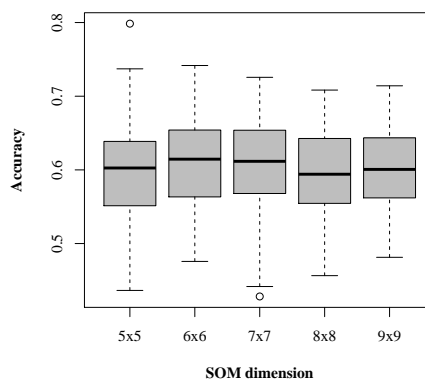
(a) 2D RECON descriptors



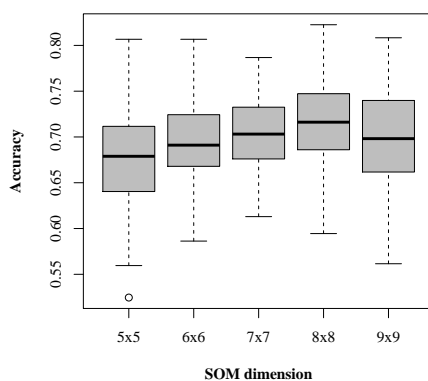
(b) 3D RECON descriptors



(c) 2D and 3D RECON descriptors



(d) TAE RECON descriptors



(e) 2D, 3D and TAE RECON descriptors

Figure 5.7: Accuracies calculated for different SOM dimensions using combinations of 2D, 3D and TAE RECON descriptors.

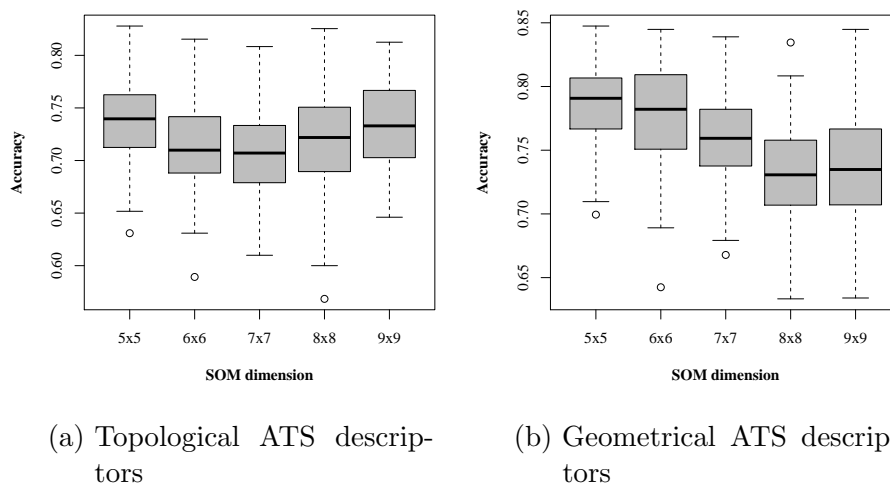
ATS descriptors

Figure 5.8: Accuracies calculated for different SOM dimensions using combinations of topological and geometrical ATS descriptors.

The SOMs generated using ATS descriptors have high accuracy values. The use of topological ATS descriptors leads to models with accuracy between 0.74 (5x5 neurons) and 0.71 (7x7 neurons). The accuracies observed for those models are not correlated with the size of the array. The models generated using geometrical ATS descriptors have values of accuracy higher than the models generated with topological ATS descriptors. The better quality of these models can be explained by the nature of the geometrical ATS descriptors: These descriptors consider the three-dimensional structure of the compounds and not only the connection matrix, as for the topological ATS descriptors. The best accuracy obtained for SOMs calculated with geometrical ATS descriptors is associated with an array size of 5x5 neurons and is equal to 0.79.

Dragon descriptors

The accuracies of SOMs generated using dragon descriptors are generally low. The models generated using 3d-MORSE and GETAWAY descriptors are not able to discriminate between BCRP inhibitors and non-inhibitors, with accuracy values lower than 0.70. Only the models based on WHIM descriptors have fair values of accuracy, increasing with the size of the array. The best value of accuracy obtained by a SOM calculated with the WHIM descriptors is equal to 0.73, and it is associated to an array of 9x9 neurons.

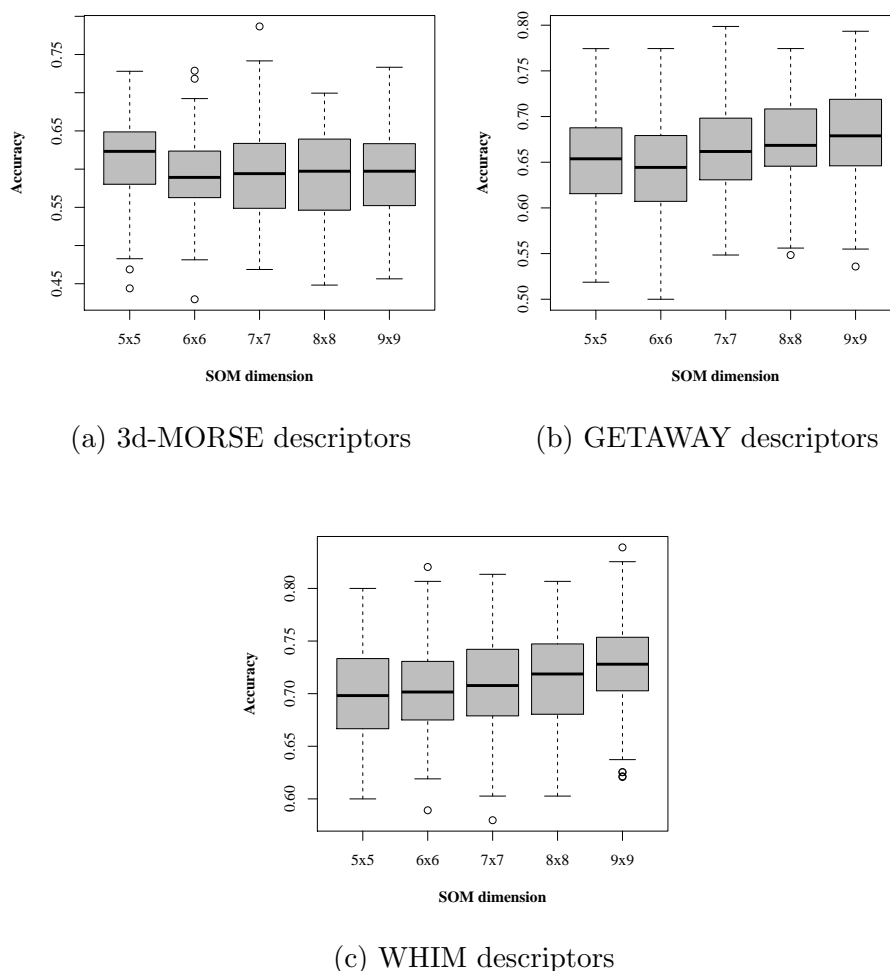


Figure 5.9: Accuracies calculated for different SOM dimensions using dragon descriptors.

Discussion of results

The results summarized in fig 5.10 shown that the best values of accuracy are obtained using RDF and geometrical ATS descriptors. In particular (table 5.3), the accuracy of the models generated using 3D RDF descriptors in a 5x5 neurons SOM is equal to 0.79, with a true positive precision equal to 0.91 and a true negative precision equal to 0.68. The accuracy of the model generated using gATS descriptors is comparable to the accuracy obtained by the model calculated using 3D RDF descriptors. Also in this case the accuracy is equal to 0.79, but the precisions of true positive and true negative are 0.90 and 0.67, respectively.

Several chemical-physical descriptors for the compounds in the database were also analyzed, in order to observe possible differences in the distribution of these descriptors between active and weakly active compounds. The best statistically relevant separation is based on descriptors codifying the volume of the molecules and the molecular surface. In particular, the difference of the mean values of van der Waals surface area

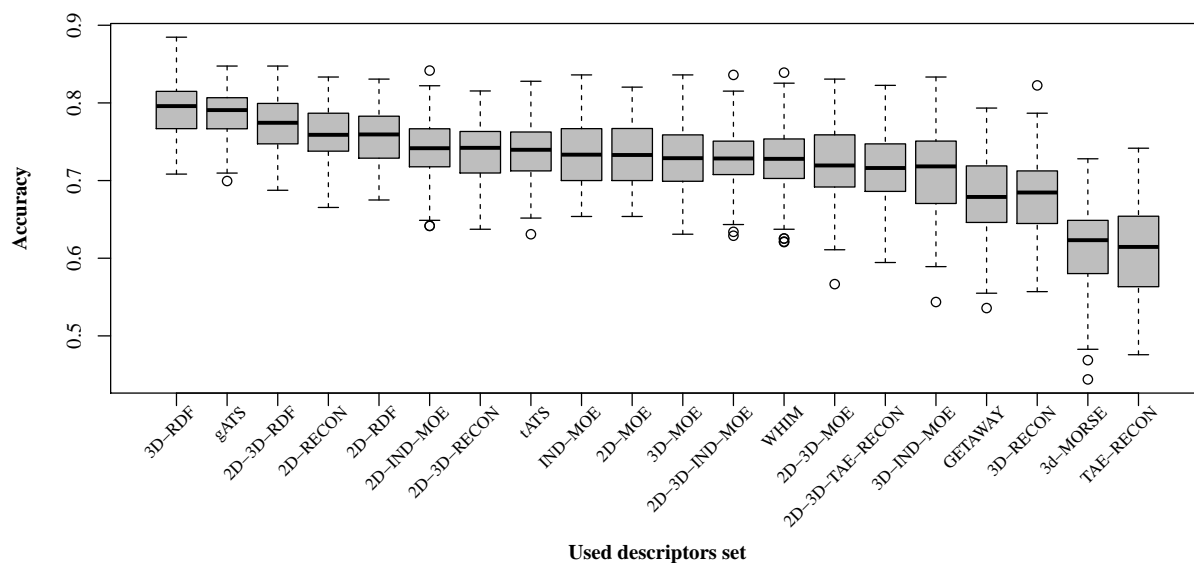


Figure 5.10: Accuracy values of the best models for each descriptor set.

between high and low active BCRP inhibitors is well rendered (Figure 5.11a). The Welch Two Sample t-test was used to evaluate the statistical significance. The result of the t-test showed a p value <0.005 and $t = 3.685$, confirming that the two population of values are statistically separated.

The van der Waals surface area values were also correlated with the first component of 3D RDF and gATS descriptors (Figure 5.11c). As expected, the values are highly correlated: the correlation between the first component of the 3D RDF descriptors and VSA has r^2 equal to 0.97 and the correlation between the first component of the gATS descriptors and VSA has an r^2 value equal to 0.98.

In order to investigate the role of logP on the activity of this class of compounds, the mean values of logP for the two populations were compared (Figure 5.11b). The result of the Welch Two sample t-test showed that the two mean values are not significantly different (p value = 0.13, $t = 1.52$). In conclusion, the capacity of this class of compounds to inhibit BCRP depends on the surface area of the molecule but not on the lipophilicity. A higher value of surface area gives higher probability that the compound is a BCRP-inhibitor.

In Fig. 5.13 is shown the distribution of active and weakly active BCRP inhibitors in the SOM. Red areas are neurons associated with weakly active BCRP-inhibitors, while beige areas are neurons associated with active BCRP inhibitors. For both SOMs, generated using 3D RDF and gATS descriptors, it is possible to distinguish well-separated clusters of neurons associated with active and weakly active inhibitors. The separation of the two clusters is already recognizable observing the principal component analysis of the 3D RDF and gATS descriptors. As illustrated in figure 5.13a, it is possible to

Table 5.3: Summary of BCRP classification powers by SOM approach.

Descriptors set	Size	Accuracy	TP rate	TN rate
3D-RDF	5x5	0.79	0.91	0.68
gATS	5x5	0.79	0.90	0.67
2D-3D-RDF	5x5	0.77	0.89	0.66
2D-RECON	8x8	0.76	0.87	0.65
2D-RDF	5x5	0.76	0.87	0.64
2D-IND-MOE	5x5	0.74	0.85	0.63
2D-3D-RECON	7x7	0.74	0.84	0.63
tATS	5x5	0.74	0.84	0.63
IND-MOE	5x5	0.74	0.84	0.63
2D-MOE	5x5	0.73	0.85	0.62
3D-MOE	8x8	0.73	0.83	0.63
2D-3D-IND-MOE	5x5	0.73	0.83	0.63
WHIM	9x9	0.73	0.82	0.63
2D-3D-MOE	5x5	0.72	0.82	0.62
2D-3D-TAE-RECON	8x8	0.72	0.81	0.62
3D-IND-MOE	6x6	0.71	0.81	0.61
GETAWAY	9x9	0.68	0.76	0.60
3D-RECON	8x8	0.68	0.76	0.60
3d-MORSE	5x5	0.61	0.67	0.56
TAE-RECON	6x6	0.61	0.67	0.55

distinguish the separation of the two clusters in the principal component analysis of 3D RDF descriptors. The first two components explain 64.4% of the variance. Figure 5.13c shows the 3D plot of the first 3 principal components of the 3D RDF descriptors: in this figure the separation of the two clusters is more evident. In this last case, the first three components explain 72.1% of the variance.

Similar results are obtained with the principal component analysis of the gATS descriptors (Figure 5.13b and 5.13d). Also using these descriptors the separation of the clusters in the 2D and 3D plot is clear. In this case, the first two components explain 65,6% of the variance and the first three components 74,5% of the variance.

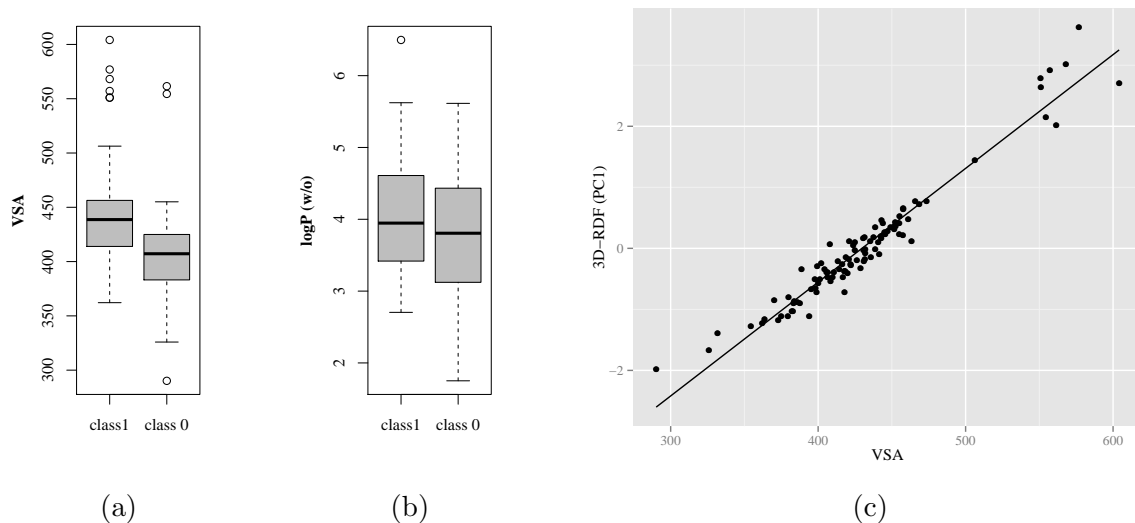


Figure 5.11: (a) Differences between inhibitors (class 1) and weak inhibitors (class 0) based on the van der Waals surface area (VSA), and (b) the logP. (c) Correlation between the 1st principal component of the 3D-RDF descriptors and the van der Waals surface area.

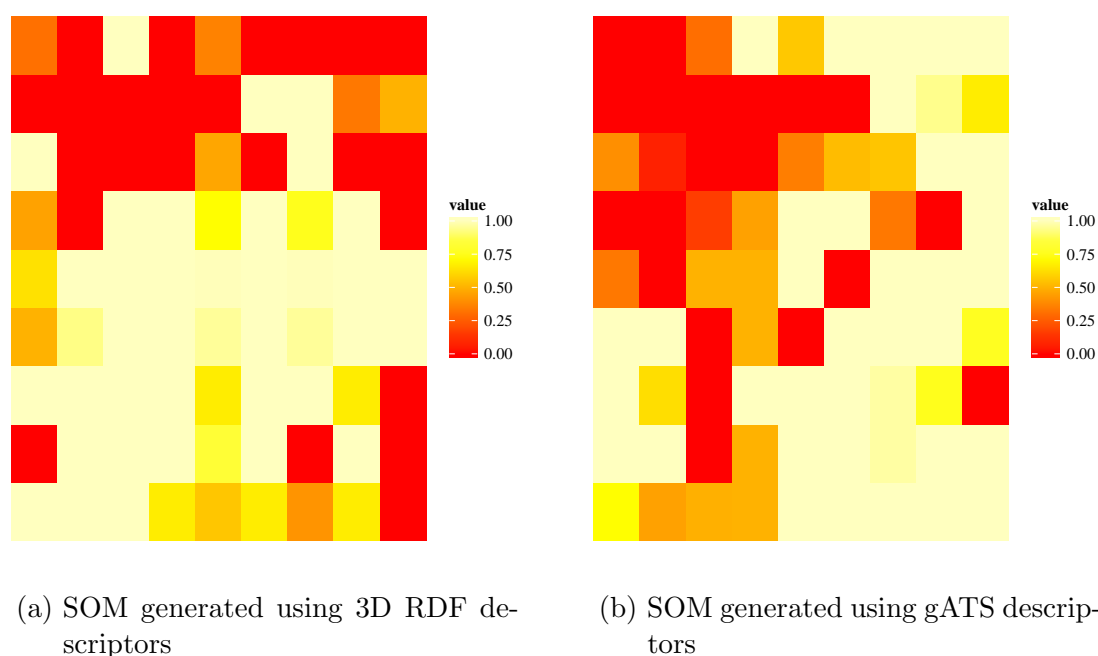


Figure 5.12: SOMs generated using the two best descriptors sets. The color of each neuron indicates the rate of association of the neuron to active (value = 1) or inactive (value = 0) compounds.

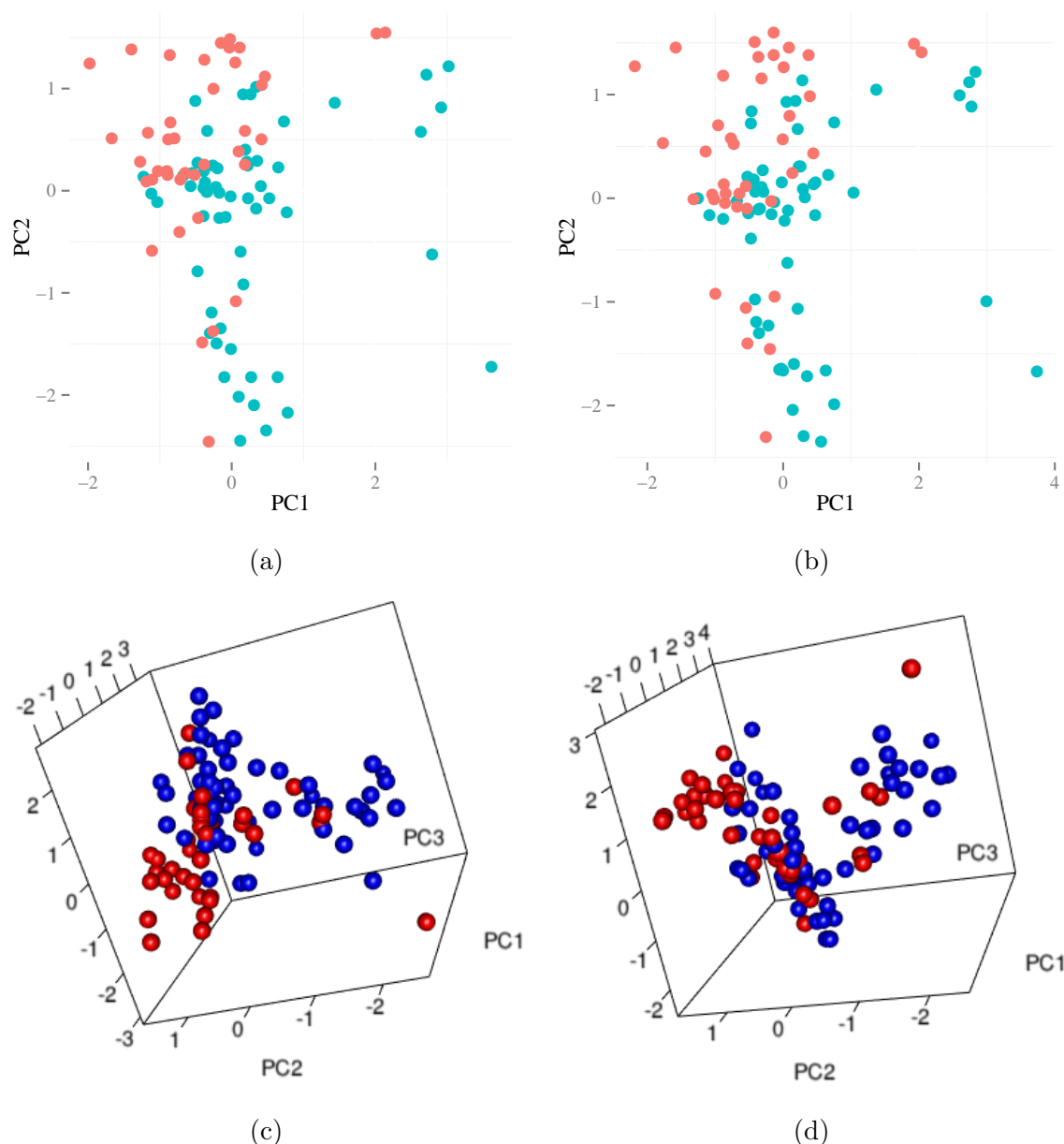


Figure 5.13: (a) Plot of the two principal components of the 3D RDF descriptors and (b) topological ATS descriptors. (c) 3D plot representing the three principal components of the 3D RDF descriptors and (d) topological ATS descriptors. The blue points and spheres are associated to active BCRP inhibitors, while red points and spheres are associated to weakly active BCRP inhibitors.

Use of an external testset

The predictivity of the developed SOM was tested using an external testset. The external dataset consists of 16 compounds synthesized and tested by Dr. Lars Möllmann [159]. 9 compounds have IC_{50} values lower than $10 \mu\text{M}$ and were classified as BCRP-inhibitors (class 1), while 7 compounds have IC_{50} values higher than $10 \mu\text{M}$ or show no activity against BCRP. These compounds were classified as non-inhibitors (class 0). The three-dimensional alignment of the active compound LM-221 with the BCRP-

5.3 Use of SOMs to discriminate between BCRP-inhibitors and non-inhibitors

inhibitor KS-176 (figure 5.14) shows that three of the four aromatic rings of compound LM-221 can be superposed with the three aromatic rings of compound KS-176. The result of the three-dimensional alignment suggests that the two classes of compounds share similar pharmacophore patterns and consequentially also a similar interaction with the protein binding site. If we assume that compounds of two different chemical classes, bind to the protein in the same binding site should have similar chemical features to be active, than we can also expect that they occupy similar positions in the chemical space and, consequentially, the same region of the SOM.

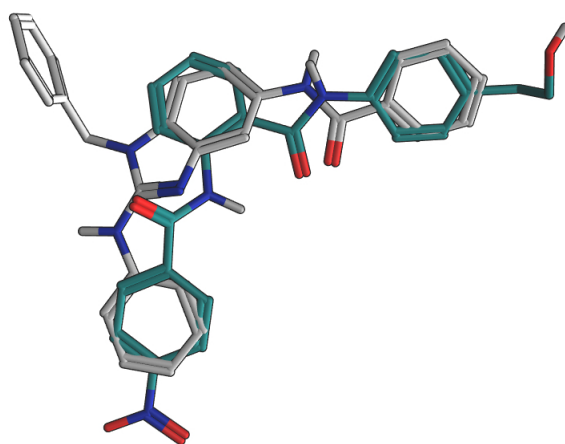


Figure 5.14: Superposition of KS-176 and LM-221.

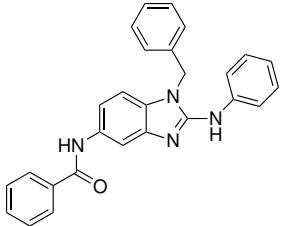
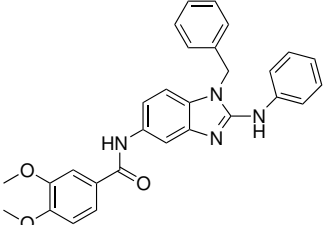
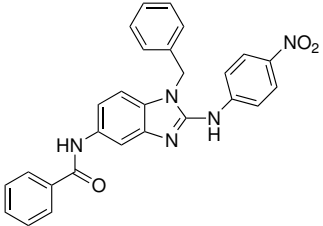
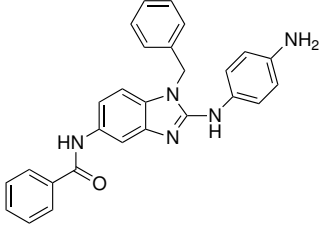
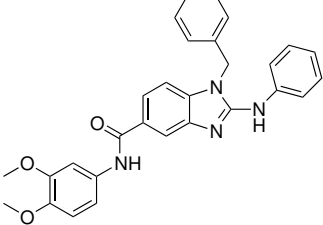
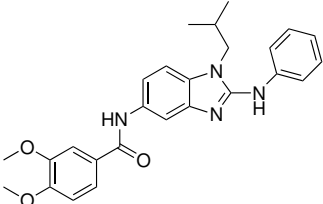
In order to use the LM dataset as an external test set, the KS dataset and the LM dataset were joined together. The combined descriptor matrix obtained by joining the two separate descriptor matrices was normalized using the software R. The normalized matrix was split again into training set and test set (the KS compounds are used as training set and the LM compounds as test set). The normalized descriptor values were used for training of the SOM and the test compounds were successively projected on the SOM model. The accuracy of the prediction was calculated according to the amount of correct predictions for the two classes.

The SOM model generated using 3D RDF descriptors is able to predict the activity of the test set with an accuracy of 0.88. All the active compounds were correctly predicted and 5 of 7 inactive compounds were also correctly predicted. The two false positive compounds are LM-234 and LM-233. Both compounds have a chemical structure very similar to the active compounds. LM-234 is an analog of compounds LM-235 but it has a thiourea as linker instead of the urea of compound LM-235. The other false positive compound, LM-233 is similar to compound LM-221, but its structure has a cyclohexane instead of a phenyl ring. The active compounds of the test set are grouped in a cluster localized in the region of the SOM corresponding to active compounds. The correct predicted inactive compounds are localized in regions of the SOM associated

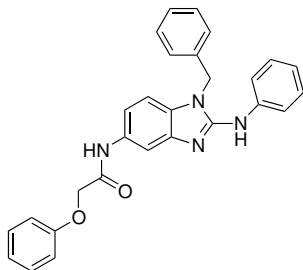
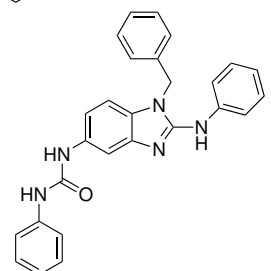
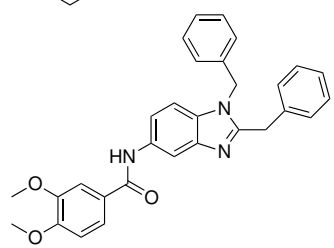
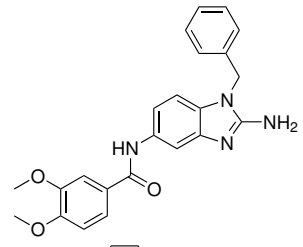
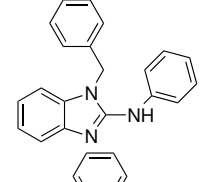
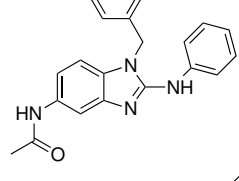
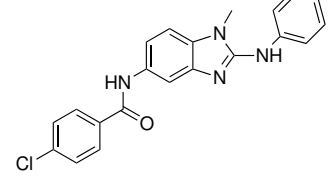
5 Development of binary QSAR models for classification of BCRP inhibitors.

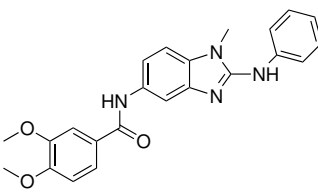
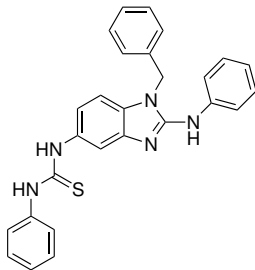
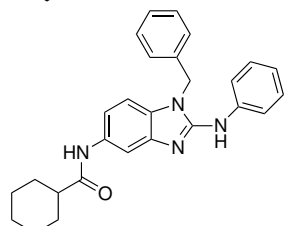
with inactive compounds.

Table 5.4: Activity of compounds used as test set. The structures and activity of these substances were extracted from the PhD thesis of Lars Mollmann [159]. Activities were measured using Hoechst 33342-assay in MCF-7 MX cells for BCRP (reference substance WK-X-24: $pIC_{50} = 6.047 \pm 0.032$).

Structures	Compounds	pIC_{50} BCRP μM \pm SD	Class
	LM-221 (1)	5.816 ± 0.032	1
	LM-222 (2)	5.527 ± 0.067	1
	LM-223 (3)	6.204 ± 0.048	1
	LM-229 (4)	5.163 ± 0.101	1
	LM-224 (5)	5.790 ± 0.056	1
	LM-228 (6)	5.432 ± 0.058	1

5.3 Use of SOMs to discriminate between BCRP-inhibitors and non-inhibitors

Structures	Compounds	pIC ₅₀ BCRP μ M \pm SD	Class
	LM-236 (7)	5.989 \pm 0.055	1
	LM-235 (8)	5.923 \pm 0.022	1
	LM-209 (9)	5.801 \pm 0.020	1
	LM-75 (10)	4.387 \pm 0.116	0
	LM-219 (11)	n.a.	0
	LM-220 (12)	n.a.	0
	LM-225 (13)	n.a.	0

Structures	Compounds	pIC ₅₀ BCRP μ M \pm SD	Class
	LM-226 (14)	4.522 \pm 0.029	0
	LM-234 (15)	n.a.	0
	LM-233 (16)	n.a.	0

The SOM model generated using gATS descriptors is able to correctly predict the activity of the test set with an accuracy value equal to 0.81. This model, as well as the model generated using 3D RDF descriptors, is able to correctly classify all the active compounds of the dataset. The active compounds are all localized in a cluster corresponding to the area of the SOM associated with active compounds. Three of the seven inactive compounds are not correctly predicted. Two of them, LM-234 and LM-233 were not correctly predicted also by the model generated using 3D RDF descriptors. The third un-correct predicted compound, LM-220 is localized on an ambiguously classified neuron.

In conclusion, self-organizing maps are able to discriminate between inhibitors and non-inhibitors in the KS-dataset. The trained SOM generated using 3D RDF and gATS descriptors can be used to predict the BCRP-inhibitory activity of other compounds not structurally directly correlated to KS compounds.

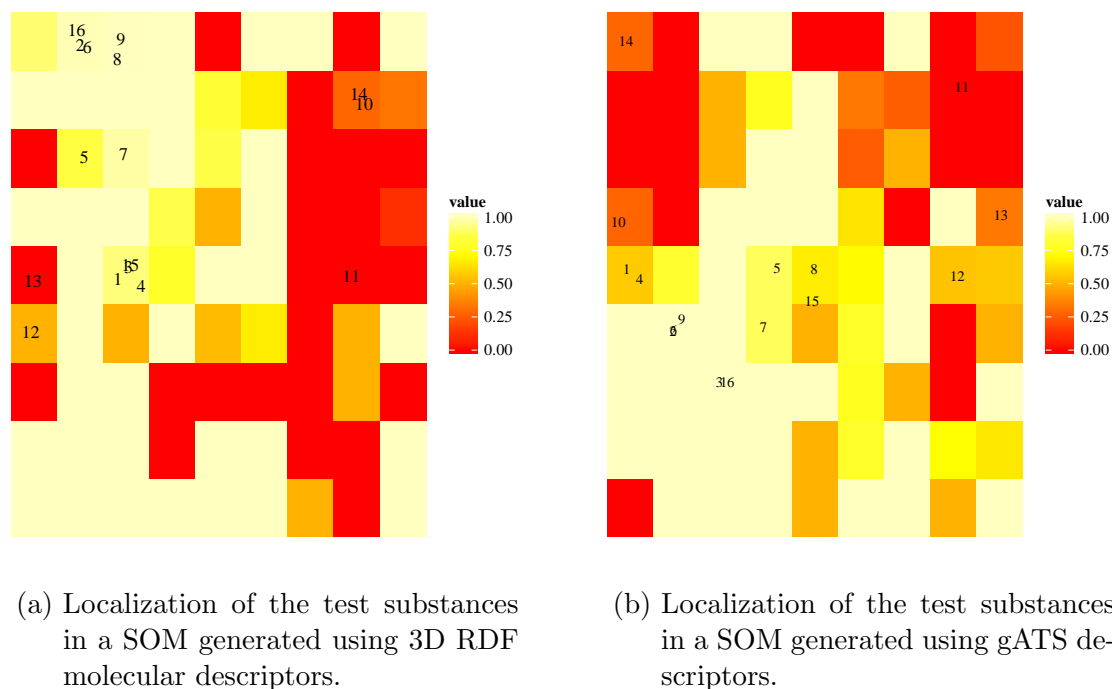


Figure 5.15: Use of externe dataset for validation. The numbers in these two illustrations are referred to the compounds in table 5.4.

5.4 Use of SVM to discriminate between BCRP-inhibitors and non-inhibitors

5.4.1 Generation of the models

In order to distinguish the BCRP-inhibitors and non-inhibitors in the KS dataset, the SVM algorithm was also applied, analogously to the SOM algorithm. The SVM models were built using the same descriptor sets previously used for the SOM models. For each SVM model based on the RBF kernel, it is necessary to find the optimal combination of the two parameters C and γ . In order to define the optimal combination of these two parameters, a grid search was performed. Several SVM models were generated, using several combinations of C and γ for values of C between 2^{-1} and 2^{12} and for values of γ between 2^{-12} and 1. The quality of each model was evaluated by 10-fold cross-validation. For each descriptor set, the combination of C and γ that generates the model with the lowest 10-fold cross-validated error was chosen to generate the final SVM model.

The workflow used to generate and evaluate the SVM models is similar to that used to generate the SOM models. First, all the descriptors with variance equal to 0 were deleted from the descriptor matrix. Then the descriptor values were normalized. Similarly to SOM models, also in SVM a down-sampling step was performed, due to

the different amount of compounds in the two classes. The SVM models were generated using the package e1071 [160] implemented in R. The downsampling and the model generation were performed 1000 times and the average of the accuracy is reported. Small differences of accuracy for each single cycle are due to the different elements randomly selected in the downsampling. For each descriptor set, the best model was evaluated not only with cross-validation, but also using an external testset, as for the SOM models. Also in this case the external testset is the LM dataset.

5.4.2 Evaluation of the results

Table 5.5: Summary of BCRP classification powers by SVM approach.

Descriptors set	C	γ	Accuracy	TP rate	TN rate
3D-RDF	16	2^{-12}	0.75	0.89	0.62
gATS	2	2^{-10}	0.77	0.93	0.60
2D-3D-RDF	2	2^{-7}	0.77	0.90	0.64
2D-RECON	4	2^{-12}	0.67	0.95	0.38
2D-RDF	1	2^{-5}	0.76	0.91	0.60
2D-IND-MOE	0.5	2^{-9}	0.70	0.90	0.50
2D-3D-RECON	64	2^{-11}	0.74	0.82	0.66
tATS	32	2^{-12}	0.75	0.91	0.59
IND-MOE	2	2^{-8}	0.75	0.90	0.60
2D-MOE	16	2^{-12}	0.72	0.90	0.55
3D-MOE	16	2^{-12}	0.71	0.88	0.54
2D-3D-IND-MOE	1	2^{-10}	0.71	0.91	0.50
WHIM	1	2^{-9}	0.70	0.93	0.48
2D-3D-MOE	8	2^{-11}	0.71	0.85	0.58
2D-3D-TAE-RECON	4	2^{-11}	0.72	0.90	0.55
3D-IND-MOE	8	2^{-11}	0.73	0.89	0.56
GETAWAY	16	2^{-8}	0.64	0.74	0.53
3D-RECON	32	2^{-12}	0.70	0.85	0.55
3d-MORSE	16	2^{-11}	0.70	0.80	0.60
TAE-RECON	8	2^{-12}	0.70	0.90	0.50

For each descriptor set, the accuracy values and the percentage of true predicted compounds for each class are reported in table 5.5. The best accuracy values were obtained using 2D and 3D RDF descriptors (and their combination), gATS, tATS and inductive descriptors. The accuracy values of the SVM models are generally lower than the accuracy values of the corresponding SOM models. The SVM model generated using 3D RDF descriptors leads to an accuracy equal to 0.75, lower than the accuracy found for the SOM model with the same descriptors set (the accuracy is equal to 0.79 for the SOM model). The SVM model generated using gATS descriptors has also a lower value of accuracy (0.77) with respect to that obtained by the SOM model (0.79).

However, in this case, the difference between the two accuracy values is rather low. The use of tATS descriptors (0.75), inductive descriptors (0.75) and the combination of 2D and 3D RDF descriptors (0.71) leads to SVM models with accuracy values comparable to the corresponding SOM models (0.74, 0.74, and 0.72, respectively). The difference between the accuracy of the SVM model obtained using 2D RECON descriptors and the accuracy of the SOM model obtained using the same descriptor set is considerable: their values are equal to 0.67 for the SVM model and 0.76 for the SOM model. The loss of predictivity of the SVM model is principally due to its very low predictivity for the non-inhibitors compounds, compared to the SOM model.

The low predictivity for the non-inhibitors is observed for almost all the SVM models (but also for SOMs models), although the downsampling step before building of the models. The low predictivity of inactive compounds can be explained by the use of a too low threshold (equal to 10 μM) that discriminate between active and inactive compounds. As a result of this, the inactive compounds share a large part of their chemical space with the active compounds. On the other hand many active compounds occupy a region of the chemical space where almost no inactive compounds are localized. It is also possible to observe the unbalanced intersection of the two clusters in the 2D and 3D representation of the chemical space for 3D RDF and gATS descriptors (Figures 5.13a, 5.13b, 5.13c, and 5.13d). For this particular case of unbalanced clusters intersection, the down-sampling is not enough to generate models with a balanced prediction for active and inactive compounds.

5.4.3 Use of an external testset

The predictivity of the generated SVM model was investigated using an external testset. Analogously to the SOM models, the LM dataset was used as test set to test the predictivity of the generated models. The SVM models used to predict the activity of the LM dataset were chosen according to the accuracy values associated with the models. The best two SVM models were generated using a combination of 2D and 3D RDF descriptors and gATS descriptors.

In contrast to what was observed for the SOM models, SVM models are not able to effectively discriminate between active and inactive compounds of the LM dataset. All the active compounds of the test set were correctly predicted, but only 2 compounds (LM-219 and LM-225) belonging to the inactive compounds are correctly predicted. The predictivity of the SVM model generated using gATS descriptors is slightly better than the model generated using 2D and 3D RDF descriptors. This model is able to correctly predict all the active compounds of the test set but only 3 of the inactive compounds (LM-219, LM-220, LM-225).

5.5 Use of k-NN to discriminate between BCRP-inhibitors and non-inhibitors

5.5.1 Generation of the models

The k-NN is the fastest and simplest classification algorithm used in this work [161]. The only parameter that needs to be optimized is the k value. It indicates the number of neighbors around an element, for which its class has to be determined. The best value of k was systematically searched generating several models using values between 1 and 9. Only odd values of k were used, in order to avoid cases of draws (where the number of neighbors of the two classes are the equal: in this case, it is not possible to determinate the class of the query element). The best value of k for each descriptor set is reported in table 5.6. The workflow used to generate and evaluate the k-NN models is analogous to that used to generate the SOM and the SVM models. The descriptors with variance equal to 0 were deleted from the descriptor matrix. Then, the matrix values were normalized. The following step is the downsampling of the elements in the majority class (active substances). The downsampling and the generation of the model were performed 1000 times and the average of the accuracy were reported. The best models were used to predict the activity of an external dataset.

5.5.2 Evaluation of the results

The best k-NN models were obtained, as already observed for SOM and SVM, using RDF and ATS descriptors. In particular, the use of 3D RDF descriptors, the combination of 2D and 3D RDF descriptors and gATS descriptors leads to the best models. The accuracy value of the k-NN model generated using 3D RDF descriptors and of the model obtained using 2D and 3D RDF descriptors together are 0.74 and 0.75, respectively. Although the second model has the best accuracy, its capacity to predict inactive compounds is lower than in the model based on only 3D RDF descriptors. The model built using gATS descriptors has an accuracy equal to 0.73. This result confirms that the gATS descriptors, as well as RDF are a good choice for building machine learning models useful to discriminate between active and inactive compounds of a dataset.

5.5.3 Use of an external testset

The best k-NN models were used to predict the activity of the LM test set. The three best model based on the k-NN algorithm are obtained using 3D RDF descriptors, 2D and 3D RDF descriptors together and gATS descriptors, the same descriptor sets

Table 5.6: Summary of BCRP classification powers by kNN approach.

Descriptors set	k	Accuracy	TP rate	TN rate
3D-RDF	7	0.74	0.85	0.63
gATS	5	0.73	0.87	0.58
2D-3D-RDF	5	0.75	0.92	0.58
2D-RECON	3	0.70	0.85	0.54
2D-RDF	5	0.71	0.88	0.53
2D-IND-MOE	3	0.69	0.91	0.46
2D-3D-RECON	3	0.63	0.74	0.52
tATS	7	0.67	0.87	0.47
IND-MOE	3	0.66	0.82	0.50
2D-MOE	3	0.63	0.79	0.47
3D-MOE	7	0.62	0.80	0.44
2D-3D-IND-MOE	3	0.63	0.77	0.49
WHIM	3	0.66	0.84	0.47
2D-3D-MOE	3	0.63	0.72	0.54
2D-3D-TAE-RECON	7	0.59	0.85	0.33
3D-IND-MOE	3	0.62	0.80	0.45
GETAWAY	7	0.55	0.74	0.36
3D-RECON	5	0.67	0.85	0.49
3d-MORSE	5	0.55	0.65	0.46
TAE-RECON	5	0.58	0.86	0.31

as in case of SOM and SVM models. The results show that, analogously to the SOM algorithm, the k-NN based models are mostly able to predict the activity of the compounds in the external testset. In particular, the model generated using 3D RDF descriptors and the model generated using gATS descriptors are both able to correctly predict the active compounds of the LM testset. The inactive compounds are only partially correctly predicted: The compounds LM-226, LM-234 and LM-233 are wrongly predicted as active compounds. The k-NN model generated using a combination of 2D and 3D RDF descriptors showed better results with respect to the other two models: The active compounds of the LM dataset were all correctly predicted, while only two compounds (LM-234 and LM-233) of the inactive class were wrongly predicted. These compounds are also predicted as active compounds by the SOM and SVM algorithms.

5.6 Comparison of the results

The comparison of the results obtained with the three machine learning algorithms shows that all three algorithms are able to discriminate between inhibitors and non-inhibitors of BCRP in a dataset of structurally highly similar compounds. The results show that the most predictive classification algorithms between the three tested in this

5 Development of binary QSAR models for classification of BCRP inhibitors.

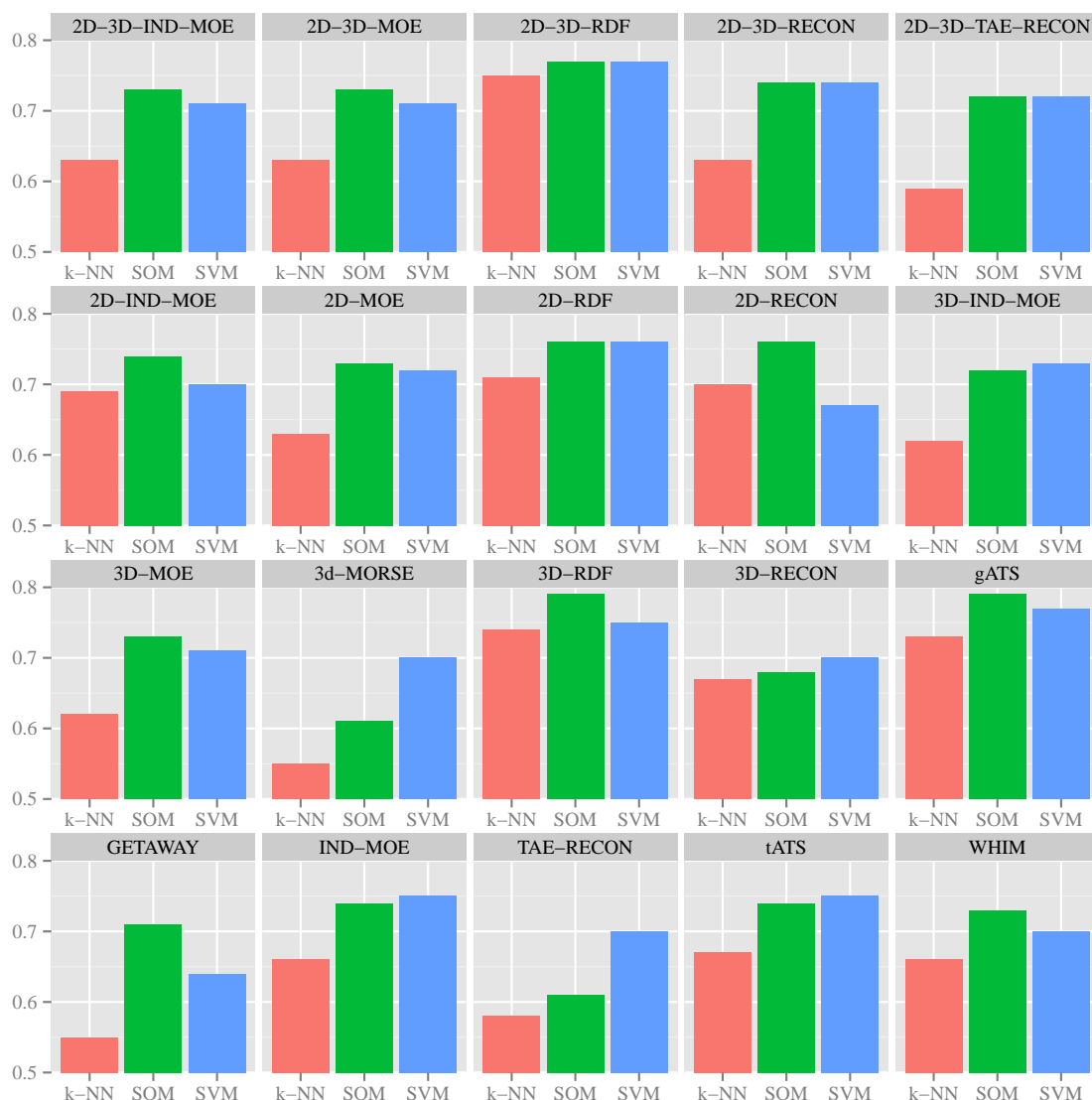


Figure 5.16: Summary of results.

work are the SOM and SVM algorithms. The best results obtained with these three algorithms were given using RDF descriptors (3D or the combination of 2D and 3D descriptors) and gATS descriptors. As shown in figure 5.16, the use of RDF descriptors and gATS descriptors leads to high values of accuracy for the training set, as well as comparable values of accuracy between the SOM and SVM classification algorithms, while the k-NN classification algorithm is slightly worse than the other two. This results suggest that the used descriptor sets are able to separate the two classes of the dataset almost independently to the used algorithm.

The use of an external testset shows that the most predictive models generated using the KS dataset are able to correctly classify also other BCRP-inhibitor candidates, with less structural similarity. The best results are obtained using the SOM algorithm and the k-NN algorithm, while the SVM algorithm gives the worst results in terms of

Table 5.7: Summary of the predictions of the test set.

Method	3D RDF				2D and 3D RDF				gATS			
	TP	TN	FP	FN	TP	TN	FP	FN	TP	TN	FP	FN
SOM	9	5	2	0	9	4	3	0	9	4	3	0
SVM	9	3	4	0	9	2	5	0	9	3	4	0
k-NN	9	4	3	0	9	5	2	0	9	4	3	0

predictivity. As reported in table 5.7, the correct number of predictions is higher for SOM and k-NN, independent of the used descriptor sets. The absolutely best values are obtained using the SOM algorithm and 3D RDF descriptors or the k-NN algorithm with the combination of 2D and 3D RDF descriptors. As shown in table 5.8, all the used training set and algorithm are able to correctly recognize the BCRP-inhibitors of the test set. Between the non-inhibitors, only two compounds are never correctly predicted: LM-234 and LM-233. This limit of the prediction capacity could be due to the extremely high structural correlation between these two compounds and the majority of the active compounds of the test set.

Table 5.8: Prediction of the class of the substances in the test set. A value of "1" indicates active compounds (BCRP inhibitors) while "0" indicates inactive compounds.

Compound	Class	3D RDF			2D and 3D RDF			gATS		
		SOM	SVM	k-NN	SOM	SVM	k-NN	SOM	SVM	k-NN
LM-221	1	1	1	1	1	1	1	1	1	1
LM-222	1	1	1	1	1	1	1	1	1	1
LM-223	1	1	1	1	1	1	1	1	1	1
LM-229	1	1	1	1	1	1	1	1	1	1
LM-224	1	1	1	1	1	1	1	1	1	1
LM-228	1	1	1	1	1	1	1	1	1	1
LM-236	1	1	1	1	1	1	1	1	1	1
LM-235	1	1	1	1	1	1	1	1	1	1
LM-209	1	1	1	1	1	1	1	1	1	1
LM-75	0	0	1	0	0	1	0	0	1	0
LM-219	0	0	0	0	0	0	0	0	0	0
LM-220	0	0	0	0	0	1	0	1	0	0
LM-225	0	0	0	0	0	0	0	0	0	0
LM-226	0	0	1	1	0	1	0	0	1	1
LM-234	0	1	1	1	1	1	1	1	1	1
LM-233	0	1	1	1	1	1	1	1	1	1

5 Development of binary QSAR models for classification of BCRP inhibitors.

In conclusion, it is possible to discriminate between inhibitors and non-inhibitors in a dataset of BCRP-inhibitors, also if the compounds are structurally highly similar, as in the case of the KS dataset. In this case, the best results are obtained using the SOM algorithm with the RDF and gATS descriptors calculated for the compounds in the dataset. The generated models could be used to predict the binary activity of other classes of substances.

6 Structure-activity relationships of quinazolines derivatives as inhibitors of BCRP.

6.1 Structural and activity data

The 36 quinazolines included in this work were synthesized by Dr. Kapil Juvale and were taken from his PhD Thesis [162]. The activity data were also measured by Kapil Juvale using the Hoechst 33342 assay at a final concentration of 1 μ M in MDCK BCRP cells. In figure 6.1 the general structure of the quinazolines included in this work is reported and in table 6.1 the activities (pIC_{50}) of the compounds are reported.

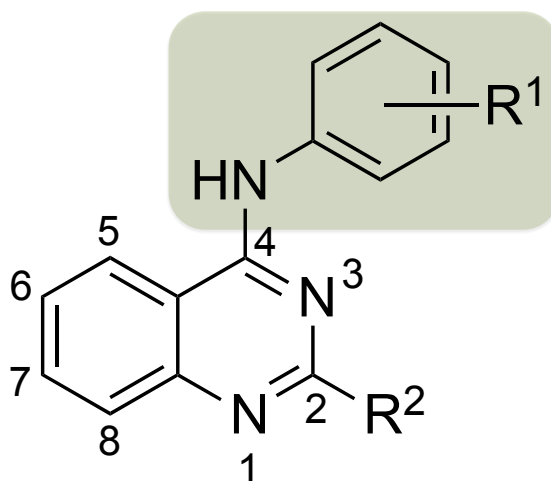


Figure 6.1: The general structure of the quinazolines used in this work. The aniline sub-structure is highlighted in green.

6 Structure-activity relationships of quinazolines derivatives as inhibitors of BCRP.

Table 6.1: The compounds included in the analysis..

Compound	pIC ₅₀	R ¹	Positions 6 and 7	R ²
KJ-98	5.59	H	H	Phenyl
KJ-99	5.43	2-Br	H	Phenyl
KJ-100	6.24	3-Br	H	Phenyl
KJ-101	5.17	4-Br	H	Phenyl
KJ-102	5.72	3-Cl	H	Phenyl
KJ-103	5.78	3-Cl,4-F	H	Phenyl
KJ-104	5.65	2-OCH ₃	H	Phenyl
KJ-105	5.88	3-OCH ₃	H	Phenyl
KJ-106	5.72	4-OCH ₃	H	Phenyl
KJ-109	6.90	3-NO ₂	H	Phenyl
KJ-110	6.01	4-NO ₂	H	Phenyl
KJ-112	6.85	3-CN	H	Phenyl
KJ-113	6.60	3-CF ₃	H	Phenyl
KJ-114	5.82	4-CF ₃	H	Phenyl
KJ-115	6.83	3-OH	H	Phenyl
KJ-116	5.70	H	H	3,4-Dimethoxyphenyl
KJ-117	6.54	3-Br	H	3,4-Dimethoxyphenyl
KJ-118	6.38	3,4-OCH ₃	H	3,4-Dimethoxyphenyl
KJ-119	6.20	2-NO ₂	H	3,4-Dimethoxyphenyl
KJ-120	7.13	3-NO ₂	H	3,4-Dimethoxyphenyl
KJ-121	6.53	4-NO ₂	H	3,4-Dimethoxyphenyl
KJ-122	6.53	3-CF ₃	H	3,4-Dimethoxyphenyl
KJ-123	6.72	3-CN	H	3,4-Dimethoxyphenyl
KJ-124	6.60	2-OH	H	3,4-Dimethoxyphenyl
KJ-125	6.87	3-OH	H	3,4-Dimethoxyphenyl
KJ-126	6.33	4-OH	H	3,4-Dimethoxyphenyl
KJ-127	5.98	3-Br	H	H
KJ-128	5.93	3-CN	H	H
KJ-129	5.96	3-NO ₂	H	H
KJ-131	5.32	3,4-OCH ₃	H	H
KJ-132	5.76	3-Br	6,7-Dimethoxy	H
KJ-133	5.64	3-CF ₃	6,7-Dimethoxy	H
KJ-134	5.98	3-CN	6,7-Dimethoxy	H
KJ-135	5.42	2-NO ₂	6,7-Dimethoxy	H
KJ-136	6.09	3-NO ₂	6,7-Dimethoxy	H

Compound	pIC ₅₀	R ¹	Positions 6 and 7	R ²
KJ-137	5.54	4-NO ₂	6,7-Dimethoxy	H

6.2 Free-Wilson analysis

The Free-Wilson analysis directly correlates structural features of the molecules with the activity of the compounds [163]. In this work, the Free-Wilson analysis was used to characterize and quantify the contribution of different substituents in the structure of quinazolines to the BCRP inhibitory activity. Furthermore, the Free-Wilson analysis can be used to verify the quality of the biological data and find possible errors in the biological data.

The first Free-Wilson analysis was obtained using all the compounds of table 6.1. The used matrix is shown in table 6.2. The principal substitutions of these compounds involve the anilin substructure (R¹). This aromatic ring is substituted at position 2, 3, and 4 by several groups. Altogether, 17 features of the matrix used for the Free-Wilson analysis describe the variations at position R¹. The R² position, in the reference compound KJ-98 is occupied by a phenyl group. This group is either replaced by an hydrogen (code name: R² (H)) or by a 3,4-dimethoxyphenyl group (code name: R² (3,4-OCH₃-Ph)). For six compounds included in the analysis, the positions 6 and 7 are also substituted by two methoxy groups.

Table 6.2: The Free-Wilson matrix. Compound KJ-98 is used as reference compound.

Compound	R ¹																R ²				
	2-Br	2-OCH ₃	2-NO ₂	2-OH	3-Br	3-OCH ₃	3-NO ₂	3-OH	4-Br	4-OCH ₃	4-NO ₂	4-OH	3-Cl	4-F	3-CN	3-CF ₃	4-CF ₃	H	3,4-OCH ₃ -Ph	6,7-OCH ₃	
KJ-98																					
KJ-99	1																				
KJ-100					1																
KJ-101									1												
KJ-102														1							
KJ-103														1	1						
KJ-104		1																			
KJ-105						1															
KJ-106										1											
KJ-109							1														
KJ-110											1										
KJ-112																1					
KJ-113																	1				
KJ-114																		1			
KJ-115								1													
KJ-116																					1
KJ-117					1																1
KJ-118						1					1										1
KJ-119			1																		1
KJ-120							1														1
KJ-121											1										1
KJ-122																	1				1
KJ-123																1					1
KJ-124				1																	1
KJ-125								1													1
KJ-126												1									1
KJ-127					1														1		
KJ-128																1				1	
KJ-129							1														1
KJ-131						1					1										1
KJ-132					1															1	1
KJ-133																	1				1
KJ-134																1					1
KJ-135			1																	1	1
KJ-136							1													1	1
KJ-137											1									1	1

Table 6.3: Result of the Free-Wilson analysis. In the table the used features, the coefficients \pm SD and the p values are reported.

Features	Coefficients \pm SD	p values
Intercept	5.5480 \pm 0.0995	< 0.001
2-Br	-0.1220 \pm 0.1802	0.5088
2-OCH ₃	0.1008 \pm 0.1802	0.5843
2-NO ₂	0.5403 \pm 0.1537	< 0.001
2-OH	0.8923 \pm 0.1863	< 0.001
3-Br	0.9038 \pm 0.1299	< 0.001
3-OCH ₃	0.3674 \pm 0.1324	0.014
3-NO ₂	1.2931 \pm 0.1299	< 0.001
3-OH	1.2218 \pm 0.1448	< 0.001
4-Br	-0.3825 \pm 0.1801	0.051
4-OCH ₃	0.2014 \pm 0.1324	0.1491
4-NO ₂	0.6656 \pm 0.1342	< 0.001
4-OH	0.6223 \pm 0.1863	0.004
3-Cl	0.1721 \pm 0.1802	0.3547
4-F	0.0600 \pm 0.2124	0.7814
3-CN	1.1431 \pm 0.1299	< 0.001
3-CF ₃	0.8943 \pm 0.1342	< 0.001
4-CF ₃	0.2720 \pm 0.1802	0.1518
R ₂ (H)	-0.7277 \pm 0.0994	< 0.001
R ₂ (3,4-OCH ₃ -Ph)	0.1597 \pm 0.0728	0.0445
6,7-OCH ₃	0.0114 \pm 0.1054	0.9157

The QSAR model obtained by this first analysis has an r^2 value of 0.964 ($s = 0.150$, $F = 19.79$ for $n = 36$). The coefficients reported in table 6.3 highlight the negative effect of the substitution with bromine at position 2 and 4 of the aniline. The substitution of the phenyl group at position R² with an hydrogen decreases the inhibitory activity of these compounds: this effect is also described by a negative coefficient in the Free-Wilson analysis. The low coefficient associated with the substitutions at positions 6 and 7 evidences that the substitution with methoxy groups at these positions does not influence the activity of these compounds.

Table 6.4: The Free-Wilson matrix for 19 compounds. Compound KJ-98 is used as reference compound.

Compound	2-Br	2-OCH ₃	2-NO ₂	2-OH	3-Br	3-OCH ₃	3-NO ₂	3-OH	4-Br	4-OCH ₃	4-NO ₂	4-OH	3,4-OCH ₃ -Ph
KJ-98													
KJ-99	1												
KJ-100					1								
KJ-101									1				
KJ-104		1											
KJ-105						1							
KJ-106										1			
KJ-109							1						
KJ-110												1	
KJ-115								1					
KJ-116													1
KJ-117					1								1
KJ-118							1			1			1
KJ-119			1										1
KJ-120							1						1
KJ-121											1		1
KJ-124				1									1
KJ-125								1					1
KJ-126												1	1

Another Free-Wilson analysis was performed including only compounds chosen according to the presence or absence of the chosen substituents at all three positions of the aniline substructure. The 19 compounds with these substituents were included in the analysis. The chosen substituents are:

- Br, NO₂, OH and OCH₃ at positions ortho, meta and para to the aniline (Code names of the features: Br, OCH₃, NO₂, OH).
- OCH₃ at positions 3 and 4 of the aromatic ring at position R² (Code name of the feature: 3,4-OCH₃-Ph).

Table 6.5 reports the features used in the Free-Wilson analysis. Compound KJ-98, with hydrogen atoms at all positions, is used as reference. The final model consists of ten significant variables. The obtained QSAR model shows a very good fit ($r^2 = 0.988$, $s = 0.1177$, $F = 30.74$ for $n = 19$). Furthermore, the model is able to explain the contribution of the substituents to the activity of the compounds: as already observed also in the first Free-Wilson model, the only two negative coefficients are associated to the presence of bromine at position 2 and 4 of the aniline.

Table 6.5: Result of the Free-Wilson analysis. In the table the used features, the coefficients \pm SD and the p values are reported.

Features	Coefficients \pm SD	p values
Intercept	5.513 \pm 0.079	< 0.001
2-Br	-0.087 \pm 0.142	0.568
2-OCH ₃	0.136 \pm 0.142	0.382
2-NO ₂	0.439 \pm 0.148	0.032
2-OH	0.839 \pm 0.148	0.002
3-Br	0.754 \pm 0.114	0.001
3-OCH ₃	0.384 \pm 0.109	0.017
3-NO ₂	1.378 \pm 0.114	< 0.001
3-OH	1.213 \pm 0.114	< 0.001
4-Br	-0.347 \pm 0.114	0.058
4-OCH ₃	0.218 \pm 0.109	0.102
4-NO ₂	0.635 \pm 0.114	0.003
4-OH	0.569 \pm 0.148	0.012
3,4-OCH ₃ -Ph	0.248 \pm 0.066	0.013

This substitution decreases the BCRP-inhibitory activity of the compounds KJ-99 and KJ-101 with respect to the reference compound. The other coefficients, all positive, indicate that substitutions on the aniline lead to an increase of the BCRP-inhibitory activity. The higher coefficients are associated with OH and NO₂ groups, in particular at position 3 (meta) of the aniline. These numerical values are confirmed by

the high inhibitory activity of compounds KJ-115 ($\text{pIC}_{50} = 6.83$) and KJ-120 ($\text{pIC}_{50} = 7.13$).

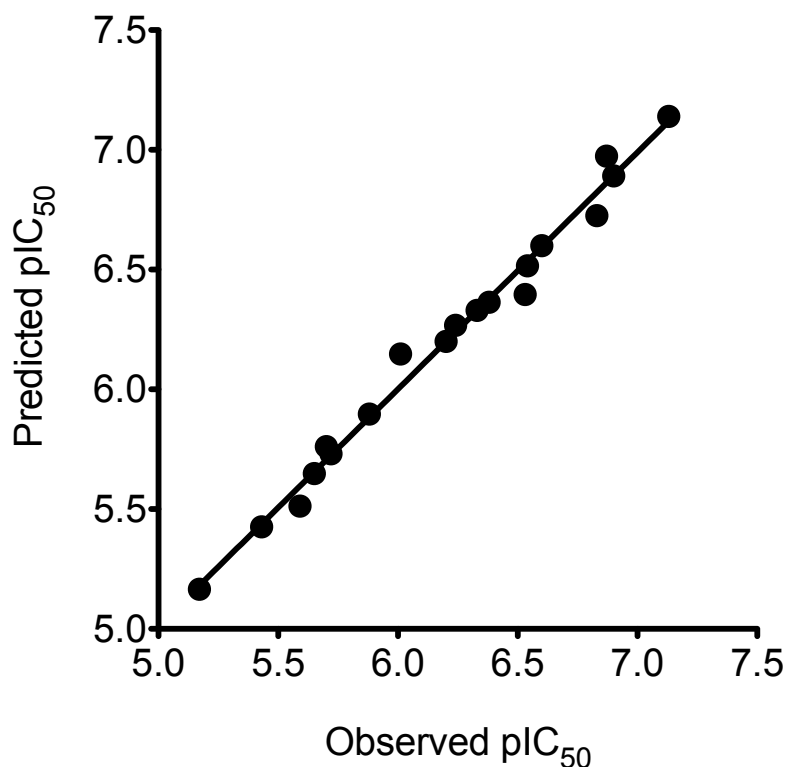


Figure 6.2: Correlation of observed versus predicted BCRP inhibition potencies using the Free-Wilson model ($n = 19$).

The presence of two methoxy groups on the aromatic ring at position R^2 does not influence the activity of this class of compounds. This observation is also confirmed by the low value of the coefficient associated with this feature (equal to 0.248). The predicted versus observed activity values calculated by the Free-Wilson model, using only 19 compounds are shown in figure 6.2.

6.3 Preparation of the dataset

The compounds presented in this work were built using the software MOE [97]. The general procedure for preparation of molecular structures used in 2D-QSAR requires the minimization of each molecule in the dataset. However, considering the low flexibility and high similarity of the compounds used in this work, it was decided to minimize only the most active compound of the dataset and to use it as template for the superposition of the other molecules. The molecular structure of compound KJ-120 was chosen to be minimized and then used as template. Firstly, a conformational search was performed in order to obtain a structure as close as possible to the absolute

energy minimum. In this first step, the charges were calculated using the MMFF94x force field [156]. A stochastic search was performed using an iteration limit value of 1000 and RMS gradient equal to 0.005. The rejection limit was fixed at a value of 100 and the RMSD limit to a value of 0.25. Finally, the structure with lower energy was chosen. At a later stage, new charges were assigned using the PM3 semi-empirical method implemented in the MOPAC 6 package [164]. Finally, the other molecular structures were superimposed on the template to obtain the final dataset.

6.4 Chemical descriptor based QSAR model

The biological activities (expressed as pIC₅₀) of the 36 compounds included in this analysis are uniformly distributed in a range between 5.17 and 7.13, as displayed in table 6.1. The range of about two logarithmic units is an important prerequisite to obtain a predictive model. The predictivity of the model was demonstrated by leave-one-out (LOO) cross validation.

The best model developed using 2D and 3D descriptors calculated with the software MOE uses only three descriptors: the fractional negative van der Waals surface area (Q_VSA_FNEG), the critical packing parameter (vsurf_CP) and the polar volume at -1.0 (vsurf_Wp3). The last two descriptors are VSURF descriptors. These descriptors were selected from the original pool of 2D and 3D included in the software MOE 2013.08. The descriptors used in the model were selected using the genetic algorithm based svl code "QuaSAR-Evolution" [165]. The multiple linear regression (MLR) analysis was performed using the software R [166], as well as the statistical analysis of the generated QSAR model. The model generated using MLR analysis is shown below:

$$\begin{aligned} \text{pIC}_{50} = & 4.902(\pm 0.407) + 2.148(\pm 0.465)Q_VSA_FNEG \\ & - 2.178(\pm 0.420)vsurf_CP + 0.028(\pm 0.003)vsurf_Wp3 \end{aligned} \quad (6.1)$$

$$n = 36, r^2 = 0.777, s = 0.254, F = 37.18, p < 0.00001, r_{cv}^2 = 0.712$$

The plot of the observed versus predicted pIC₅₀ values is reported in figure 6.3. The relative importance of the used descriptors is equal to 0.565 for the fractional negative van der Waals surface area, 0.633 for the critical packing parameter and 1.000 for the polar volume at -1.0. These three descriptors are not correlated with each other.

The results shown above indicate the high quality of the model and its high predictiv-

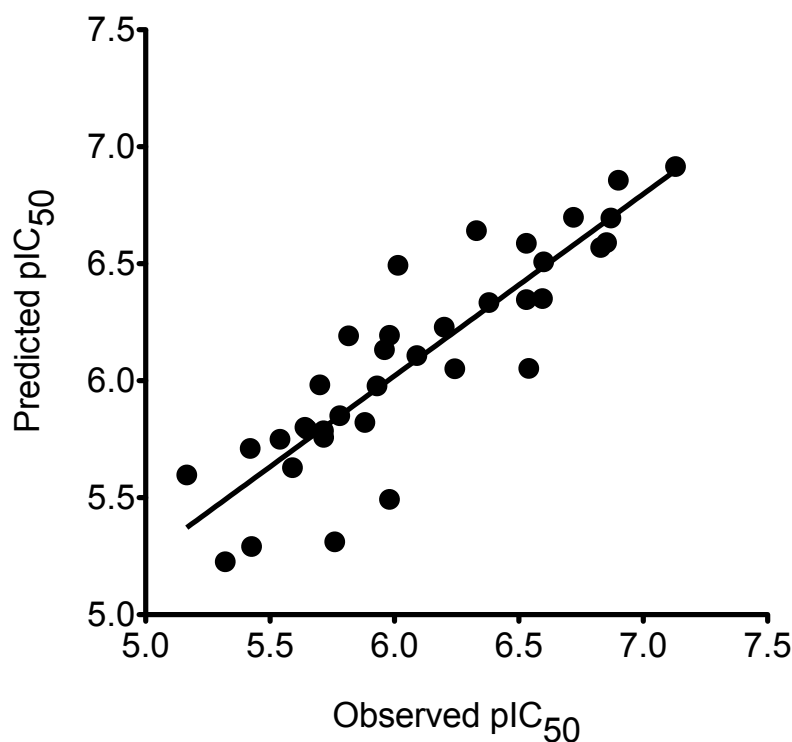


Figure 6.3: Correlation of observed versus predicted BCRP inhibition potencies using the multiparametric QSAR model.

ity. The positive sign of the coefficient for the Q_VSA_FNEG descriptor indicates that an increase of the molecular surface area characterized by negative charges increases the BCRP inhibitory activity of these compounds. The critical packing parameter is the ratio between the volumes of the hydrophobic regions and the surface of the hydrophilic regions times the length of the hydrophobic regions [98]. This parameter is correlated with the lipophilicity ($SLogP$, calculated using MOE) of the molecule. An increase of the critical packing parameter leads to a decrease of the BCRP inhibitory activity of these compounds. The positive sign of the coefficient for the polar volume suggests that the activity of the compounds is directly correlated with the polar area of the molecule, as already indicated by the first parameter of the equation (Q_VSA_FNEG).

6.5 3D-QSAR analysis

The molecular alignment is an essential step for CoMFA and CoMSIA studies. The minimized structure of the most active compound of this series, compound KJ-120, which was obtained by stochastic search and consecutive geometrical optimization, was used as a template for the molecular alignment. After the alignment of the remaining compounds, the charges were recalculated using the PM3 semi-empirical method, without a fresh optimization of the molecular geometry. The 3D-QSAR was performed

using the software SYBYL-X 1.2 (TRIPOS) [167].

For the CoMFA calculations, the standard settings were used: The grid size was extended 4 Å beyond every molecule in the three dimensions, with a spacing value of 2 Å. The probe atom was a sp^3 carbon probe with +1 charge. The steric (s), electrostatic (e), both (b), and hydrogen bond (h+b) CoMFA fields were calculated. The PM3 charges, calculated with MOPAC, were used for the calculation of the electrostatic fields. The steric and electrostatic cutoff values were both fixed to the standard value of 30.0 kcal/mol.

The steric (s), electrostatic (e), hydrophobic (h), hydrogen bond donor (d), and hydrogen bond acceptor (a) CoMSIA fields were also calculated. The grid box for the CoMSIA calculation was generated analogously to that of the CoMFA calculation. The probe atom with a radius of 1 Å, hydrophobicity equal to +1, and hydrogen bond donor-acceptor properties also equal to +1 were used.

The QSAR models for the CoMFA and CoMSIA analysis were calculated using the Partial Least Square (PLS) method. The internal predictivity of the models was evaluated using the Leave-One-Out (LOO) cross-validation. The quality of the non-cross-validated model was quantified by the squared correlation coefficient (r^2) and the standard error of estimation (s).

The produced CoMFA and CoMSIA models and the statistical parameters are summarized in table 6.6. The best models are reported in bold. The best values of squared correlation coefficient and cross-validated squared correlation coefficients (q^2) are obtained for the CoMSIA models. In CoMSIA analysis, the electrostatic field gives the best q^2 value (equal to 0.497, $n_{opt} = 5$) among the single field models. On the other hand, the best single field model obtained by CoMFA analysis uses the hydrogen bond field ($q^2 = 0.517$, $n_{opt} = 3$). The other fields alone are not able to generate predictive CoMSIA models. The combined CoMSIA models with highest q^2 values are a combination of hydrogen bond donor and electrostatic fields ($q^2 = 0.750$, $n_{opt} = 6$), steric, electrostatic and hydrogen bond donor fields ($q^2 = 0.731$, $n_{opt} = 6$), and hydrogen bond donor-acceptor, electrostatic and hydrophobic fields ($q^2 = 0.723$, $n_{opt} = 7$).

Due to the inability of the LOO-cross-validation to estimate the model capacity to predict the activity of completely new compounds [168], further validations of the generated models are needed. In the leave-many-out (LMO) cross-validation a number of compounds greater than one is left out for the validation. Different authors suggest that this cross-validation variant could give more realistic results with respect to the leave-one-out cross-validation [168][169]. The number of groups in which the dataset has to be divided depends on the number of compounds in the dataset. Gramatica [168] suggests that if the number of compounds in the dataset is lower than 50, that the number of groups must be greater than two, otherwise the predictivity of the generated model is underestimated. In this work 10-, 5-, and 3- folds cross-validation

Table 6.6: 3D-QSAR models using CoMFA and CoMSIA analysis and their statistical characteristics. The best models are reported in bold.

	q^2	s	n_{opt}	r^2	s	F
CoMFA						
Both	0.565	0.361	4	0.768	0.263	5.651
s	0.455	0.404	4	0.658	0.320	14.915
e	0.490	0.397	5	0.744	0.281	17.451
hb	0.517	0.374	3	0.703	0.293	25.220
s+e+hb	0.640	0.345	7	0.926	0.157	49.895
CoMSIA						
a	-0.023	0.528	1	0.137	0.485	5.391
d	0.129	0.487	1	0.194	0.469	8.163
d+a	0.122	0.504	3	0.405	0.415	7.264
e	0.497	0.394	5	0.733	0.287	16.437
e+a	0.691	0.320	7	0.915	0.167	43.287
e+d	0.750	0.283	6	0.912	0.167	50.362
e+h	0.480	0.408	6	0.731	0.293	13.113
e+h+a	0.719	0.305	7	0.931	0.151	54.213
e+h+d	0.703	0.313	7	0.914	0.169	42.496
e+h+d+a	0.723	0.303	7	0.940	0.141	62.509
h	0.461	0.422	7	0.728	0.300	10.699
h+a	0.596	0.366	7	0.845	0.227	21.743
h+d	0.390	0.450	7	0.751	0.287	12.082
h+d+a	0.591	0.368	7	0.841	0.229	21.149
s	0.460	0.408	5	0.628	0.339	10.133
s+a	0.513	0.376	3	0.629	0.328	18.055
s+d	0.462	0.422	7	0.726	0.301	10.575
s+e	0.512	0.388	5	0.751	0.278	18.068
s+e+a	0.685	0.323	7	0.901	0.181	36.479
s+e+d	0.731	0.293	6	0.906	0.174	46.320
s+e+h	0.476	0.403	5	0.721	0.293	15.528
s+e+h+a	0.715	0.307	7	0.928	0.154	51.741
s+e+h+a+d	0.724	0.302	7	0.936	0.145	58.768

of the best models obtained by LOO-cross-validation was performed, in order to verify their predictivity. In the case of the LMO-cross-validation the population of each group is randomly selected, therefore the q^2 values obtained are always different. For this reason, each LMO-cross-validation was repeated 10 times and the means of the obtained $q^2 \pm SD$ are reported in table 6.7. The e+d model, that shows the highest q^2 value with the LOO-cross-validation ($q^2 = 0.750$), shows a decrease of predictivity already for the 10-folds cross-validation ($q^2 = 0.694$). A reduction of the predictivity is also observed for the e+h+d+a model (from $q^2 = 0.723$ to $q^2 = 0.658$). A slight decrease of the q^2 value is observed also for the s+e+d model.

Table 6.7: The results of the leave-many-out (LMO) cross-validation for the best 3D-QSAR models. The average q^2 and the SD are based on 10 repetitions of the LMO protocol.

Number of groups	CoMSIA fields		
	e+d $q^2 \pm SD$	s+e+d $q^2 \pm SD$	e+h+d+a $q^2 \pm SD$
LOO	0.750	0.731	0.723
10	0.694 \pm 0.069	0.725 \pm 0.016	0.658 \pm 0.121
5	0.674 \pm 0.069	0.715 \pm 0.053	0.645 \pm 0.150
3	0.648 \pm 0.087	0.624 \pm 0.062	0.485 \pm 0.199

A further validation method, called Y-randomization test, was applied for the best models. It consists of randomly shuffling the activity values of the compounds and then calculating new q^2 values, obtained by LOO cross-validation [170]. The aim of this test is to show if a correlation between the activities and the descriptors really exists, or if the q^2 values are obtained by chance. The results of this test are reported in table 6.8. It is seen that all the best generated models have mean q^2 values lower than 0, suggesting that the predictivities obtained by LOO cross-validation are not obtained by chance.

Table 6.8: The results of the Y-scrambling analysis. The q^2 values and the SD values are obtained from 10 repetitions of the analysis.

CoMSIA fields	$q^2 \pm SD$
e+d	-0.244 \pm 0.133
s+e+d	-0.220 \pm 0.134
e+h+d+a	-0.198 \pm 0.163

In the light of these results the e+d and the "s+e+d" models were chosen to explain the activity of this class of compounds. Finally, figure 6.4 shows the correlations between observed and predicted pIC_{50} models for the studied compounds. The predicted

6 Structure-activity relationships of quinazolines derivatives as inhibitors of BCRP.

activity values are obtained using the "e+d" CoMSIA model ($r^2 = 0.912$) and the "s+e+d" CoMSIA model ($r^2 = 0.906$). The plots show that for all the compounds the predicted values are well correlated with the observed values.

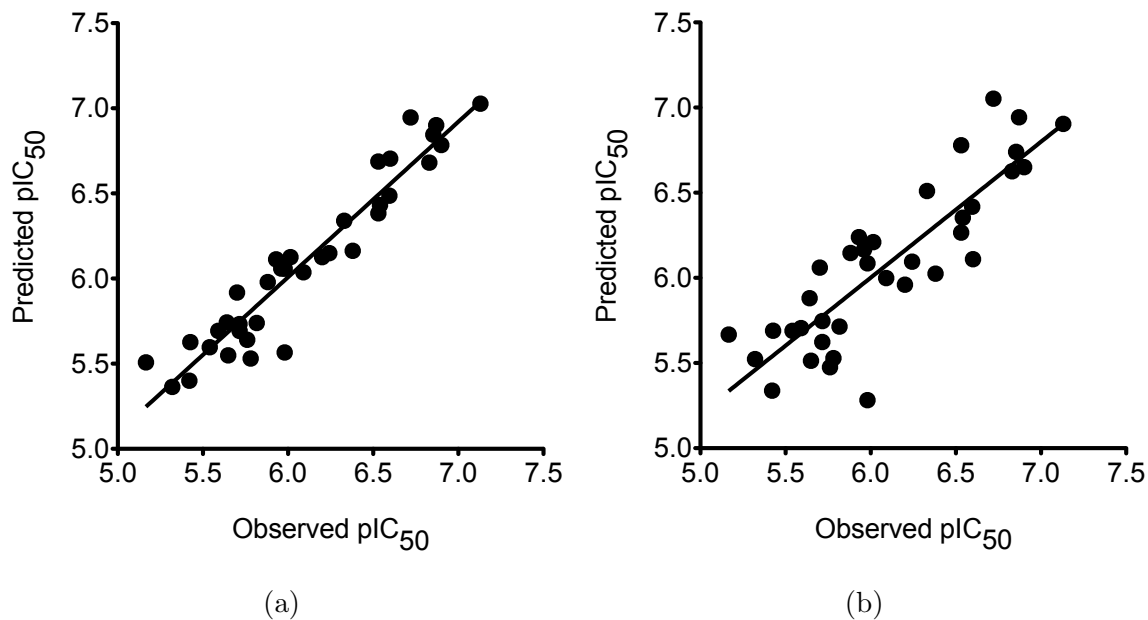


Figure 6.4: Correlation of observed versus predicted BCRP inhibition potencies using the 3D-QSAR model based on (a) the "e+d" fields and on (b) the "s+e+d" fields.

6.6 Contour plot

The contour (STDEV*COEFF) plots of the "e+d" CoMSIA model and of the "s+e+d" CoMSIA model are illustrated in figure 6.5. In particular, the figures 6.5a and 6.5b show the contour (STDEV*COEFF) plots of the "e+d" CoMSIA model drawn together with the most active compound of the series, compounds KJ-120 ($pIC_{50} = 7.13$) and with the low active compound KJ-131 ($pIC_{50} = 5.32$), respectively. The figures 6.5c and 6.5d show the contour (STDEV*COEFF) plots of the "s+e+d" CoMSIA model drawn together with the compounds KJ-120 and KJ-131, respectively. These plots describe the contribution of the fields to the compounds activity. The green contour areas mark the favorable influences of the steric field, while the yellow areas are associated with regions where the steric field is unfavorable. It can be seen in figure 6.5c that this contour area is occupied by a phenyl rest of compound KJ-120, while this phenyl residue is absent in the structure of compound KJ-131. An unfavorable area for the steric field is also reported at position 4 of the aniline. This position is not occupied by any substituent in compound KJ-120, while compound KJ-131 has a methoxy group at the same position. Regarding the electrostatic field, blue contour areas indicate favorable electrostatic interactions, while red contour areas unfavorable electrostatic interactions.

The contour plot indicates that generally, the presence of polar substituents increases the activity of the compounds if they are localized at position 4 of the aniline, while these substituents slightly decrease the inhibitory activity if localized at position 3 of the same aromatic ring. The favorable influence of hydrogen bond donor groups is indicated by turquoise areas and the unfavorable influence by purple areas. The most important contribution of this field, as it is possible to observe in the contour plot, is given by substituents at position 3 of the aniline: The presence of hydrogen-bond donor groups at this position leads to an increase of the inhibitory activity of the compounds. Two examples are represented by the hydroxy group of compounds KJ-115 ($pIC_{50} = 6.83$) and KJ-125 ($pIC_{50} = 6.87$).

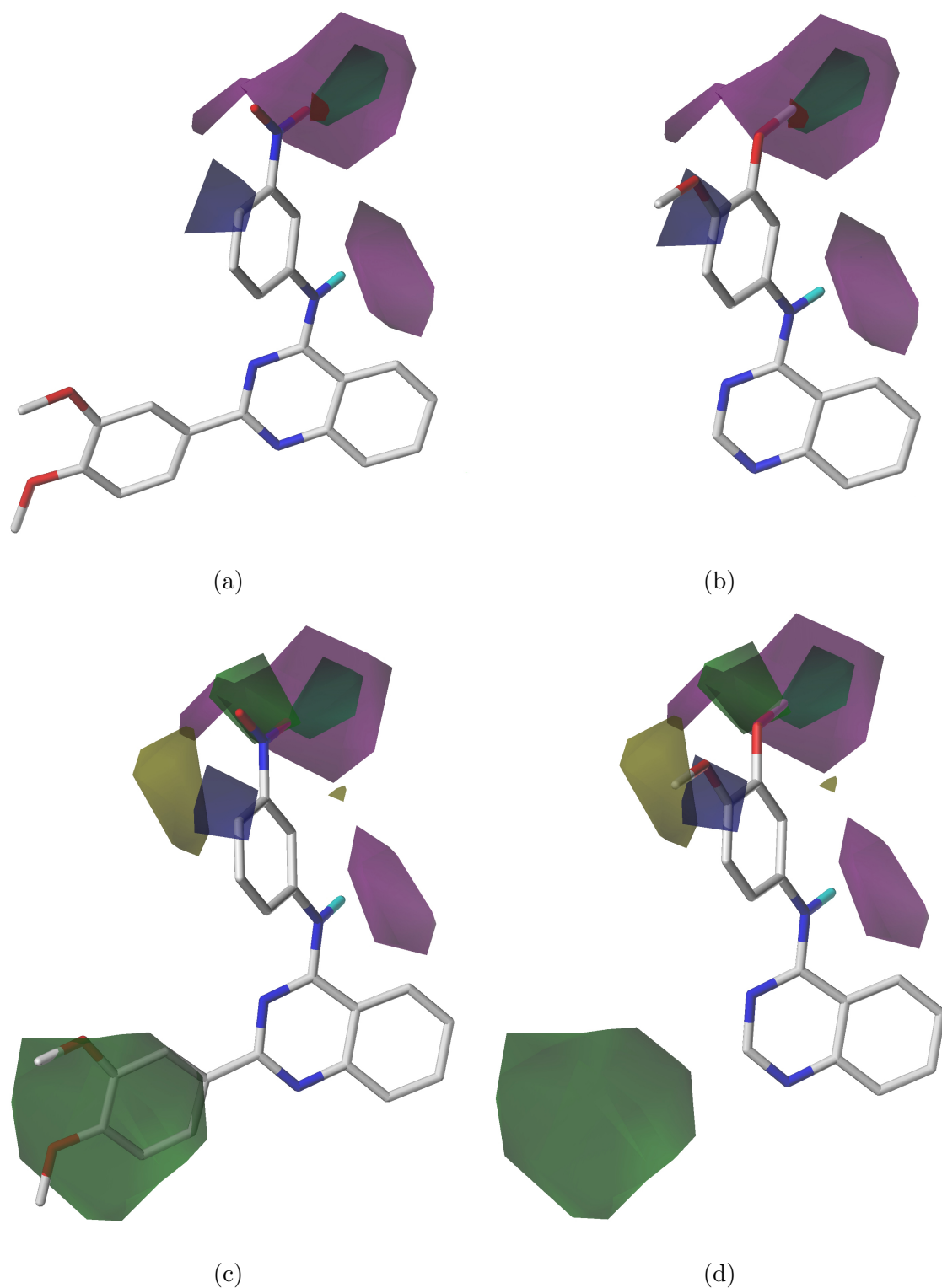


Figure 6.5: CoMSIA contour plot (STDEV*COEFF) of the 3D-QSAR model generated using (a) the "e+d" fields and the most active compound KJ-120, (b) the "e+d" fields and the low active compound KJ-131, (c) the "s+e+d" fields and the most active compound KJ-120, and (d) the "s+e+d" fields and the low active compound KJ-131. The steric bulk is favored at green (60 %) and disfavored at yellow (10 %) regions. Polar groups are favored at blue (90 %) and disfavored at red (15 %). H-bond donor groups are favored at turquoise (75 %) and disfavored at purple (45 %) regions.

7 Conclusions

The first aim of this work was the characterization of a new class of BCRP inhibitors, structurally derived from tariquidar. Furthermore, the Hoechst 33342-assay was optimized and validated and then used to determine the inhibitory activity values (IC_{50}) of the studied compounds. The optimization of the assay and the subsequent biological tests were performed using MCF-7 MX cells. Considering WK-X-27 (KS-176) as reference and starting point, several compounds with variations of the two linkers and of the three aromatic rings were synthesized by Dr. Steggemann.

The results of the biological tests confirmed that the deletion of the tetrahydroisoquinoline moiety, originally included in the tariquidar structure, generally leads to loss of inhibitory activity against P-gp, while the inhibitory activity against BCRP is not compromised. The relative position of the two linkers in the central aromatic ring was also found to be essential for the inhibitory activity of these compounds: Indeed, only the *ortho* substituted compounds are BCRP inhibitors. The modifications of the first linker leads to active and selective compounds only in case of KS-311 ($IC_{50} = 0.644 \mu\text{M}$), where the length of the linker is increased by one single carbon atom. Also variations of the second linker generally lead to worsening of the inhibitory BCRP activity except in case of compound KS-246, with an IC_{50} value of $0.52 \mu\text{M}$.

The variations on the third aromatic ring can be explained using 2D-QSAR. The obtained correlation shows the importance of the σ values of the substituents and their interaction surface area for the BCRP inhibitory activity of these compounds. The most active compound with a modification of the third aromatic ring is KS-407 ($IC_{50} = 1.16 \mu\text{M}$) and its activity is well predicted by the 2D-QSAR model. The substitution of the second aromatic ring generally leads to worsening of BCRP inhibitory activity. This effect is stronger if the substituent is at position R^1 of the ring, it is weaker if the substituent is at position R^2 . The inhibitory activity also decreases with increasing van der Waals volume of the substituent. The substitution of the hydroxyethyl group in the first aromatic ring with an hydroxy group directly connected to the ring (KS-251) leads to a potent (IC_{50} BCRP = $1.76 \mu\text{M}$) and selective (IC_{50} P-gp = $27.5 \mu\text{M}$) BCRP inhibitor, suggesting that the geometrical dimension of the pharmacophore of this compounds class can be reduced.

Several members of this new class of BCRP inhibitors, like the potent and selective BCRP inhibitors KS-407 and KS-251 were shown to be able to reverse the resistance against several cytotoxic compounds like Hoechst 3342, mitoxantrone, and SN-38 in

BCRP overexpressing cells.

Three classification algorithms, the Self-Organizing Maps (SOM), the Support Vector Machine (SVM) and the k-Nearest Neighbor (k-NN) were used to understand the reason why several compounds do not show any inhibitory activity against BCRP. The best results were obtained using the 3D RDF and gATS descriptors and building the model using the SOM algorithm. These molecular descriptors codify properties like the molecular surface or the molecular dimension. It was also shown that the lipophilicity does not play a major role to discriminate between active or inactive BCRP inhibitors. Using the combination of the SOM algorithm and 3D RDF or gATS descriptors, it is possible to discriminate between active and inactive BCRP inhibitors with an accuracy of 0.79 using either the descriptors. The 3D RDF and gATS descriptors were also successfully used in combination with the SVM and k-NN algorithms. The SVM algorithm was able to discriminate between the two classes with an accuracy of 0.75 for the 3D RDF descriptors and of 0.77 for the gATS descriptors. The use of the k-NN algorithm leads to lower accuracy values: an accuracy of 0.74 was obtained using the 3D-RDF descriptors and of 0.73 using the gATS descriptors. The trained model was also used to predict the activity of an external test set, consisting of 16 compounds not structurally similar with the compounds included in the training set. The results show that the classified activity data of the test set was predicted with an accuracy of 0.88 using the SOM algorithm and 3D RDF descriptors. This result suggests that this model can be used to predict classified activity values of newly designed BCRP inhibitor candidates.

In the last project the 2D- and 3D- quantitative structure activity relationship (QSAR) models of 36 quinazoline derivatives were generated. The Free-Wilson model highlighted the importance of hydroxy and nitro groups at position *meta* of the aniline for the activity of these compounds. The chemical descriptor based QSAR model, that was built using only three molecular descriptors has a r^2 value of 0.777 and suggests that the inhibitory activity is correlated with the distribution of the partial charges on the molecular surface and with the lipophilicity of the molecules. Furthermore, CoMFA and CoMSIA models were also generated. The best models are the CoMSIA models that use the combination of electrostatic and hydrogen bond donor CoMSIA fields (e+d model, $q^2 = 0.750$), steric, electrostatic and hydrogen bond donor CoMSIA fields (s+e+d model, $q^2 = 0.713$) and electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor CoMSIA fields (e+h+d+a model, $q^2 = 0.723$). The generated models are stable also for 10- 5- and 3- folds cross-validation, in particular the e+d and s+e+d models. Furthermore, the Y-randomization test demonstrated that the obtained q^2 values of these models are not obtained by chance. In conclusion, these 2D- and the 3D- QSAR models are highly predictive, able to explain the inhibitory activity of the studied compounds, and can be used for later design of new

BCRP inhibitors.

Bibliography

- [1] Jemal, A.; Bray, F.; Center, M. Global cancer statistics. *CA: a cancer journal ...* **2011**, *61*, 69–90.
- [2] Umar, A.; Dunn, B. K.; Greenwald, P. Future directions in cancer prevention. *Nature reviews. Cancer* **2012**, *12*, 835–848.
- [3] Fuster, V.; Kelly, B. *Promoting Cardiovascular Health in the Developing World: A Critical Challenge to Achieve Global Health*; 2010.
- [4] Boumendjel, A.; Boutonnat, J. ABC Transporters and Multidrug Resistance **2009**.
- [5] Xia, C. Q.; Smith, P. G. Drug efflux transporters and multidrug resistance in acute leukemia: therapeutic impact and novel approaches to mediation. *Molecular pharmacology* **2012**, *82*, 1008–21.
- [6] Leonard, G.; Leonard GD,; Fojo T,; Bates SE, The Role of ABC Transporters in Clinical Practice. *The Oncologist* **2003**, *8*, 411–424.
- [7] Sarkadi, B.; Homolya, L.; Szakács, G.; Váradi, A. Human multidrug resistance ABCB and ABCG transporters: participation in a chemoinnity defense system. *Physiological reviews* **2006**, *86*, 1179–236.
- [8] Dean, M.; Hamon, Y.; Chimini, G. The human ATP-binding cassette (ABC) transporter superfamily. *Journal of lipid research* **2001**, *42*, 1007–17.
- [9] Gottesman, M. M.; Ambudkar, S. V. Overview: ABC transporters and human disease. *Journal of bioenergetics and biomembranes* **2001**, *33*, 453–8.
- [10] Chang, X.-b. A molecular understanding of ATP-dependent solute transport by multidrug resistance-associated protein MRP1. *Cancer metastasis reviews* **2007**, *26*, 15–37.
- [11] Vlaming, M. L. H.; Lagas, J. S.; Schinkel, A. H. Physiological and pharmacological roles of ABCG2 (BCRP): recent findings in Abcg2 knockout mice. *Advanced drug delivery reviews* **2009**, *61*, 14–25.

Bibliography

- [12] Chen, Z.-S.; Tiwari, A. K. Multidrug resistance proteins (MRPs/ABCCs) in cancer chemotherapy and genetic diseases. *The FEBS journal* **2011**, *278*, 3226–45.
- [13] Solbach, T. F.; Paulus, B.; Weyand, M.; Eschenhagen, T.; Zolk, O.; Fromm, M. F. ATP-binding cassette transporters in human heart failure. *Naunyn-Schmiedeberg's archives of pharmacology* **2008**, *377*, 231–43.
- [14] Wu, C.-P.; Hsieh, C.-H.; Wu, Y.-S. The emergence of drug transporter-mediated multidrug resistance to cancer chemotherapy. *Molecular pharmaceutics* **2011**, *8*, 1996–2011.
- [15] Tusnády, G. E.; Sarkadi, B.; Simon, I.; Váradi, A. Membrane topology of human ABC proteins. *FEBS letters* **2006**, *580*, 1017–22.
- [16] Seeger, M. a.; van Veen, H. W. Molecular basis of multidrug transport by ABC transporters. *Biochimica et biophysica acta* **2009**, *1794*, 725–37.
- [17] Velamakanni, S.; Wei, S. L.; Janvilisri, T.; van Veen, H. W. ABCG transporters: structure, substrate specificities and physiological roles : a brief overview. *Journal of bioenergetics and biomembranes* **2007**, *39*, 465–71.
- [18] Hollenstein, K.; Dawson, R. J. P.; Locher, K. P. Structure and mechanism of ABC transporter proteins. *Current opinion in structural biology* **2007**, *17*, 412–8.
- [19] Dawson, R. J. P.; Locher, K. P. Structure of the multidrug ABC transporter Sav1866 from *Staphylococcus aureus* in complex with AMP-PNP. *FEBS letters* **2007**, *581*, 935–8.
- [20] Kage, K.; Fujita, T.; Sugimoto, Y. Role of Cys-603 in dimer/oligomer formation of the breast cancer resistance protein BCRP/ABCG2. *Cancer science* **2005**, *96*, 866–72.
- [21] Wang, H.; Lee, E.-W.; Cai, X.; Ni, Z.; Zhou, L.; Mao, Q. Membrane topology of the human breast cancer resistance protein (BCRP/ABCG2) determined by epitope insertion and immunofluorescence. *Biochemistry* **2008**, *47*, 13778–87.
- [22] Juliano, R.; Ling, V. A surface glycoprotein modulating drug permeability in chinese hamster ovary cell mutants. *Biochimica et biophysica acta* **1976**, *455*, 152–162.
- [23] Ling, V.; Thompson, L. H. Reduced permeability in CHO cells as a mechanism of resistance to colchicine. *Journal of cellular physiology* **1974**, *83*, 103–16.

- [24] Szakács, G.; Váradi, A.; Ozvegy-Laczka, C.; Sarkadi, B. The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). *Drug discovery today* **2008**, *13*, 379–93.
- [25] Zhou, S.-F. *Xenobiotica; the fate of foreign compounds in biological systems*; 2008; Vol. 38, pp 802–32.
- [26] Rosenberg, M. F.; Mao, Q.; Holzenburg, a.; Ford, R. C.; Deeley, R. G.; Cole, S. P. The structure of the multidrug resistance protein 1 (MRP1/ABCC1). crystallization and single-particle analysis. *The Journal of biological chemistry* **2001**, *276*, 16076–82.
- [27] Pajeva, I. K.; Globisch, C.; Wiese, M. Structure-function relationships of multidrug resistance P-glycoprotein. *Journal of medicinal chemistry* **2004**, *47*, 2523–33.
- [28] Globisch, C.; Pajeva, I. K.; Wiese, M. Identification of putative binding sites of P-glycoprotein based on its homology model. *ChemMedChem* **2008**, *3*, 280–95.
- [29] Aller, S. G.; Yu, J.; Ward, A.; Weng, Y.; Chittaboina, S.; Zhuo, R.; Harrell, P. M.; Trinh, Y. T.; Zhang, Q.; Urbatsch, I. L.; Chang, G. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science (New York, N.Y.)* **2009**, *323*, 1718–22.
- [30] Jin, M. S.; Oldham, M. L.; Zhang, Q.; Chen, J. Crystal structure of the multidrug transporter P-glycoprotein from *Caenorhabditis elegans*. *Nature* **2012**, *490*, 566–569.
- [31] Li, J.; Jaimes, K. F.; Aller, S. G. Refined structures of mouse P-glycoprotein. *Protein Science* **2014**, *23*, 34–46.
- [32] Shapiro, a. B.; Ling, V. Positively cooperative sites for drug transport by P-glycoprotein with distinct drug specificities. *European journal of biochemistry / FEBS* **1997**, *250*, 130–7.
- [33] Holland, I.; Cole, S.; Kuchler, K.; Higgins, C. *ABC Proteins: From Bacteria to Man*; 2003.
- [34] Raviv, Y.; Pollard, H.; Bruggemann, E. Photosensitized labeling of a functional multidrug transporter in living drug-resistant tumor cells. *Y. Journal of Biological ...* **1990**, *265*, 3975–3980.
- [35] Higgins, C. F.; Gottesman, M. M. Is the multidrug transporter a flippase? *Trends in biochemical sciences* **1992**, *17*, 18–21.

Bibliography

- [36] Altenberg, G. a.; Vanoye, C. G.; Horton, J. K.; Reuss, L. Unidirectional fluxes of rhodamine 123 in multidrug-resistant cells: evidence against direct drug extrusion from the plasma membrane. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 4654–7.
- [37] Hill, C. R.; Jamieson, D.; Thomas, H. D.; Brown, C. D. a.; Boddy, A. V.; Veal, G. J. Characterisation of the roles of ABCB1, ABCC1, ABCC2 and ABCG2 in the transport and pharmacokinetics of actinomycin D in vitro and in vivo. *Biochemical Pharmacology* **2013**, *85*, 29–37.
- [38] Seelig, A. A general pattern for substrate recognition by P-glycoprotein. *European journal of biochemistry / FEBS* **1998**, *251*, 252–61.
- [39] Cianchetta, G.; Singleton, R. W.; Zhang, M.; Wildgoose, M.; Giesing, D.; Fravolini, A.; Cruciani, G.; Vaz, R. J. A pharmacophore hypothesis for P-glycoprotein substrate recognition using GRIND-based 3D-QSAR. *Journal of medicinal chemistry* **2005**, *48*, 2927–35.
- [40] Huang, J.; Ma, G.; Muhammad, I.; Cheng, Y. Identifying P-glycoprotein substrates using a support vector machine optimized by a particle swarm. *Journal of chemical information and modeling* **2007**, *47*, 1638–47.
- [41] Wessler, J. D.; Grip, L. T.; Mendell, J.; Giugliano, R. P. The P-glycoprotein transport system and cardiovascular drugs. *Journal of the American College of Cardiology* **2013**, *61*, 2495–502.
- [42] Sharom, F. J. The P-glycoprotein multidrug transporter. *Essays in biochemistry* **2011**, *50*, 161–78.
- [43] Amin, M. L. P-glycoprotein Inhibition for Optimal Drug Delivery. *Drug target insights* **2013**, *7*, 27–34.
- [44] Leslie, E. M.; Deeley, R. G.; Cole, S. P. C. Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense. *Toxicology and applied pharmacology* **2005**, *204*, 216–37.
- [45] Fletcher, J. I.; Haber, M.; Henderson, M. J.; Norris, M. D. ABC transporters in cancer: more than just drug efflux pumps. *Nature reviews. Cancer* **2010**, *10*, 147–56.
- [46] Slovak, M.; Ho, J.; Bhardwaj, G.; Kurz, E. Localization of a Novel Multidrug Resistance-associated Gene in the HT1080 / DR4 and H69AR Human Tumor Cell Lines. *Cancer research* **1993**, 3221–3225.

- [47] Hipfner, D. R. Membrane Topology of the Multidrug Resistance Protein (MRP). A STUDY OF GLYCOSYLATION-SITE MUTANTS REVEALS AN EXTRA-CYTOSOLIC NH₂ TERMINUS. *Journal of Biological Chemistry* **1997**, *272*, 23623–23630.
- [48] Munoz, M.; Henderson, M.; Haber, M.; Norris, M. Role of the MRP1/ABCC1 multidrug transporter protein in cancer. *IUBMB life* **2007**, *59*, 752–7.
- [49] Wong, I. L. K.; Chan, K.-F.; Tsang, K. H.; Lam, C. Y.; Zhao, Y.; Chan, T. H.; Chow, L. M. C. Modulation of multidrug resistance protein 1 (MRP1/ABCC1)-mediated multidrug resistance by bivalent apigenin homodimers and their derivatives. *Journal of medicinal chemistry* **2009**, *52*, 5311–22.
- [50] Gekeler, V.; Ise, W.; Sanders, K. The Leukotriene LTD₄ Receptor Antagonist MK571 Specifically Modulates MRP associated multidrug resistance. . . . *and biophysical research . . .* **1995**.
- [51] Nakano, R.; Oka, M.; Nakamura, T.; Fukuda, M.; Kawabata, S.; Terashi, K.; Tsukamoto, K.; Noguchi, Y.; Soda, H.; Kohno, S. A leukotriene receptor antagonist, ONO-1078, modulates drug sensitivity and leukotriene C₄ efflux in lung cancer cells expressing multidrug resistance protein. *Biochemical and biophysical research communications* **1998**, *251*, 307–12.
- [52] Häcker, H.-G.; Leyers, S.; Wiendlocha, J.; Gütschow, M.; Wiese, M. Aromatic 2-(thio)ureidocarboxylic acids as a new family of modulators of multidrug resistance-associated protein 1: synthesis, biological evaluation, and structure-activity relationships. *Journal of medicinal chemistry* **2009**, *52*, 4586–4595.
- [53] Pajeva, I. K.; Globisch, C.; Wiese, M. Combined pharmacophore modeling, docking, and 3D QSAR studies of ABCB1 and ABCC1 transporter inhibitors. *ChemMedChem* **2009**, *4*, 1883–96.
- [54] Doyle, L. a.; Yang, W.; Abruzzo, L. V.; Krogmann, T.; Gao, Y.; Rishi, a. K.; Ross, D. D. A multidrug resistance transporter from human MCF-7 breast cancer cells. *Proceedings of the National Academy of Sciences* **1998**, *95*, 15665–15670.
- [55] Doyle, L. A.; Ross, D. D. Multidrug resistance mediated by the breast cancer resistance protein BCRP (ABCG2). *Oncogene* **2003**, *22*, 7340–58.
- [56] Kage, K.; Tsukahara, S. Dominant-negative inhibition of breast cancer resistance protein as drug efflux pump through the inhibition of S-S dependent homodimerization. . . . *journal of cancer* **2002**, *630*, 626–630.

Bibliography

- [57] Wakabayashi, K.; Nakagawa, H.; Tamura, A.; Koshihara, S.; Hoshijima, K.; Komada, M.; Ishikawa, T. Intramolecular disulfide bond is a critical check point determining degradative fates of ATP-binding cassette (ABC) transporter ABCG2 protein. *The Journal of biological chemistry* **2007**, *282*, 27841–6.
- [58] Allen, J. D.; Schinkel, A. H. Multidrug resistance and pharmacological protection mediated by the breast cancer resistance protein (BCRP/ABCG2). *Molecular Cancer Therapeutics* **2002**, *1*, 427–434.
- [59] McDevitt, C. a.; Collins, R. F.; Conway, M.; Modok, S.; Storm, J.; Kerr, I. D.; Ford, R. C.; Callaghan, R. Purification and 3D Structural Analysis of Oligomeric Human Multidrug Transporter ABCG2. *Structure* **2006**, *14*, 1623–1632.
- [60] McDevitt, C. a.; Collins, R.; Kerr, I. D.; Callaghan, R. Purification and structural analyses of ABCG2. *Advanced drug delivery reviews* **2009**, *61*, 57–65.
- [61] Staud, F.; Pavsek, P. Breast cancer resistance protein (BCRP/ABCG2). *The International Journal of Biochemistry & Cell Biology* **2005**, *37*, 720–725.
- [62] Robey, R. W.; To, K. K. K.; Polgar, O.; Dohse, M.; Fetsch, P.; Dean, M.; Bates, S. E. ABCG2: a perspective. *Advanced drug delivery reviews* **2009**, *61*, 3–13.
- [63] Krishnamurthy, P.; Ross, D. D.; Nakanishi, T.; Bailey-Dell, K.; Zhou, S.; Mercer, K. E.; Sarkadi, B.; Sorrentino, B. P.; Schuetz, J. D. The stem cell marker Bcrp/ABCG2 enhances hypoxic cell survival through interactions with heme. *The Journal of biological chemistry* **2004**, *279*, 24218–25.
- [64] Ishikawa, T. The role of human ABC transporter ABCG2 (BCRP) in pharmacotherapy. *Advanced drug delivery reviews* **2009**, *61*, 1–2.
- [65] Mao, Q.; Unadkat, J. D. Role of the breast cancer resistance protein (ABCG2) in drug transport. *The AAPS journal* **2005**, *7*, E118–33.
- [66] Rabindran, S.; Ross, D.; Doyle, L. Fumitremorgin C Reverses Multidrug Resistance in Cells Transfected with the Breast Cancer Resistance Protein. *Cancer research* **2000**, 47–50.
- [67] Robey, R.; Medina-Pérez, W. Overexpression of the ATP-binding Cassette Half-Transporter , ABCG2 (MXR / BCRP / ABCP1), in Flavopiridol-resistant Human Breast Cancer Cells. *Clinical Cancer ...* **2001**, *7*, 145–152.
- [68] Honjo, Y.; Hrycyna, C. A. C.; Yan, Q.-w. Q.; Medina-pe, W. Y.; Robey, R. W.; Laar, A. V. D.; Litman, T.; Dean, M.; Bates, S. E. Acquired Mutations in the

- MXR/BCRP/ABCP Gene Alter Substrate Specificity in MXR/BCRP/ABCP-overexpressing Cells. *Cancer research* **2001**, 6635– 6639.
- [69] Chen, Z.-S.; Robey, R. W.; Belinsky, M. G.; Shchaveleva, I.; Ren, X.-Q.; Sugimoto, Y.; Ross, D. D.; Bates, S. E.; Kruh, G. D. Transport of methotrexate, methotrexate polyglutamates, and 17beta-estradiol 17-(beta-D-glucuronide) by ABCG2: effects of acquired mutations at R482 on methotrexate transport. *Cancer research* **2003**, *63*, 4048–54.
- [70] Robey, R. W.; Honjo, Y.; van de Laar, a.; Miyake, K.; Regis, J. T.; Litman, T.; Bates, S. E. A functional assay for detection of the mitoxantrone resistance protein, MXR (ABCG2). *Biochimica et biophysica acta* **2001**, *1512*, 171–82.
- [71] Bachmeier, C. J.; Trickler, W. J.; Miller, D. W. Drug Efflux Transport Properties of 20,70-Bis(2-carboxyethyl)- 5(6)-carboxyfluorescein Acetoxymethyl Ester (BCECF-AM) and Its Fluorescent Free Acid, BCECF **2004**, *93*, 932–942.
- [72] Robey, R. W. Pheophorbide a Is a Specific Probe for ABCG2 Function and Inhibition. *Cancer Research* **2004**, *64*, 1242–1246.
- [73] Allen, J. D.; van Loevezijn, A.; Lakhai, J. M.; van der Valk, M.; van Tellingen, O.; Reid, G.; Schellens, J. H. M.; Koomen, G.-J.; Schinkel, A. H. Potent and specific inhibition of the breast cancer resistance protein multidrug transporter in vitro and in mouse intestine by a novel analogue of fumitremorgin C. *Molecular cancer therapeutics* **2002**, *1*, 417–25.
- [74] Ahmed-Belkacem, A.; Pozza, A.; Macalou, S.; Pe??rez-Victoria, J. M.; Boumendjel, A.; Di Pietro, A. Inhibitors of cancer cell multidrug resistance mediated by breast cancer resistance protein (BCRP/ABCG2). *Anti-Cancer Drugs* **2006**, *17*, 239–243.
- [75] Woehlecke, H.; Osada, H.; Herrmann, A.; Lage, H. Reversal of breast cancer resistance protein-mediated drug resistance by tryprostatin A. *International journal of cancer. Journal international du cancer* **2003**, *107*, 721–8.
- [76] Shiozawa, K.; Oka, M.; Soda, H.; Yoshikawa, M.; Ikegami, Y.; Tsurutani, J.; Nakatomi, K.; Nakamura, Y.; Doi, S.; Kitazaki, T.; Mizuta, Y.; Murase, K.; Yoshida, H.; Ross, D. D.; Kohno, S. Reversal of breast cancer resistance protein (BCRP/ABCG2)-mediated drug resistance by novobiocin, a coumermycin antibiotic. *International journal of cancer. Journal international du cancer* **2004**, *108*, 146–51.

Bibliography

- [77] Bruin, M. D.; Miyake, K.; Litman, T.; Robey, R.; Bates, S. Reversal of resistance by GF120918 in cell lines expressing the ABC half-transporter MXR. *Cancer letters* **1999**, *146*, 117–126.
- [78] Modok, S.; Mellor, H. R.; Callaghan, R. Modulation of multidrug resistance efflux pump activity to overcome chemoresistance in cancer. *Current opinion in pharmacology* **2006**, *6*, 350–4.
- [79] Minderman, H.; O’Loughlin, K. VX-710 (Biricodar) Increases Drug Retention and Enhances Chemosensitivity in Resistant Cells Overexpressing P-Glycoprotein , Multidrug Resistance Protein , and Breast Cancer Resistance Protein. *Clinical Cancer Research* **2004**, *10*, 1826–1834.
- [80] Takara, K.; Matsubara, M.; Yamamoto, K.; Minegaki, T.; Takegami, S.; Takahashi, M.; Yokoyama, T.; Okumura, K. Differential effects of calcium antagonists on ABCG2/BCRP-mediated drug resistance and transport in SN-38-resistant HeLa cells. *Molecular medicine reports* **2012**, *5*, 603–9.
- [81] Zhang, S.; Yang, X.; Morris, M. E. Flavonoids are inhibitors of breast cancer resistance protein (ABCG2)-mediated transport. *Molecular pharmacology* **2004**, *65*, 1208–16.
- [82] Imai, Y.; Tsukahara, S.; Asada, S.; Sugimoto, Y. Phytoestrogens / Flavonoids Reverse Breast Cancer Resistance Protein / ABCG2-Mediated Multidrug Resistance. *Cancer Research* **2004**, *64*, 4346–4352.
- [83] Katayama, K.; Masuyama, K.; Yoshioka, S.; Hasegawa, H.; Mitsuhashi, J.; Sugimoto, Y. Flavonoids inhibit breast cancer resistance protein-mediated drug resistance: transporter specificity and structure-activity relationship. *Cancer chemotherapy and pharmacology* **2007**, *60*, 789–97.
- [84] An, G.; Morris, M. E. Effects of single and multiple flavonoids on BCRP-mediated accumulation, cytotoxicity and transport of mitoxantrone in vitro. *Pharmaceutical research* **2010**, *27*, 1296–308.
- [85] Nicolle, E.; Boccard, J.; Guilet, D.; Dijoux-Franca, M.-G.; Zelefac, F.; Macalou, S.; Grosselin, J.; Schmidt, J.; Carrupt, P.-A.; Di Pietro, A.; Boumendjel, A. Breast cancer resistance protein (BCRP/ABCG2): new inhibitors and QSAR studies by a 3D linear solvation energy approach. *European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences* **2009**, *38*, 39–46.

- [86] Juvale, K.; Pape, V. F. S.; Wiese, M. Investigation of chalcones and benzochalcones as inhibitors of breast cancer resistance protein. *Bioorganic & medicinal chemistry* **2012**, *20*, 346–55.
- [87] Pick, A.; Müller, H.; Mayer, R.; Haenisch, B.; Pajeva, I. K.; Weigt, M.; Bönisch, H.; Müller, C. E.; Wiese, M. Structure-activity relationships of flavonoids as inhibitors of breast cancer resistance protein (BCRP). *Bioorganic & medicinal chemistry* **2011**, *19*, 2090–102.
- [88] Juvale, K.; Stefan, K.; Wiese, M. Synthesis and biological evaluation of flavones and benzoflavones as inhibitors of BCRP/ABCG2. *European journal of medicinal chemistry* **2013**, *67*, 115–26.
- [89] Erlichman, C.; Boerner, S.; Hallgren, C. The HER Tyrosine Kinase Inhibitor CI1033 Enhances Cytotoxicity of 7-Ethyl-10-hydroxycamptothecin and Topotecan by Inhibiting Breast Cancer Resistance Protein-mediated Drug Efflux. *Cancer research* **2001**, *61*, 739–748.
- [90] Stewart, C.; Leggas, M.; Schuetz, J. Gefitinib Enhances the Antitumor Activity and Oral Bioavailability of Irinotecan in Mice. *Cancer research* **2004**, *64*, 7491–7499.
- [91] Houghton, P.; Germain, G.; Harwood, F. Imatinib Mesylate Is a Potent Inhibitor of the ABCG2 (BCRP) Transporter and Reverses Resistance to Topotecan and SN-38 in Vitro. *Cancer research* **2004**, *64*, 2333–2337.
- [92] Burger, H.; van Tol, H.; Boersma, A. W. M.; Brok, M.; Wiemer, E. a. C.; Stoter, G.; Nooter, K. Imatinib mesylate (STI571) is a substrate for the breast cancer resistance protein (BCRP)/ABCG2 drug pump. *Blood* **2004**, *104*, 2940–2.
- [93] Gupta, A.; Zhang, Y.; Unadkat, J.; Mao, Q. HIV protease inhibitors are inhibitors but not substrates of the human breast cancer resistance protein (BCRP/ABCG2). *Journal of Pharmacology and ...* **2004**, *310*, 334–341.
- [94] Janneh, O.; Owen, A.; Chandler, B.; Hartkoorn, R. Modulation of the intracellular accumulation of saquinavir in peripheral blood mononuclear cells by inhibitors of MRP1, MRP2, P-gp and BCRP. *Aids* **2005**, 2097–2102.
- [95] Chang, C.; Ekins, S.; Bahadduri, P.; Swaan, P. W. Pharmacophore-based discovery of ligands for drug transporters. *Advanced drug delivery reviews* **2006**, *58*, 1431–50.

Bibliography

- [96] Matsson, P.; Englund, G.; Ahlin, G.; Bergström, C.; Norinder, U.; Artursson, P. A Global Drug Inhibition Pattern for the Human ATP-Binding Cassette Transporter Breast Cancer Resistance Protein. *Journal of Pharmacology and Experimental Therapeutics* **2007**, *323*, 19–30.
- [97] Molecular Operating Environment (MOE), 2012.10. *Chemical Computing Group Inc.*, 2012.
- [98] Todeschini, R. *Molecular Descriptors for Chemoinformatics, Volumes I & II*; Wiley-VCH, 2009.
- [99] Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228.
- [100] Wildman, S.; Crippen, G. Prediction of Physicochemical Parameters by Atomic Contributions. *Journal of Chemical Information and Modeling* **1999**, *39*, 868–873.
- [101] Labute, P. A widely applicable set of descriptors. *Journal of molecular graphics & modelling* **2000**, *18*, 464–77.
- [102] Pearlman, R. S.; Smith, K. M. Novel Software Tools for Chemical Diversity. *Perspectives in drug discovery and design* **1998**, 339–353.
- [103] Burden, F. Molecular Identification Number for Substructure Searches. *Journal of Chemical Information and Computer ...* **1989**, 225–227.
- [104] Halgren, T. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of computational chemistry* **1996**, *17*, 490–519.
- [105] Meyer, a. Y. The size of molecules. *Chemical Society Reviews* **1986**, *15*, 449.
- [106] Cruciani, G.; Pastor, M.; Guba, W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences* **2000**, *11 Suppl 2*, S29–39.
- [107] Breneman, C. M. C.; Rhem, M. QSPR analysis of HPLC column capacity factors for a set of high energy materials using electronic van der waals surface property descriptors computed by. *Journal of computational chemistry* **1997**, *18*, 182–197.
- [108] Lavine, B. K.; Davidson, C. E.; Breneman, C.; Katt, W.; Sundling, C. M. Electronic van der Waals surface property descriptors and genetic algorithms

- for developing structure-activity correlations in olfactory databases. *Journal of chemical information and computer sciences* **2003**, *43*, 1890–905.
- [109] Hemmer, M. Deriving the 3D structure of organic molecules from their infrared spectra. *Vibrational Spectroscopy* **1999**, *19*, 151–164.
- [110] Cherkasov, A. Inductive QSAR Descriptors. Distinguishing Compounds with Antibacterial Activity by Artificial Neural Networks. *International Journal of Molecular Sciences* **2005**, *6*, 63–86.
- [111] Moreau, G.; Broto, P. The autocorrelation of a topological structure: A new molecular descriptor. *Nouv. J. Chim.* **1980**, *4*, 359–360.
- [112] Gasteiger, J.; Sadowski, J. Chemical information in 3D space. *Journal of chemical information and computer sciences* **1996**, *2338*, 1030–1037.
- [113] VCCLAB, *Virtual Computational Chemistry Laboratory*.
- [114] Todeschini, R.; Gramatica, P. The Whim Theory: New 3D Molecular Descriptors for Qsar in Environmental Modelling. *SAR and QSAR in Environmental Research* **1997**, *7*, 89–115.
- [115] Consonni, V.; Todeschini, R.; Pavan, M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. *Journal of chemical information and computer sciences* **2002**, *42*, 682–92.
- [116] Kohonen, T. Self-Organized Formation of Topologically Correct Feature Maps. *Biological cybernetics* **1982**, *69*, 59–69.
- [117] Jure Zupan, J. G. *Neural Networks in Chemistry and Drug Design*; Wiley-VCH, 1999.
- [118] Anzali, S.; Gasteiger, J.; Holzgrabe, U. The use of self-organizing neural networks in drug design. *Perspectives in drug discovery and design* **1998**, 273–299.
- [119] Kohonen, T. The Self-Organizing Map. *Proceedings of the IEEE* **1990**, *78*, 1464–1480.
- [120] Gasteiger, J.; Teckentrup, A.; Terfloth, L.; Spycher, S. Neural networks as data mining tools in drug design. *Journal of Physical Organic Chemistry* **2003**, *16*, 232–245.
- [121] Guha, R.; Serra, J. R.; Jurs, P. C. Generation of QSAR sets with a self-organizing map. *Journal of molecular graphics & modelling* **2004**, *23*, 1–14.

Bibliography

- [122] Vapnik, V.; Cortes, C. Support-vector networks. *Machine learning* **1995**, *20*, 273–297.
- [123] Vapnik, V. *Estimation of dependences based on empirical data*; 2006.
- [124] Ben-Hur, A.; Ong, C. S.; Sonnenburg, S.; Schölkopf, B.; Rätsch, G. Support vector machines and kernels for computational biology. *PLoS computational biology* **2008**, *4*, e1000173.
- [125] Moss, G. P.; Shah, a. J.; Adams, R. G.; Davey, N.; Wilkinson, S. C.; Pugh, W. J.; Sun, Y. The application of discriminant analysis and Machine Learning methods as tools to identify and classify compounds with potential as transdermal enhancers. *European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences* **2012**, *45*, 116–27.
- [126] Li, H.; Yap, C. W.; Ung, C. Y.; Xue, Y.; Li, Z. R.; Han, L. Y.; Lin, H. H.; Chen, Y. Z. Machine learning approaches for predicting compounds that interact with therapeutic and ADMET related proteins. *Journal of Pharmaceutical Sciences* **2007**, *96*, 2838–2860.
- [127] Burbidge, R.; Trotter, M.; Buxton, B.; Holden, S. Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Computers & chemistry* **2001**, *26*, 5–14.
- [128] Noble, W. What is a support vector machine? *Nature biotechnology* **2006**, *24*, 1565–1567.
- [129] Tarca, A. L.; Carey, V. J.; Chen, X.-w.; Romero, R.; Drăghici, S. Machine learning and its applications to biology. *PLoS computational biology* **2007**, *3*, e116.
- [130] Witten, Ian H. & Eibe, F. *Data mining*; 2005.
- [131] Kubinyi, H. QSAR and 3D QSAR in drug design Part 1: methodology. *Drug Discovery Today* **1997**, *2*, 457–467.
- [132] Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society* **1988**, *110*, 5959–67.
- [133] Zhu, R.; Liu, Q.; Tang, J.; Li, H.; Cao, Z. Investigations on inhibitors of hedgehog signal pathway: a quantitative structure-activity relationship study. *International journal of molecular sciences* **2011**, *12*, 3018–33.

- [134] Wang, R.; Gao, Y.; Liu, L.; Lai, L. All-Orientation Search and All-Placement Search in Comparative Molecular Field Analysis. *Journal of Molecular Modeling* **1998**, *4*, 276–283.
- [135] Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *Journal of medicinal chemistry* **1994**, *37*, 4130–46.
- [136] Klebe, G.; Abraham, U. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *Journal of computer-aided molecular design* **1999**, *13*, 1–10.
- [137] Lindl, T.; Lewandowski, B. An Evaluation of the In Vitro Cytotoxicities of 50 Chemicals by using an Electrical Current Exclusion Method versus the Neutral Red Uptake and MTT Assays. *ATLA. Alternatives to ...* **2005**, *33*, 1–11.
- [138] Tawar, U.; Jain, A. K.; Dwarakanath, B. S.; Chandra, R.; Singh, Y.; Chaudhury, N. K.; Khaitan, D.; Tandon, V. Influence of phenyl ring disubstitution on bisbenzimidazole and terbenzimidazole cytotoxicity: synthesis and biological evaluation as radioprotectors. *Journal of medicinal chemistry* **2003**, *46*, 3785–92.
- [139] Shapiro, A.; Ling, V. Reconstitution of drug transport by purified P-glycoprotein. *Journal of biological chemistry* **1995**, *270*, 16167–75.
- [140] Müller, H.; Klinkhammer, W.; Globisch, C.; Kassack, M. U.; Pajeva, I. K.; Wiese, M. New functional assay of P-glycoprotein activity using Hoechst 33342. *Bioorganic & medicinal chemistry* **2007**, *15*, 7470–9.
- [141] Holló, Z.; Homolya, L.; Davis, C.; Sarkadi, B. Calcein accumulation as a fluorometric functional assay of the multidrug transporter. *Biochimica et Biophysica Acta (...)* **1994**, *1191*, 384–388.
- [142] Liminga, G.; Nygren, P.; Larsson, R. Microfluorometric evaluation of calcein acetoxymethyl ester as a probe for P-glycoprotein-mediated resistance: effects of cyclosporin A and its nonimmunosuppressive. *Experimental cell research* **1994**, *212*, 291–296.
- [143] Mosmann, T. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *Journal of immunological methods* **1983**, *65*, 55–63.
- [144] Krishnamurthy, P.; Schuetz, J. D. Role of ABCG2/BCRP in biology and medicine. *Annual review of pharmacology and toxicology* **2006**, *46*, 381–410.

Bibliography

- [145] Kim, M.; Turnquist, H.; Jackson, J.; Sgagias, M. The Multidrug Resistance Transporter ABCG2 (Breast Cancer Resistance Protein 1) Effluxes Hoechst 33342 and Is Overexpressed in Hematopoietic Stem Cells. *Clinical Cancer ...* **2002**, *2*, 22–28.
- [146] Müller, H. Ph.D. thesis, University of Bonn, 2007.
- [147] Martin, C.; Berridge, G.; Mistry, P.; Higgins, C.; Charlton, P.; Callaghan, R. The molecular interaction of the high affinity reversal agent XR9576 with P-glycoprotein. *British journal of pharmacology* **1999**, *128*, 403–11.
- [148] Pick, A.; Müller, H.; Wiese, M. Structure-activity relationships of new inhibitors of breast cancer resistance protein (ABCG2). *Bioorganic & medicinal chemistry* **2008**, *16*, 8224–36.
- [149] Klinkhammer, W.; Müller, H.; Globisch, C.; Pajeva, I. K.; Wiese, M. Synthesis and biological evaluation of a small molecule library of 3rd generation multidrug resistance modulators. *Bioorganic & medicinal chemistry* **2009**, *17*, 2524–35.
- [150] Pick, A.; Klinkhammer, W.; Wiese, M. Specific inhibitors of the breast cancer resistance protein (BCRP). *ChemMedChem* **2010**, *5*, 1498–505.
- [151] Cramer, J.; Kopp, S.; Bates, S. E.; Chiba, P.; Ecker, G. F. Multispecificity of drug transporters: probing inhibitor selectivity for the human drug efflux transporters ABCB1 and ABCG2. *ChemMedChem* **2007**, *2*, 1783–8.
- [152] Mueller, H.; Kassack, M. U.; Wiese, M. Comparison of the usefulness of the MTT, ATP, and calcein assays to predict the potency of cytotoxic agents in various human cancer cell lines. *Journal of biomolecular screening* **2004**, *9*, 506–15.
- [153] Chiba, P.; Burghofer, S.; Richter, E.; Tell, B.; Moser, a.; Ecker, G. Synthesis, pharmacologic activity, and structure-activity relationships of a series of propafenone-related modulators of multidrug resistance. *Journal of medicinal chemistry* **1995**, *38*, 2789–93.
- [154] Klinkhammer, W. Ph.D. thesis, University of Bonn, 2006.
- [155] Steggemann, K. Ph.D. thesis, University of Bonn, 2011.
- [156] Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry* **1996**, *17*, 490–519.

- [157] Wehrens, R.; Buydens, L. Self- and Super-organising Maps in R: the kohonen package. *J. Stat. Softw.* **2007**, *21*, year.
- [158] Kohonen, T. Essentials of the self-organizing map. *Neural networks : the official journal of the International Neural Network Society* **2012**, *37*, 52–65.
- [159] Möllmann, L. Ph.D. thesis, University of Bonn, 2011.
- [160] Meyer, D.; Wien, T. U. *Support Vector Machines. The Interface to libsvm in package e1071. Online-Documentation of the package e1071 R*; 2001.
- [161] Melville, J. L.; Burke, E. K.; Hirst, J. D. Machine learning in virtual screening. *Combinatorial chemistry & high throughput screening* **2009**, *12*, 332–43.
- [162] Juvale, K. Ph.D. thesis, University of Bonn, 2013.
- [163] Kubinyi, H. Free Wilson Analysis . Theory , Applications and its Relationship to Hansch Analysis. *Quantitative Structure-Activity Relationships* **1988**, *133*, 121–133.
- [164] Clark, T. *A Handbook of Computational Chemistry*; 1985.
- [165] *QuaSAR-Evolution*; Scientific Vector Language (SVL) source code provided by Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite nr. 910, Montreal, QC, Canada, H3A 2R7, 2015.
- [166] R, version 2.15.3. *R Development Core Team*, Vienna, Austria, 2012.
- [167] Tripos (2011) SYBYL-X 1.2. *Tripos International*, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
- [168] Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR & Combinatorial Science* **2007**, *26*, 694–701.
- [169] Veerasamy, R.; Rajak, H.; Jain, A. Validation of QSAR Models-Strategies and Importance. *International Journal of Drug Design and Discovery* **2011**, *2*, 511–519.
- [170] Tropsha, A.; Gramatica, P.; Gombar, V. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR & Combinatorial Science* **2003**, *22*, 69–77.