



UNIVERSITÉ DE STRASBOURG



en cotutelle avec

INSTITUT BOGATSKY DE CHIMIE PHYSIQUE, ODESSA, UKRAINE

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUE

UMR 7140 Unistra/CNRS

THÈSE

présentée par

Tetiana KHRISTOVA

soutenue le 15 novembre 2013

pour obtenir le grade de

Docteur de l'Université de Strasbourg

Discipline: Chimie

Spécialité: Chemoinformatique

## **Computer-aided design of novel antithrombotic agents**

Directeurs de thèse:

M. VARNEK Alexandre

Professeur, Université de Strasbourg

M. KUZMIN Victor  
Ukraine

Professeur, Institut Bogatsky de Chimie Physique, Odessa,

Rapporteurs:

M. VILLOUTREIX Bruno

Docteur, Université Paris-Diderot

M. TETKO Igor

Docteur, Institut of Structural Biology, Helmholtz Zentrum  
München, Neuherberg, Allemagne

Examineurs:

M. STOTE Roland

Docteur, Université de Strasbourg

M. LANGER Thierry

Professeur, Prestwick Chemical, Strasbourg

*Ms Tetiana Khristova was a member of the European Doctoral College of the University of Strasbourg during the preparation of her PhD, from 2010 to 2013, class name Jane Goodall. She has benefited from specific financial supports offered by the College and, along with her mainstream research, has followed a special course on topics of general European interests presented by international experts. This PhD research project has been led with the collaboration of two universities: the A.V.Bogatsky Physico-Chemical Institute, Ukraine and the University of Strasbourg, France.*

## **Acknowledgment**

I would like to thank French Embassy in Ukraine for the PhD fellowship. As well I want to express my gratitude to my supervisors Professor Alexandre Varnek and Professor Victor Kuz'min for the opportunity to participate in this exciting French-Ukrainian project. Without their guidance and persistent help this PhD thesis would not have been possible.

Then I would like to express my gratitude to Dr. Bruno Villoutreix, Dr. Igor Tetko, Dr. Roland Stote and Pr. Thierry Langer for their time and effort in reviewing this work.

Furthermore I would like to thank Dr. Pavel Polishchuk without whose help and moral support this work would have never been finished successfully.

I would like to thank my colleagues at the laboratory Dr. Dragos Horvath, Dr. Gilles Marcou, Dr. Olga Klimchuk, Dr. Vladimir Chupakhin. I am grateful for their constant support and help.

I cannot find words to express my gratitude to Dr. Ioana Oprisiu who have supported me throughout entire process. She was always there for me when I needed her.

I also want to thank some of my fellow PhD students: Laurent Hoffer, Christophe Muller, Fiorella Ruggiu, H el ena Gaspar. They each helped to make my time in the PhD program more fun and interesting.

I am deeply and forever indebted to my parents and friends for their love and support. Thank you for believing in me.

## Abbreviation

<b>ACS</b>	: Acute coronary syndrome
<b>AD</b>	: Applicability domain
<b>ADP</b>	: Adenosine diphosphate
<b>AUC</b>	: Area under ROC curve
<b>BA</b>	: Balanced Accuracy
<b>CoMFA</b>	: Comparative Molecular Field Analysis
<b>FPT</b>	: ISIDA fuzzy pH-dependent pharmacophore triplets
<b>ISIDA</b>	: In Silico Design and Data Analysis
<b>MIDAS</b>	Metal Ion-Dependent Adhesion Site
<b>MOE</b>	: Molecular Operating Environment
<b>PASS</b>	: Prediction of Activity Spectra for Substances
<b>PLANTS</b>	: Protein-Ligand ANT System
<b>QSAR</b>	: Quantitative structure-activity relationship
<b>RF</b>	: Random forest method
<b>RMSD</b>	: Root-mean-square deviation
<b>RMSE</b>	: Root mean-squared error
<b>ROC</b>	: Receiver Operating Characteristic
<b>ROCS</b>	: Rapid Overlay of Chemical Structures
<b>SiRMS</b>	: Simplex representation of molecular structure
<b>SMF</b>	: Substructure molecular fragments
<b>TF</b>	: Tissue factor
<b>TP</b>	: Thromboxane A2 receptor
<b>TXA2</b>	: Thromboxane A2
<b>vWF</b>	: von Willebrand factor
<b>XED</b>	: eXtended Electron Distribution force field

## Contents

Introduction.....	8
Chapter 1. Bibliographic overview .....	10
1. Mechanism of thrombosis.....	11
2. Known agents for antithrombotic therapy .....	18
Chapter 2. Used methodologies and their applications.....	23
3. Virtual screening funnel.....	24
3.1. QSAR approach .....	25
3.1.1. Molecular descriptors .....	25
3.1.1.1. Simplex representation of molecular structure .....	25
3.1.1.2. ISIDA substructure molecular fragments .....	26
3.1.1.3. ISIDA fuzzy pH-dependent pharmacophore triplets .....	27
3.1.2. Machine-learning methods .....	28
3.1.2.1. Random Forest.....	28
3.1.2.2. Applicability domain .....	29
3.1.2.3. Criteria for estimation of the performance of QSAR models.....	29
3.1.2.3.1. Regression models .....	29
3.1.2.3.2. Classification models.....	30
3.1.2.3.3. The ROC curve (Receiving Operating Characteristics) .....	31
3.1.2.4. Validation of QSAR models .....	31
3.1.3. ADME/Tox assessments of novel compounds .....	32
3.1.4. PASS.....	32
3.2. Molecular shape-similarity .....	33
3.2.1. ROCS. Rapid Overlay of Chemical Structures.....	33
3.3. Molecular field similarity .....	34
3.3.1. Cresset FieldAlign approach.....	35
3.4. Pharmacophore modeling .....	37
3.4.1. LigandScout .....	38
3.5. Omega.....	40
3.6. Docking.....	41
3.6.1. MOE, FlexX, PLANTS .....	42
Chapter 3. RGD-peptidomimetics – antagonists of integrin $\alpha_{IIb}\beta_3$ receptor .....	46
4. Previously reported models .....	47

4.1. Pharmacophore models .....	47
4.2. QSAR modeling.....	50
4.3. Molecular docking .....	54
4.4. Conclusions.....	59
5. Computer-aided design of new RGD-peptidomimetics.....	60
5.1. Dataset description.....	60
5.2. QSAR modeling.....	63
5.3. Pharmacophore models .....	67
5.3.1. Structure-based models.....	68
5.3.2. Ligand-based models .....	69
5.4. Docking.....	72
5.4.1. Choice of the docking software and binding site preparation.....	72
5.4.2. Docking experiments on selected subsets of RGD peptidomimetics. ....	76
5.5. Molecular shape-based comparison.....	82
5.6. Molecular fields similarity .....	84
5.7. Virtual screening of databases of commercially available compounds .....	85
5.8. Design and screening of $\alpha_{IIb}\beta_3$ -focused virtual compound library .....	86
5.9. Experimental validation.....	89
5.10. Conclusions.....	89
Chapter 4. Development of RUC-1 and RUC-2 analogs - antagonists of $\alpha_{IIb}\beta_3$ receptor .....	91
6. Development of RUC-1 and RUC-2 ligands .....	92
6. 1. Previously reported <i>in silico</i> screening study .....	93
7. Computer-aided design of antagonists of $\alpha_{IIb}\beta_3$ receptor found in closed form.....	95
7.1. Pharmacophore models.....	95
7.2. Ligand to Protein docking.....	97
7.3. Virtual screening of commercially available compounds.....	98
7.4. Design of focused libraries .....	101
7.5. Virtual screening of designed compounds.....	103
7.6. Experimental validation.....	103
7.7. Conclusions.....	105
Chapter 5. Development of antagonists of thromboxane A2 receptor.....	107
8. Previously reported models .....	108
8.1. Pharmacophore models.....	108
8.2. QSAR modeling.....	111

8.3. Ligand to Protein docking.....	116
8.4. Conclusions.....	118
9. Computer-aided design of thromboxane A2 receptor antagonists.....	120
9.1. Dataset preparation .....	120
9.2. QSAR modeling.....	121
9.3. Pharmacophore models .....	123
9.4. Molecular shape-based comparison.....	126
9.5. Molecular fields similarity .....	128
9.6. Virtual screening of BioInfoDB .....	129
9.7. Library design by pharmacophore fragment-based approach.....	131
9.8. Library design based on molecular fields similarity.....	131
9.9. Virtual screening of designed libraries .....	132
9.10. Conclusions.....	134
General conclusions .....	135
References.....	139
APPENDIX 1. Biological experiments for antagonists of integrin $\alpha_{IIb}\beta_3$ .....	154
APPENDIX 2. Publications .....	160
APPENDIX 3. Résumé de la thèse de doctorat.....	180

## Introduction

Thrombotic diseases cause high mortality worldwide. The thrombogenic process is complex and multi-staged. A lot of receptor systems are involved in pathogenic thrombosis. Receptors found on the surface of platelets are actively participating in this process, in particular, integrin  $\alpha_{IIb}\beta_3$  and thromboxane A2 receptors. The first one is responsible for the interaction of platelets with fibrinogen to form clots, and the second one is responsible for the platelet activation by one of agonists excreted by adjacent platelets - thromboxane A2.

The classical antagonists of the  $\alpha_{IIb}\beta_3$  receptor, RGD peptidomimetics, have proven their usefulness in reducing the risk of periprocedural myocardial infarction and urgent target vessel revascularization during catheterization and have claimed a place in therapy for these indications. Nowadays three antagonists of the  $\alpha_{IIb}\beta_3$  receptor are commercial drugs: Abciximab, Eptifibatid and Tirofiban. They are pretty expensive and have some severe side effects, such as thrombocytopenia and bleeding. This motivates researchers to develop novel potent  $\alpha_{IIb}\beta_3$  receptor antagonists.

Thrombocytopenia is associated with the conformational changes of the  $\alpha_{IIb}\beta_3$  receptors, which are induced by binding of known antagonists. Recently, Coller et al. [1-2] reported development of new non-classical  $\alpha_{IIb}\beta_3$  receptor antagonists RUC-1 and RUC-2. Unlike the above mentioned marketed drugs, which bind the open form of the  $\alpha_{IIb}\beta_3$  receptor, the discovered ligands bind its closed form. This doesn't induce the conformational changes of the protein and, hence, the risk of undesirable side effects is reduced. Thus, the development of ligands having binding mechanisms similar to RUC1 and RUC-2 represents a promising way to design novel antithrombotic agents.

Antagonists of thromboxane A2 receptors (TP receptor) could be particularly useful in treatment of acute myocardial ischemia, heart failure and exhibit cardioprotective effects. However, up to now no marketed antagonists of the TP receptor are available. All investigated agents are less efficient than aspirin. The latter is not an antagonist of TP receptor as such, but targets platelet thromboxane A2 synthesis by irreversibly acetylating cyclooxygenase-1 and 2. Aspirin has some adverse effects: gastrointestinal toxicity, resistance, hemorrhagic events, especially minor, gastrointestinal, and total bleeding, and stroke. At the pharmacological level, thromboxane A2 receptor antagonists could be more advantageous than low-dose aspirin. Indeed, thromboxane A2 receptor antagonists inhibit the deleterious effects of other endogenous thromboxane A2 receptor ligands such as endoperoxides, prostanoids and isoprostanes, whose



formation would not be affected by aspirin, other COX inhibitors or thromboxane synthase inhibitors.

The given work has been carried out in the framework of collaboration between the University of Strasbourg and the A.V. Bogatsky Physico-Chemical Institute of National Academy of Sciences of Ukraine (PCI). The Medicinal department of PCI deals with the development of new anti-thrombotic agent, having less pronounced side effects and lower price than commercial drugs. During last decade, a lot of row data for the molecules possessing concerning anti-thrombotic agents has been accumulated in PCI. However, this information was never used for computer-aided design of new compounds. Discovery of drugs involves difficult, expensive and time consuming procedures, which require a lot of human effort. Chemoinformatics may minimize many of the above efforts. For example, identification of potential lead compounds in the drug discovery life cycle helps to reduce the cost and time of clinical trials. There are a lot of well established tools in chemoinformatics, which have both advantages as well as disadvantages. The consensus approach should be used to reduce the disadvantages and obtain more reliable results. In this study, the ensemble of modern computational approaches was used – QSAR, pharmacophore modeling, molecular docking, molecular fields and molecular shapes similarity analysis integrated in virtual screening workflow. This work represents the first (and successful) attempt of an *in silico* design of new anti-thrombotic agents using both experimental data collected in PCI and available literature data.

Our study resulted in suggestion of new potential antagonists of  $\alpha_{IIb}\beta_3$  and TP receptors. Suggested antagonists of  $\alpha_{IIb}\beta_3$  able to bind either open or closed form of the receptor have been synthesized and tested experimentally in PCI. Experiments show that some of theoretically designed compounds are more efficient than Tirofiban – the commercialized drug molecule. They are expected to be inexpensive and, hence, more accessible for the population of Ukraine. The recommended antagonists of TP receptor have been already synthesized in PCI but biological tests have not been completed yet.

This manuscript consists of four parts. The first one represents a bibliographic review describing the mechanism of thrombosis and known agents for antithrombotic therapy. The second part describes the virtual screening methods used in this study. Two other parts concern computer-aided design of (i) new antagonists of  $\alpha_{IIb}\beta_3$  receptor and (ii) antagonists of thromboxane A2 receptor.

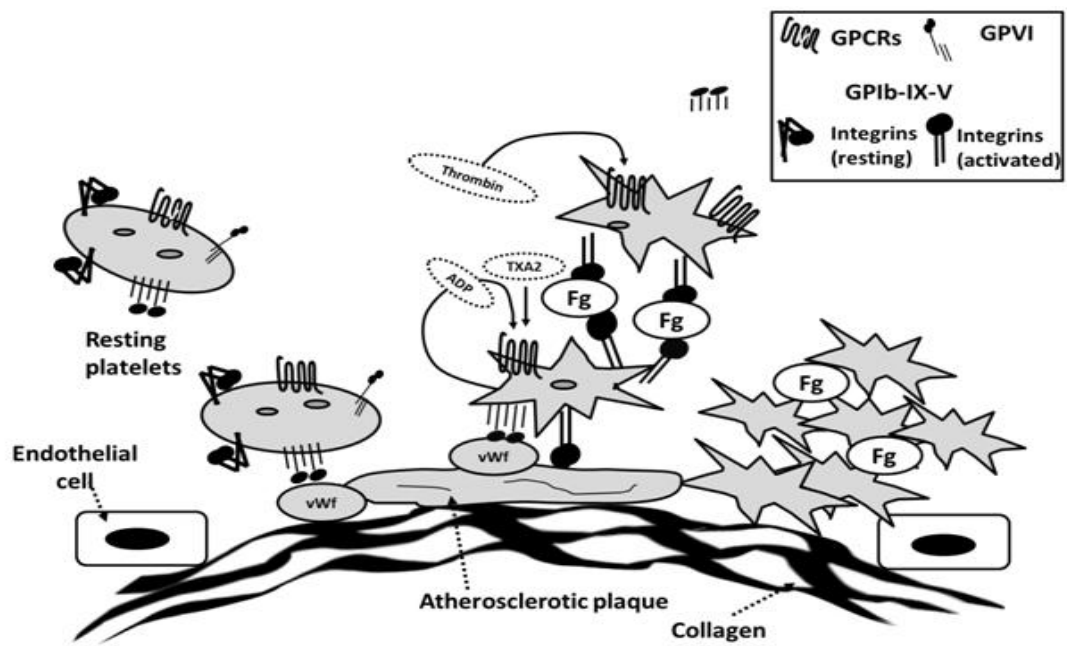
*Chapter 1.*

**Bibliographic overview**

## 1. Mechanism of thrombosis

Hemostasis is the host defense system aimed at preserving the integrity of the circulatory system in mammals. It is a highly complex and tightly regulated process, which maintains the balance between pro-coagulation and anti-coagulation factors. The endothelium is a surface that separates blood cells and soluble plasma proteins involved in blood coagulation from subendothelial vessel components possessing pro-coagulation properties. Physiologically, hemostasis is activated after injury of the vessel wall. Exposure of the subendothelial matrix is one of the most important events following vessel injury. After the exposure of an endothelial surface to the blood flow, which is rich in negatively charged phospholipids, they can bind coagulation factors and other hemostatic molecules (such as the tissue factor), thereby promoting prothrombotic signaling. Once tissue factor was released from injured endothelial cells, inactive blood coagulation proteins (zymogens) are sequentially converted into the corresponding active enzymes, through a cascade of sequential, calcium-dependent enzymatic reactions. The rapid production of thrombin marks the initiation of thrombus formation.[3]

Pathological thrombosis occurs when the hemostatic pathway is so strongly activated that it exceeds the normal regulatory counterbalance by anticoagulant factors [3]. Arterial and venous thromboses differ in both their composition and conditions of formation. Thrombi in the arterial circulation are mainly platelet-rich and are formed under high shear stress, whereas thrombi in the venous circulation are fibrin-rich and are formed under low shear stress conditions[4]. Pathological arterial thrombosis represents the most frequent cause of death worldwide and is preceded by a complex interplay between environmental and genetic factors[5]. It is supposed that the main reason for such strong activation of the coagulation pathway in arteries can be the rupture of atherosclerotic plaque [6-9]. Plaque erosion and rupture are promoted by the relative abundance of the various subgroups of inflammatory leukocytes, a paucity of collagen, high expression of tissue factor, matrix-degrading proteinases, and oxidative stress-mediated apoptosis of cells contained in the atherosclerotic plaque[8]. Plaque rupture exposes the subendothelial matrix and a variety of thrombogenic material located in the core of the plaque to the arterial circulation. Atherosclerotic plaques contain such platelet activating molecules as tissue factor, von Willebrand factor (vWF), various types of collagens, fibronectin, vitronectin and many others [10-12]. When they are released into the arterial circulation, these molecules actively stimulate adhesion and aggregation of platelets, as well as their activation and secretion of their dense granules [9, 13] (Figure 1).



**Figure 1.** General scheme of pathological coagulation

Initially, platelets bind to the exposed extracellular matrix by interaction of GPIb-V-IX receptor complex with vWF, which itself is absorbed by collagen. This is the main primary interaction capable of tethering platelets to a surface of the vessel [14-15]. This interactions help to stabilize the platelet attachment at sites of injury or plaque rupture by reducing the velocity of platelets contacting the surface under high flow conditions, thereby increasing the time for other interactions[16]. The vWF molecule can act as signal transducer, in particular it has capacity to activate integrin  $\alpha_{IIb}\beta_3$ [14].

Platelets bind to collagen via the immunoglobulin superfamily receptor, GPVI, and integrin  $\alpha_2\beta_1$ . GPVI was found to be important for aggregate formation, but not for primary adhesion unlike to GPIb- and  $\alpha_2\beta_1$  receptors [17], which both are involved in primary hemostasis [18]. By itself, collagen is unable to mediate adhesion as it has low affinity to the receptor, but it plays important role in signaling. It triggers the intracellular signal that activates platelets. Activated platelets release the content of their  $\alpha$ - and dense granules and switch platelets integrins, such as  $\alpha_2\beta_1$  and  $\alpha_{IIb}\beta_3$  to high-affinity state[19].

The platelets contain two classes of secretory granules. The first ones are dense granules that secrete ADP and calcium ions, which reinforce platelet-platelet and platelet-surface coagulation. The second ones are  $\alpha$  granules, which secrete such proteins as vWF, platelet factor 4, fibrinogen and plasma proteins, such as albumin and IgG [20-21].

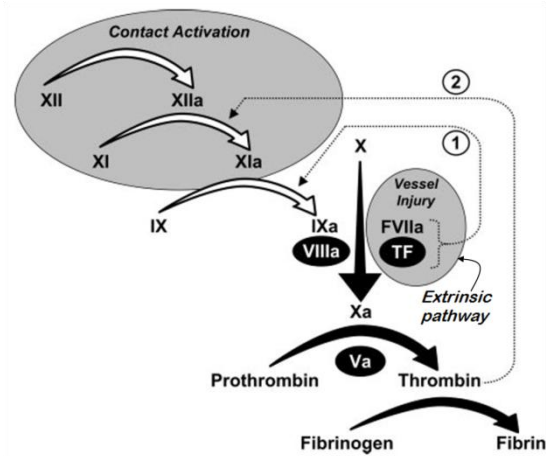
The agonists from dense granules, secondary mediators such as adenosine diphosphate (ADP) together with locally produced thrombin contribute to cellular activation by stimulating corresponding receptors that couple to heterodimeric G proteins, which induce different

signalling events and act synergistically to induce full platelet activation. If these agonists present in the blood flow in high concentration platelets can be activated independently of collagen [14].

ADP activates platelets both in an autocrine and a paracrine fashion [22]. Two types of ADP GPCR receptors (P2Y<sub>1</sub> and P2Y<sub>12</sub>) coupled with G<sub>q</sub>[23] and G<sub>12</sub>[24] proteins, respectively, exist on the surface of platelets: the P2Y<sub>1</sub> receptor initiates platelet shape change and ADP-induced aggregation through the mobilization of internal calcium stores, and the P2Y<sub>12</sub> receptor is necessary for full aggregation response to ADP and the stabilization of aggregates [25]. Concomitant intracellular signaling from both ADP receptors is essential for normal ADP-induced platelet aggregation and inhibition of signaling through either receptor, by specific antagonists, is sufficient to block ADP-induced platelet aggregation [26].

Activated platelets produce another positive-feedback mediator – thromboxane A<sub>2</sub> (TXA<sub>2</sub>), which binds with thromboxane A<sub>2</sub> receptor (TP) [27-28]. One of the most important physiological and pathophysiological actions of TXA<sub>2</sub> is platelet activation, that leads to platelet shape change, secretion, increasing of the expression of  $\alpha_{\text{IIb}}\beta_3$  receptors [27-28]. TP is a G protein-coupled receptor [29] as well as ADP receptors. It couples with multiple G proteins, but signalling through G<sub>q</sub> and G<sub>12/13</sub> proteins appear most relevant to TP function [30-31]. This leads to different biological responses, such as an increase in level of intracellular free calcium ions and exposure of  $\alpha_{\text{IIb}}\beta_3$  binding sites [28, 32]

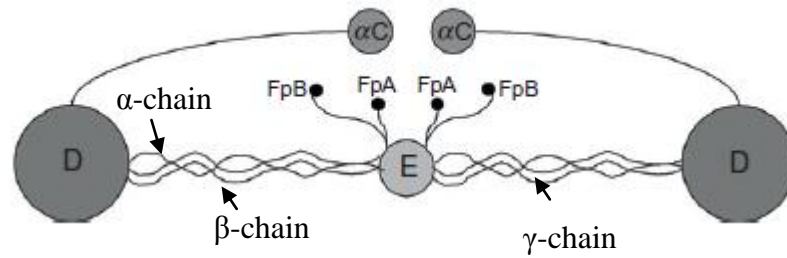
Thrombin is a critical contributor to secondary hemostasis via coagulation cascades. Its generation springs from tissue factor (TF), a type-I integral membrane protein, which circulates in the blood [33] and is contained in the endothelium and atherosclerotic plaques [34-35]. In the extrinsic coagulation pathway, TF plays a role of obligate cofactor for activation of zymogen factor X by formed complex with factor VIIa, so it converts to Xa factor. Factor Xa itself forms a complex with factor Va, which converts prothrombin to active thrombin [36] (Figure 2). The intrinsic (contact) coagulation pathway can be triggered by different ways, in particular *via* activation of factor IX with complex of TF-VIIa [37] and *via* activation of factor XI by thrombin [38]. This pathway leads to fast thrombin generation and is considered as an amplification of the extrinsic pathway [39].



**Figure 2.** Extrinsic and intrinsic coagulation pathways [40]. Roman numerals indicate unactivated coagulation factors, and activated factors are indicated by a lower case “a”. Nonenzymatic cofactors are indicated by numerals in black ovals. White arrows indicate reactions of the intrinsic pathway, and dotted lines indicate reactions that are not part of the standard cascade/waterfall model. The black arrows indicate the common pathway. TF - tissue factor.

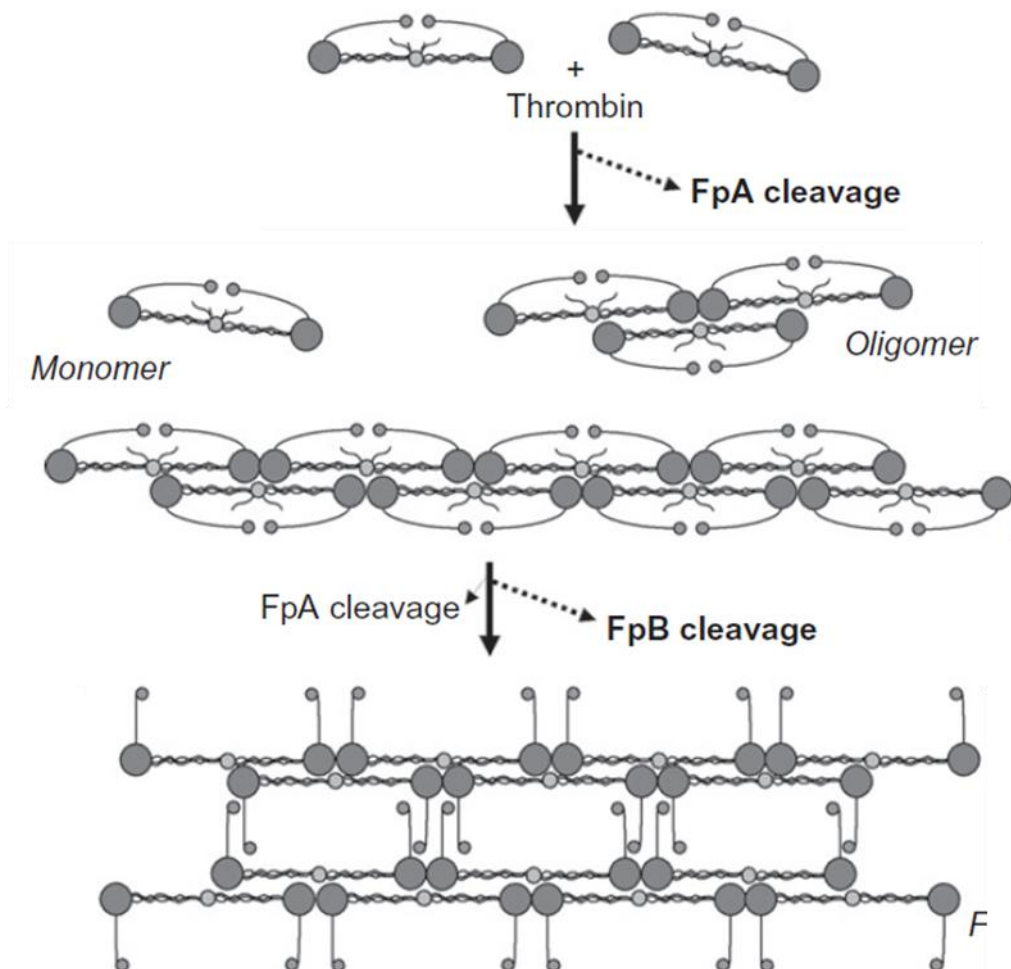
Another thrombin function is activation of platelets through two protease-activated G-protein-coupled receptors (PAR1 and PAR4). PARs are activated when thrombin cleaves its N-terminal extracellular domain, which unmask a new N terminus that acts as a tethered ligand “binding” to the receptor body (changing its conformation & orientation with respect to the rest of the protein), to effect trans-membrane signaling. Cleavage of the PAR receptors is irreversible, so they intervene only once, after which they are recycled by lysosomes [41]. Both PARs can couple to the  $G_{12/13}$ ,  $G_q$  and  $G_i$  proteins, which triggers a lot of intracellular effects [42], such as platelets shape changes,  $Ca^{2+}$  release [43-44].

One of the main functions of thrombin is a conversion of fibrinogen to fibrin. Fibrinogen is a glycoprotein synthesized by hepatocytes. It contains two symmetric parts, each consisting of three different polypeptide chains termed  $A\alpha$ ,  $B\beta$  and  $\gamma$  [45] that are held together by a series of disulphide bonds, where A and B designate corresponding fibrinopeptides. The fibrinogen molecule (Figure 3) has three distinct domains: two terminal D domains, linked to a single central E domain by a triple-stranded array of the polypeptide chains that are thought to exist in the form of  $\alpha$  helical coiled coils[46]. The E domain contains the N-terminal regions of  $\alpha$ - and  $\beta$ -chains designated as fibrinopeptide A (FPA) and B (FPB), which are cleaved by the thrombin [47]. These cleavages expose new N-terminal sequences, and convert fibrinogen into fibrin monomers which can spontaneously polymerize to form fibrin protofibrils [48].



**Figure 3.** Scheme of fibrinogen structure [48]

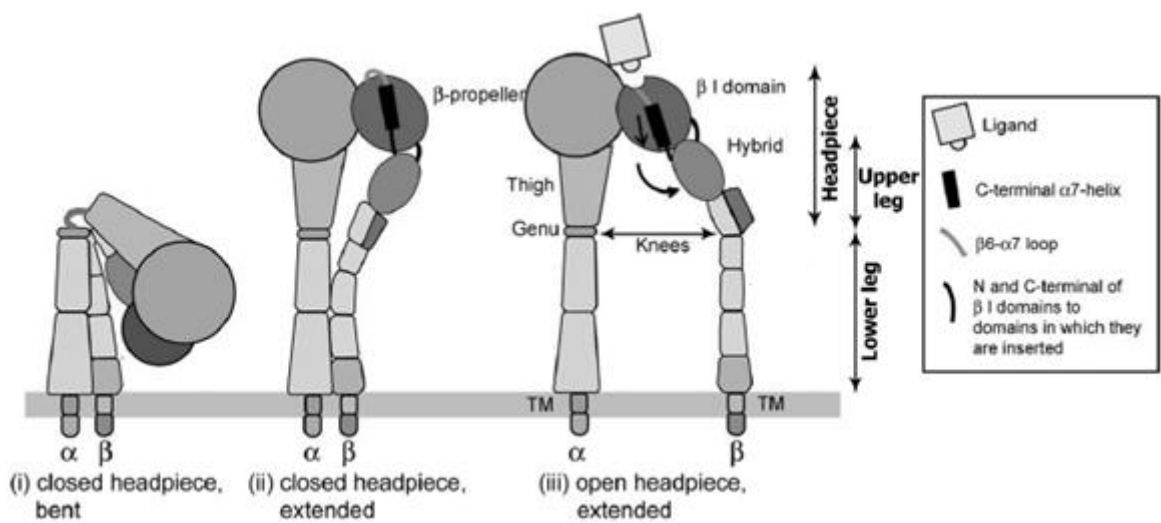
Fibrin can polymerize in different ways (Figure 4), what depends on relative rates of release of the FPA and FPB, the solution environment, including pH, ionic strength, temperature and presence or absence of other small molecules and proteins [49]. The linear growth of the fibrils predominates until they exceed 600 nm in length, afterwards lateral aggregation of fibrils into thicker fibers begins and fibers subsequently branch to form the three-dimensional fibrin network[50].



**Figure 4.** Schematic diagram of fibrin polymerization [48].

The last step in thrombus formation is a stabilization of formed clot, one of the main roles in this process plays integrin  $\alpha_{IIb}\beta_3$  receptor. It is one of the most abundant glycoprotein on platelets. Resting platelets contain about 80,000 surface copies of  $\alpha_{IIb}\beta_3$ , with additional pools of  $\alpha_{IIb}\beta_3$  in the membranes of  $\alpha$ -storage granules and the open-canalicular system [51-53].

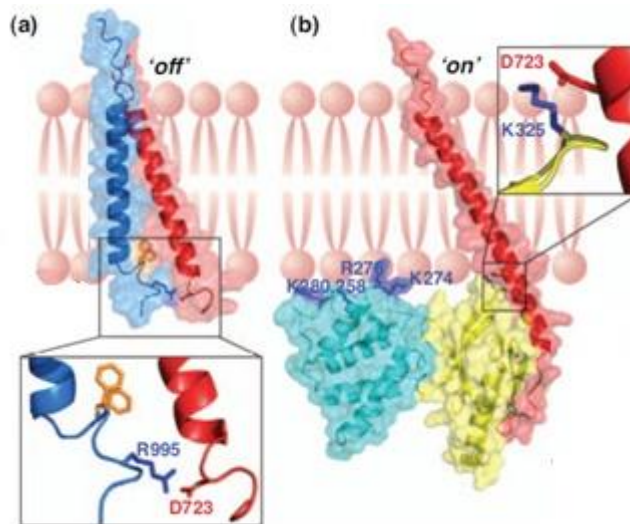
Integrins are cell surface-adhesion receptors consisting of noncovalently associated  $\alpha$ - and  $\beta$ -subunits [54] (Figure 5). In the rest state of platelets,  $\alpha_{IIb}\beta_3$  receptors are found in a closed form, such as the  $\alpha_{IIb}$ - and  $\beta_3$ -subunits are each bent at their knees. The bend brings receptor head and upper-leg domains into intimate contact with the lower-leg domains [55-56]. The bent conformation represents the low affinity state of the integrin. Regulation of affinity of integrin is associated with the transition from closed to open headpiece conformation [55, 57].



**Figure 5.** Conformational changes in  $\alpha_{IIb}\beta_3$  during the activation. [58]

In resting platelets, low-affinity conformation of  $\alpha_{IIb}\beta_3$  is stabilized by a “clasp” formed between the GFFKR sequence in  $\alpha_{IIb}$  and the NPxY motif in  $\beta_3$  [59], basically a salt bridge between  $\alpha_{IIb}$ Arg995 and  $\beta_3$ Asp723 [60]. The integrin can be activated by such cytoskeleton proteins as talin and kindlin which interact with cytoplasmic tail of  $\beta_3$ -subunit [61]. Talin activation (see Figure 6) can be caused by increase of cytoplasmic concentration of  $Ca^{2+}$  and diacylglycerol. Talin disrupts the clasp between  $\alpha_{IIb}$  and  $\beta_3$  subunits, what provokes conformational changes in the structure of the receptor and switches the integrin in “high”-affinity activated state. [62]





**Figure 6.** Integrin activation at the membrane-cytoplasm interface. (a) The inactive state of the integrin is represented by the complex formed between the transmembrane segments of the  $\alpha_{IIB}$  (light blue) and  $\beta_3$  (red) integrin subunits. (b) The complex formed between the  $\beta$  cytoplasmic tail (red) and the F2 (cyan) and F3 (yellow) subdomains of the talin head illustrates the activated integrin. [63]

In their high-affinity state  $\alpha_{IIB}\beta_3$  receptors can bind ligands containing an Arg-Gly-Asp (RGD) sequence, such as fibrinogen, fibrin, vWF, fibronectin, thrombospondin, and vitronectin [54]. The major ligand is fibrinogen, which binds to integrin via three recognition sequences RGDS [64], RGDF [65] that are placed on  $\alpha$ -chain of fibrinogen and HHLGGAKQAGDV on  $\gamma$ -chain [66]. Binding of platelets to fibrin network completes the thrombus formation.

## 2. Known agents for antithrombotic therapy

There are 3 main subgroups of antithrombotic agents which differ by their mechanism of action: anticoagulants, antiplatelet, and thrombolytic drugs [67]. Anticoagulant therapies primarily limit fibrin formation by decreasing the production of thrombin[68]. They are expected to be effective in preventing and treating venous thromboembolism, acute ischemic stroke, deep vein thrombosis, pulmonary embolism [69-71]. There are several marketed anticoagulant drugs: heparin and its derivatives, vitamin K antagonists (such as warfarin or coumadin) [69], dabigatran [72], rivaroxaban [73] and apixaban [74].

Heparin works by activating antithrombin III, which controls activities of thrombin inhibition and activates clotting factors IX, X, XI, and the tissue factor (VIIa complex); therefore, heparin affects the intrinsic clotting cascade [75]. It is effective in the initial treatment of patients with acute coronary syndromes [76-77], and in the prevention and treatment of venous thromboembolism [78-79].

The vitamin K antagonists (such as warfarin, coumadin, acenocoumarol or phenprocoumon) are the most frequently used anticoagulant agents for long-term prevention and treatment of a wide range of cardiovascular diseases [80]. For example, warfarin is indicated for the primary and secondary prevention of venous thromboembolism, prevention of embolism in patients with atrial fibrillation and prosthetic heart valves, and reduction of the risk of recurrent myocardial infarction.[81] Warfarin produces its anticoagulant effect by inhibiting the synthesis of the vitamin K-dependent clotting factors (factors II, VII, IX, and X) as well as the coagulation proteins C and S.[81-83]

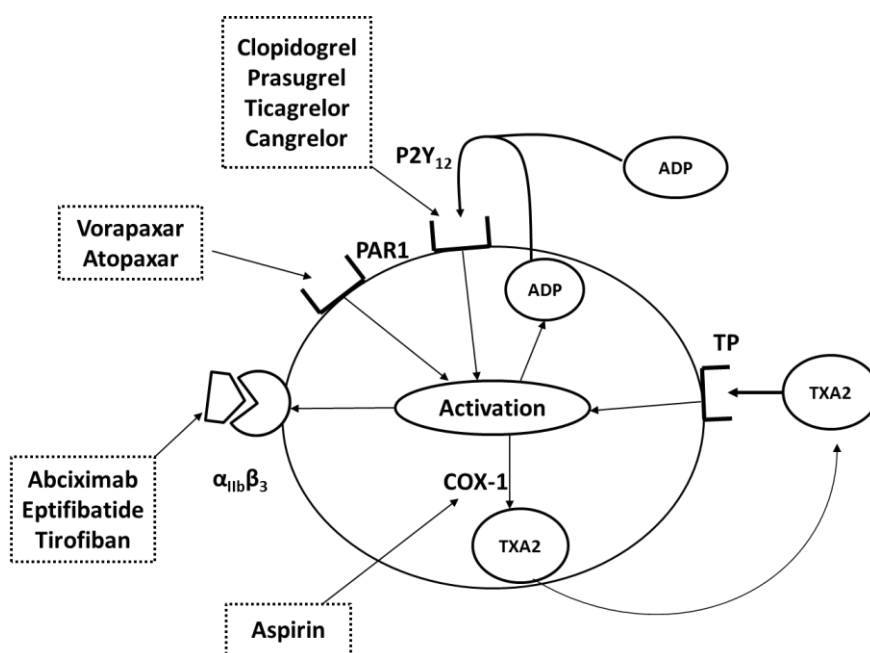
Dabigatran [72] is a low molecular weight prodrug that directly inhibits both free and fibrin-bound thrombin, which allows it to exert effects on both the coagulation cascade and platelets.

Rivaroxaban [73] is a synthetic molecule that selectively inhibits factor Xa, disrupting interactions with platelets, thrombi, or prothrombinase complex. Apixaban [74] is another direct factor Xa inhibitor; it binds factor Xa whether free or bound to platelets, thrombi, or prothrombinase complexes.

Thrombolytic drugs accelerate the transition of plasminogen to the active enzyme plasmin which degrades fibrin clots [84]. The purpose of their usage in clinical practice is the rapid lysis of formed clots in the body. The first-generation drug streptokinase (Streptase) binds to fibrin-bound plasminogen as well as free plasminogen that create a systemic fibrinolysis. The principal adverse effect associated with thrombolytic therapy is bleeding due to fibrinogenolysis or fibrinolysis at the site of vascular injury. By contrast to streptokinase, the second- and third-

generation agents (alteplase, reteplase (Retavase recombinant), tenecteplase) cause less extensive fibrinogenolysis, but bleeding occurs with a similar incidence for all agents [83]. These agents are indicated for patients with ST-segment elevation myocardial infarction, ischemic stroke, pulmonary embolism [69-70, 77, 85].

Antiplatelet therapy is the main therapy for the prevention and treatment of arterial thrombosis [86-87]. Aspirin[88], inhibitors of ADP receptors [85, 89-90] and antagonists of  $\alpha_{IIb}\beta_3$  [91-92] are effective in the treatment of patients with acute coronary syndromes, an acute myocardial infarction or ischemic stroke, unstable or stable angina, and in those undergoing percutaneous coronary intervention (Figure 7).

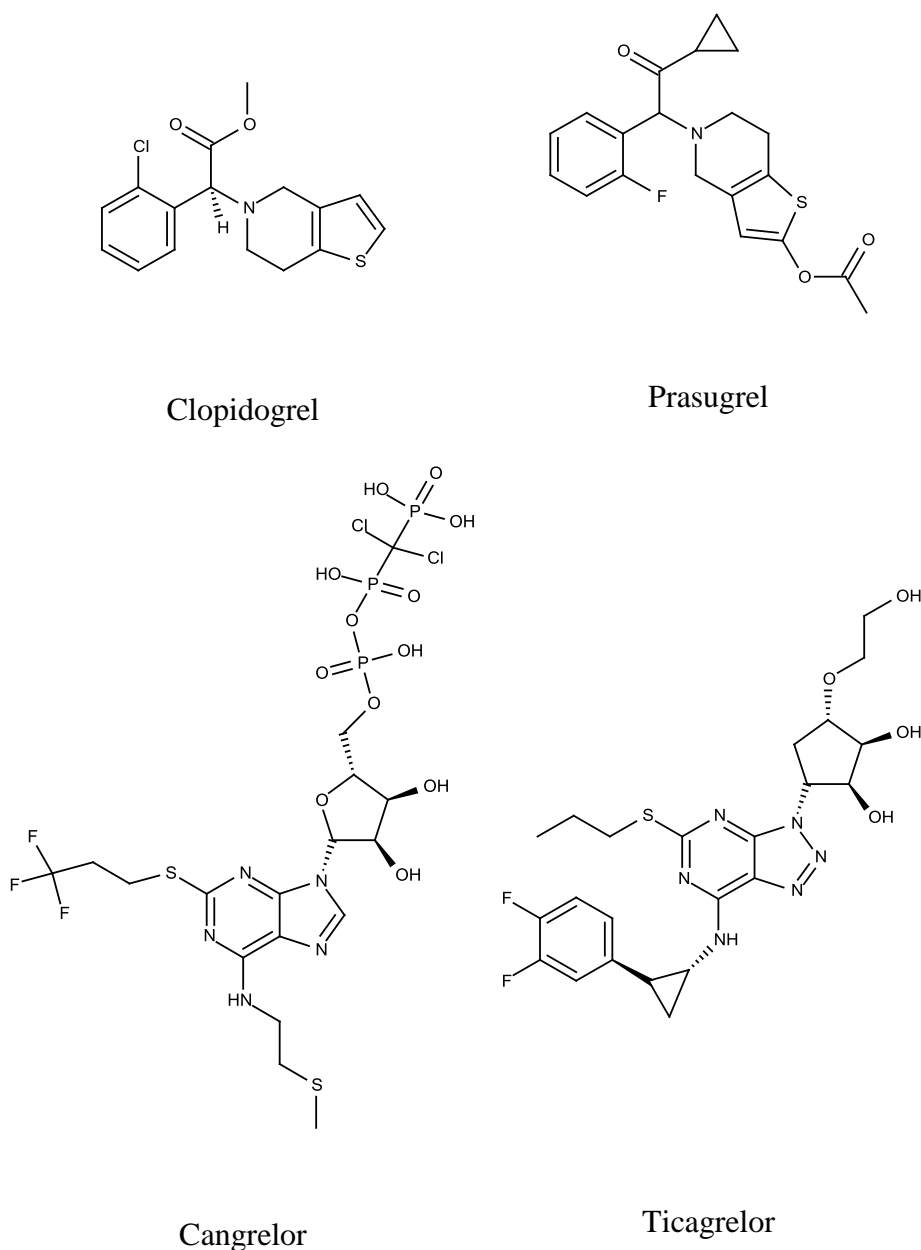


**Figure 7.** Known drugs and antagonist for antiplatelet therapy.

The known inhibitors that were approved as drugs are clopidogrel, prasugrel, cangrelor, and ticagrelor (Figure 8). These drugs inhibit P2Y<sub>12</sub> receptors. Clopidogrel is widely used in the treatment of cardiovascular diseases, including prevention of stroke, vascular disease, and myocardial infarction [93]. It is a second generation [94] prodrug that needs to be metabolized to its active derivative [95]. Afterwards it irreversibly interacts and blocks the function of the P2Y<sub>12</sub> receptors. The disadvantage of clopidogrel is a high variation of inter-individual pharmacological response [96] and its delayed onset and offset of action [86].

Prasugrel is a third generation prodrug, which as well as clopidogrel is metabolized, and its metabolite blocks the P2Y<sub>12</sub> receptor irreversibly by interaction with Cys97 and Cys175 residues[97]. In comparison to clopidogrel it has faster onset of action [98]. Although, prasugrel has benefits to clopidogrel, it has greater rate of life-threatening bleeding, including the fatal and nonfatal bleeding in comparison with clopidogrel [90].

In contrast to prasugrel and clopidogrel, cangrelor is a reversible drug with intravenous administration and rapid effect, because it doesn't need to be metabolized[99]. Despite these favourable pharmacodynamics properties, cangrelor was not superior to clopidogrel in reducing the incidence of ischemic events [100].As well as cangrelor, ticagrelor is characterized with rapid onset and reversibility [101]. It is better than clopidogrel in preventing major adverse cardiac events in ACS patients [102], but similar to prasugrel, ticagrelor is associated with high frequencies of bleeding complications [103-104].

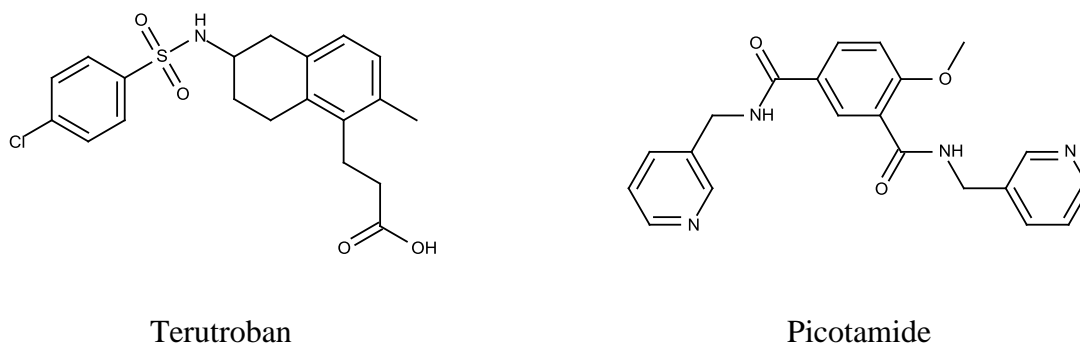


**Figure 8.** Antagonists of P2Y<sub>12</sub> receptor

No PAR-1 antagonist is currently approved for clinical use [105]. For now there are two orally active PAR1 antagonists, which are assessed in clinical trials for the prevention of arterial thrombosis, vorapaxar and atopaxar (Figure 9). Development of PAR4 antagonists is not thought

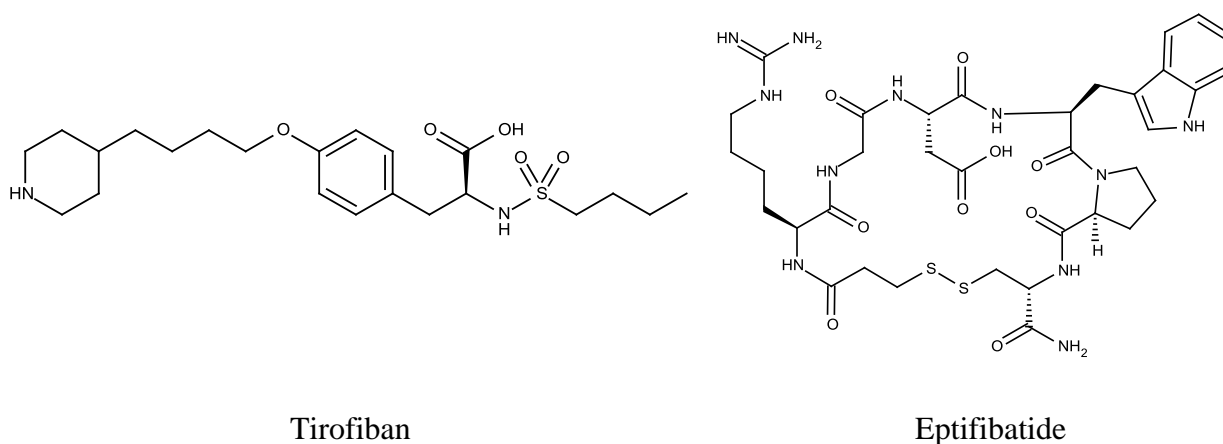


its study was stopped prematurely [120]. The second antagonist is picotamide, as well as terutroban inhibits TP, but also inhibits thromboxane synthase at equivalent concentration [121].



**Figure 10.** Antagonists of TP receptor

The last group, antagonists of  $\alpha_{IIb}\beta_3$ , is represented by three commercial drugs (Figure 11): abciximab, eptifibatid, tirofiban. Abciximab is a non-competitive antagonist of  $\alpha_{IIb}\beta_3$ . It is the humanised chimeric Fab-fragment of the monoclonal mouse antibody 7E3 [122]. Abciximab also interacts with two other integrins  $\alpha_v\beta_3$  [123] and  $\alpha_M\beta_2$  [124]. Eptifibatid [125] and tirofiban [126-127] are small molecules mimicking the RGD sequence of fibrinogen and thus block the binding site. Unlike abciximab, they are competitive antagonist of  $\alpha_{IIb}\beta_3$ . All these drugs are pretty expensive [128] and have some severe side effects like thrombocytopenia [129-131].



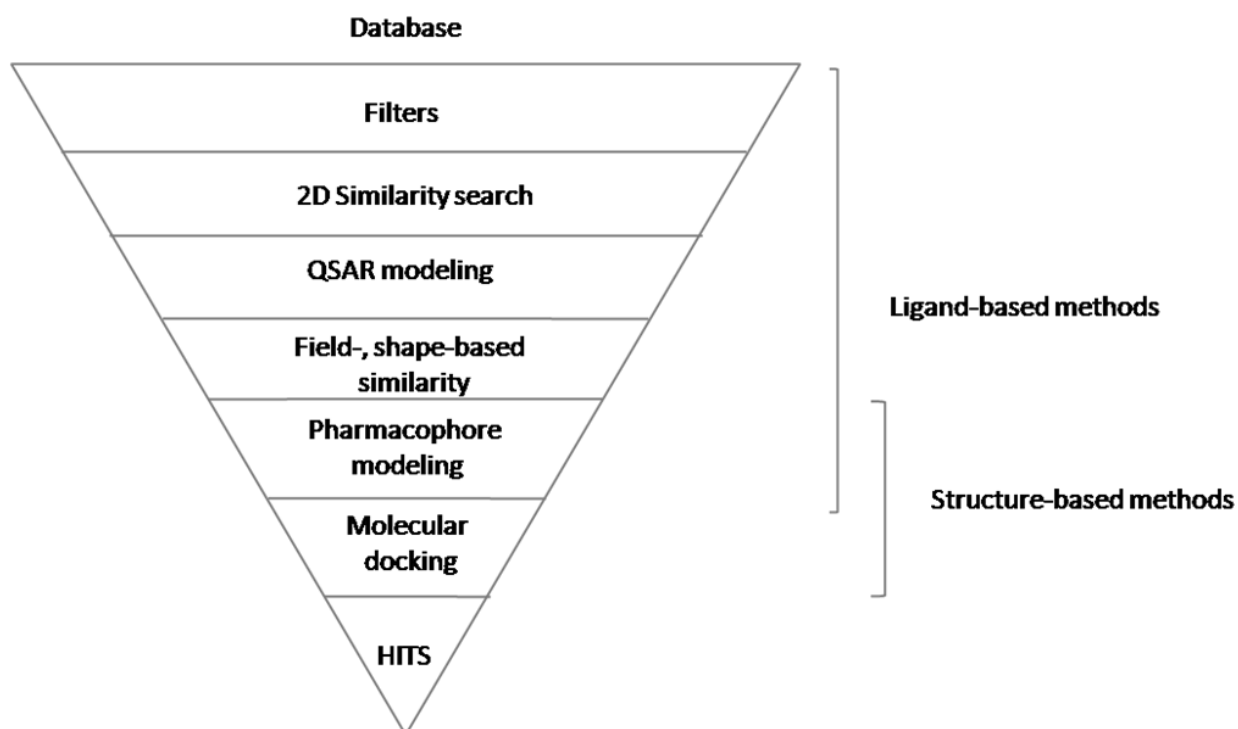
**Figure 11.** Antagonists of  $\alpha_{IIb}\beta_3$  receptor

## *Chapter 2*

# **Used methodologies and their applications**

### 3. Virtual screening funnel

Virtual screening is the computational or *in silico* analogue of biological screening. The aim of virtual screening is to score, rank and/or filter a set of structures using one or more computational procedures. Virtual screening can be used, for example, to help to decide which compounds to synthesize and which ones to purchase from a library. It may also be used for analysis of the results of experimental high-throughput screening data [132]. There exist many tools available to perform virtual screening (Figure 12). They can be categorized as being ligand-based or structure-based. The ligand-based methods use information provided by a compound or set of compounds, which are known to bind to the desired target, in order to identify other compounds in the corporate database or external databases with similar properties. They include such methods as filters (e.g., Lipinski's "rules of five"), similarity search, QSAR, pharmacophore modeling, 3D shape matching, etc. The structure-based methods can be used only in the case when the information about the target protein is available. They involve pharmacophore modeling and molecular docking. [133]



**Figure 12.** General scheme of virtual screening funnel

In this work we used both ligand-based (pharmacophore models, QSAR models, molecular fields and shapes comparison models) and structure-based (pharmacophore models, molecular docking) modeling methods.



### 3.1. QSAR approach

The QSAR approach can be generally described as an application of data analysis methods and statistics to developing models that could accurately predict biological activities or properties of compounds based on their structures.

The development of a QSAR model requires three components: (1) a dataset providing both chemical structures of the compounds and their experimental biological activity or property values; (2) molecular descriptors encoding chemical structures; and (3) machine-learning methods to obtain and validate the relationship between structures (encoded by descriptors) and activities [134].

#### 3.1.1. Molecular descriptors

The QSAR methods differ mainly by the principles and levels of representation and the description of molecular structure. It varies from the one-dimensional (1D) to four-dimensional (4D) level [135]:

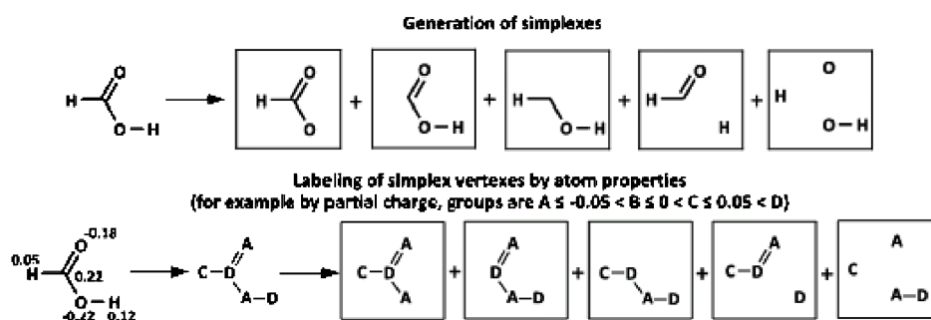
- One-dimensional molecular descriptors consider only the brute formula of a molecule (for example, alanine:  $C_3H_7NO_2$ ).
- Two-dimensional ones describe the information stored in the molecular graph, i.e., they reflect only the topology of the molecule.
- Three-dimensional molecular descriptors depend on 3D spatial coordinates of atoms and characterize one conformer only.
- Four-dimensional ones are similar to 3D molecular descriptors with the difference that they consider the structural information for a set of conformers.

Three different fragmental approaches for representation of molecular structure at 2D level have been used: Simplex representation (SiRMS) [136-138] and two types of ISIDA descriptors – Substructure molecular fragments (SMF) [139] and fuzzy pH-dependent pharmacophore triplets (FPT) [140-141].

##### 3.1.1.1. Simplex representation of molecular structure

Two-dimensional (2D) simplexes [137-138, 142] are four-atom fragments with fixed composition and topology. Simplexes are called “bounded” if all vertices are connected (Figure 13). The descriptor vector is defined as the number of occurrences of each simplex in a molecule.

Besides atom types, different physico-chemical characteristics of atoms can be used for atom labeling in simplexes, e.g. partial charge, lipophilicity, refraction and donor/acceptor propensity in hydrogen-bond formation (Figure 13). For atom properties the binning procedure are used to transform real values to four categories corresponding to their: (i) partial charge  $A \leq -0.05 < B \leq 0 < C \leq 0.05 < D$ , (ii) lipophilicity  $A \leq -0.5 < B \leq 0 < C \leq 0.5 < D$ , (iii) refraction  $A \leq 1.5 < B \leq 3 < C \leq 8 < D$ , and (iv) electronegativity  $A \leq 2.19 < B \leq 2.5 < C \leq 3 < D$ . Three characteristics of atom H-bond formation ability were specified A (acceptor of hydrogen in H-bond), D (donor of hydrogen in H-bond), and I (indifferent atom).

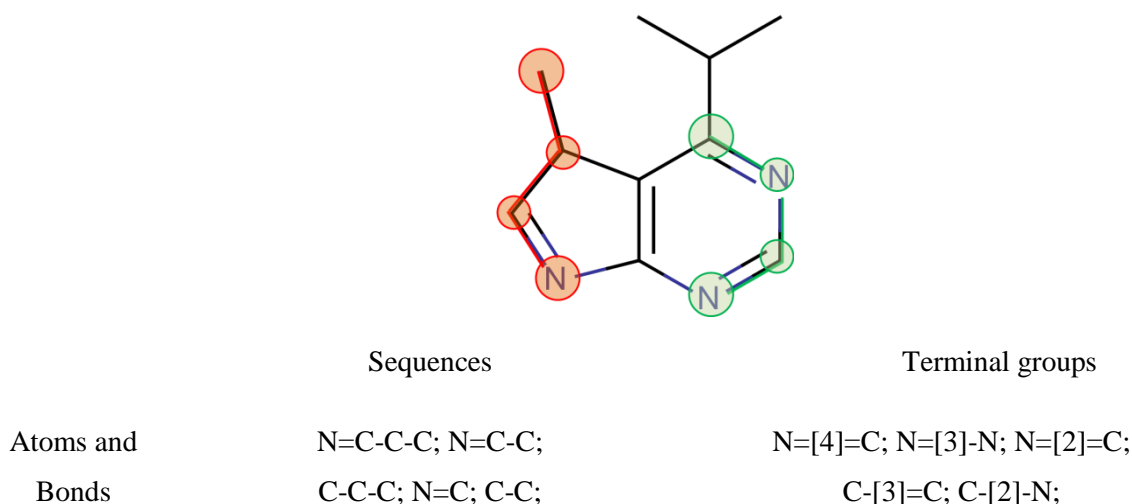


**Figure 13.** Example of generation of simplex descriptors.

### 3.1.1.2. ISIDA substructure molecular fragments

The Substructure molecular fragment method [139, 143] was implemented in ISIDA/QSPR software and is used to generate molecular fragments (Figure 14). Several different types of molecular subgraphs such as atom/bond “sequences”, “augmented” atoms and bonds and “terminal groups” are considered.

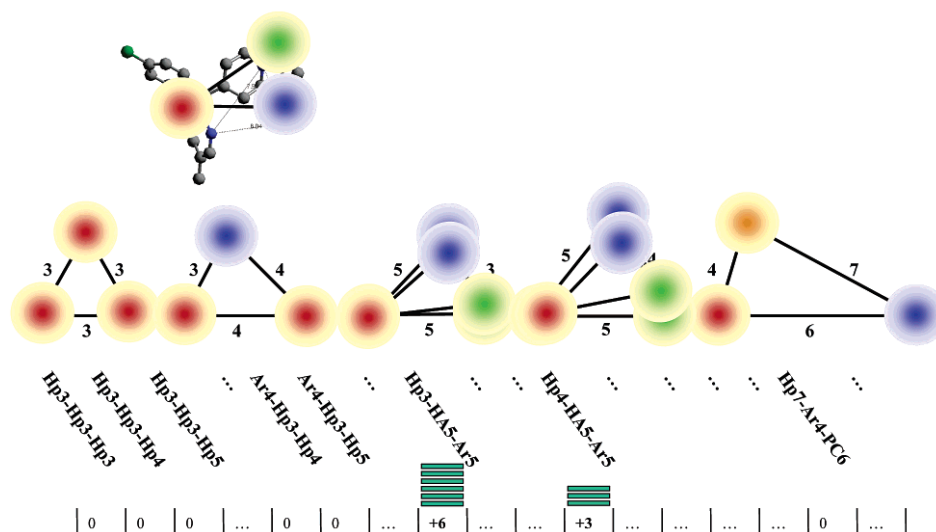
The sequences are represented by consecutive set of atoms linked by chemical bonds, where either atom types (C, N, O, ...) or bond types (single, double, ...) or both of them are considered explicitly. Only shortest paths for each pairs of atoms were used. The terminal groups correspond to the sequences with the shortest paths but defined by length (the number of atoms in square brackets, for example  $N=[4]=C$ ) and explicit identification of beginning atom and bond and ending bond and atom [144]. For searching the shortest paths, the Floyd algorithm [145] is used. For each type of sequences, the minimal ( $n_{min}$ ) and maximal ( $n_{max}$ ) number of constituent atoms is defined. For the given combination  $n_{min}$  and  $n_{max}$ , all intermediate shortest paths with  $n$  atoms ( $n_{min} < n < n_{max}$ ) are also generated. Each type of fragment is considered as descriptors whereas its occurrence is the descriptor’s value.



**Figure 14.** ISIDA fragment descriptors. Two classes of substructural fragments ( $2 < n < 4$ ): atom/bond sequences and terminal groups.

### 3.1.1.3. ISIDA fuzzy pH-dependent pharmacophore triplets

A basis set of reference pharmacophore triplets is chosen, enumerating all possible combinations of pharmacophore features (H-donor, H-acceptor, aromatic, hydrophobic, anion, cation) of the triplet corners, with defined minimal and maximal edge length. Next, all triplets of features represented in a molecule are analyzed, following a protonation state-dependent pharmacophore typing of the atoms, using shortest-path topological interatomic distances as actual edge lengths. Molecular triplets are then mapped onto basis triplets, using fuzzy logic (each molecular triplet may contribute to the population levels of several similar basis triplets, by increments directly related to their degree of similarity). Total population levels of basis triplets form a sparse vector (Figure 15) [140].



**Figure 15.** Example of fuzzy pH-dependent pharmacophore triplet vector in which every element stands for occurrence count (fuzzy) of given basis triplets.[140]

### 3.1.2. Machine-learning methods

The role of a machine-learning method is to establish an empirical relationships of the form  $P_i = \Phi(D_1, D_2, \dots, D_n)$ , where  $P_i$  are biological activities (or other properties of interest) of molecules,  $D_1, D_2, \dots, D_n$  are calculated (or, sometimes, experimentally measured) molecular descriptors, and  $\Phi$  is the empirically established mathematical transformation that should be applied to descriptors in order to estimate the property values for the given molecule [146].

The most widely used methods include multiple linear regression (MLR), partial least squares (PLS), artificial neural networks (ANNs), k-Nearest Neighbor (kNN) and Random Forest (RF) methods. In this work only the latter approach was used.

#### 3.1.2.1. Random Forest

Random forest method [147] (implemented in the CF software [148]) was used for QSAR modeling. RF is an effective statistical technique which is appropriate for analysis of large databases [149-152]. The main features of RF are listed below:

- a) it is possible to analyse compounds with different mechanism of action within one dataset;
- b) there is no need to pre-select descriptors;
- c) the method has its own reliable procedure for the estimation of model quality and internal measure of its predictive ability;
- d) obtained models are tolerant to “noise” in experimental data.

RF is an ensemble of single decision trees built by a Classification and Regression Trees algorithm (CART) [153]. Each tree has been grown according to the following rules:

1. Bootstrap samples which will be used as a training set for the current tree are produced from the whole training set of  $N$  compounds. About one-third of the compounds which aren't in the current training set are placed in the out-of bag (OOB) set. This is used to get an unbiased estimate of the model error and variable importance.

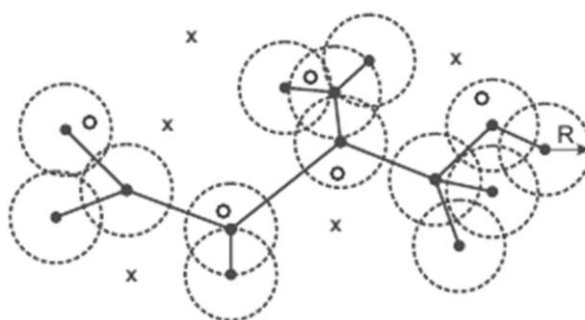
2. The best split among the  $m$  randomly selected descriptors taken from the whole set of  $M$  ones in each node is chosen. The value of  $m$  is the only tuning parameter for which RF models are sensitive.

3. Each tree is grown to the largest possible extent. There is no pruning.

RF possesses its own reliable statistical characteristics, which is used for validation and model selection. The major criterion for estimation of internal predictive ability of the RF models and model selection is the prediction performance for the out-of-bag set.

### 3.1.2.2. Applicability domain

The minimum spanning tree applicability domain (AD) approach [149, 152] implemented in CF software was used. The tree is built in the space of decision tree predictions for the given RF model using Kruskal's algorithm [154]. Then average distance ( $d_{av}$ ) and root-mean-square deviation ( $\sigma$ ) among all minimum spanning tree edges are calculated. Substantially, such distance is the characteristic of average density of molecules distribution in the considered space. If any of external set molecules is situated on the distance bigger than  $d_{av}+3\sigma$  from the nearest training set point, it means that this external set molecule is situated outside of AD (Figure 16).



**Figure 16.** Scheme of local applicability domain (AD) approach based on minimum spanning tree. ● – training set molecules, ○ – test set compounds within AD, × - test set compounds outside AD, radius of each sphere is  $R = d_{av} + 3\sigma$ .

To obtain more accurate predictions the consensus approach, which has proven itself in other studies [155], was applied. The consensus approach consists in averaging of the predictions from all individual models. To estimate the applicability domain for the consensus models the standard deviation of the prediction, obtained from the ensemble of models, was used. The general idea is that if different models yield significantly different predictions for a particular compound, then the prediction for this compound is more likely to be unreliable.

### 3.1.2.3. Criteria for estimation of the performance of QSAR models

#### 3.1.2.3.1. Regression models

The predictive ability of obtained regression models is estimated by coefficient of determination ( $R^2$ ) and root mean-squared error (RMSE):

$$R^2_{cv/oob} = 1 - \frac{\sum_{i=1}^n (y_{pred,i} - y_{exp,i})^2}{\sum_{i=1}^n (y_{exp,i} - \bar{y}_{training})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_{pred,i} - y_{exp,i})^2}, \quad (2)$$

where  $n$  – number of compounds in an external set;  $y_{exp,i}$  – observed activity value of  $i$ -th compound in an external set;  $y_{pred,i}$  – predicted activity value of  $i$ -th compounds in an external set;  $\bar{y}_{training}$  – mean activity value for compounds of a training set.

### 3.1.2.3.2. Classification models

There are several ways of evaluating the performance of classification models. Measures of the quality of classification is built from a confusion matrix, shown on Table 1, where TP are number of true positive, FP - false positive, FN - false negative, and TN - true negative.

Table 1. A confusion matrix for binary classification

Class \ Recognized	as Positive	as Negative
Positive	TP	FN
Negative	FP	TN

Among the various evaluation criteria, the measures which were used in the current work are precision, recall, sensitivity, specificity, ROC curve and AUC.

- I. Precision is defined as the proportion of the true positives against all the positive results (both true positives and false positives)

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

- II. Recall, (or Sensitivity) measures the proportion of actual positives which are correctly identified as such.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

- III. Specificity measures the proportion of negatives which are correctly identified as such.

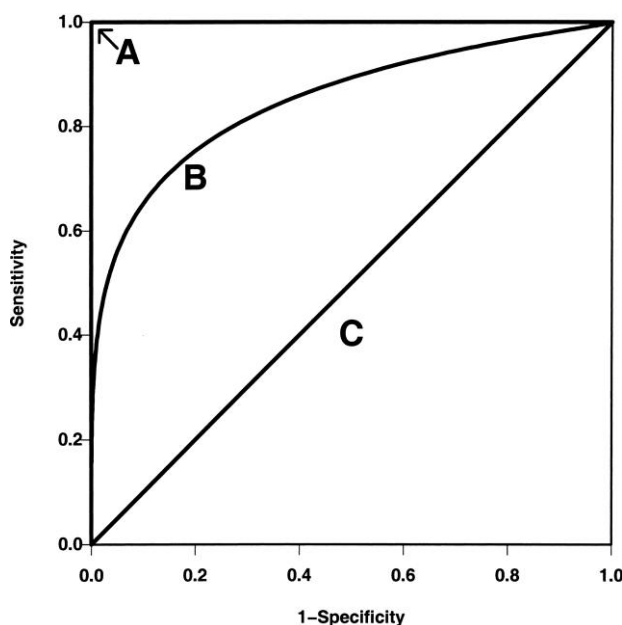
$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

IV. Balanced accuracy assesses the overall effectiveness of the classifier

$$\text{Balanced Accuracy (BA)} = \frac{0.5 \cdot \text{TP}}{\text{TP} + \text{FN}} + \frac{0.5 \cdot \text{TN}}{\text{TN} + \text{FP}} \quad (6)$$

### 3.1.2.3.3. The ROC curve (Receiving Operating Characteristics)

The ROC curves (Figure 17) serve to evaluate and to compare the developed models. On the curve, the performance of a binary classifier varies as function of the discrimination threshold; namely, it plots the fraction of true positives out of the positives (TPR = true positive rate) vs. the fraction of false positives out of the negatives (FPR = false positive rate), at various threshold settings. TPR is also known as sensitivity (also called recall), and FPR is one minus the specificity or true negative rate. It is possible to calculate the parameter from the obtained curve, called AUC (the area under the curve). It defines the probability of a model to rank active molecules higher than expectable from blind chance [156].



**Figure 17.** Three hypothetical ROC curves representing the accuracy: the perfect classification, curve A (AUC=1), a typical ROC curve B (AUC=0.85), and a diagonal line C corresponding to random chance (AUC=0.5).

### 3.1.2.4. Validation of QSAR models

Model validation implies quantitative assessment of model robustness and its predictive power. In this work the validation of the obtained models was estimated by a 5-fold external cross-validation procedure [157-158]. In this procedure, the entire modelling data set is divided in 5 non-overlapping pairs of training and test sets. Each training set covers  $\frac{4}{5}$  of the data set,

while the related test set covers the remaining  $\frac{1}{5}$ . Since each molecule belongs to only one of 5 test sets, all molecules from the initial data set are predicted. In such a way, one can avoid an ambiguity linked to selection of one only test set.

### 3.1.3. ADME/Tox assessments of novel compounds

As an additional filter in the virtual screening systems some ADME/Tox properties, such as mutagenicity, and solubility were estimated using previously developed QSAR models. The classification 2D QSAR model [155] based on the Ames mutagenicity dataset [159] was applied for prediction of mutagenicity of novel compounds. Both models were developed using Random Forest method and based on the simplex representation of molecular structure.

The aqueous solubility of screened compounds was predicted by two 2D QSAR models in order to make average consensus prediction. The former is based on ISIDA descriptors and developed using Multiple Linear Regression method [160] and the latter was developed using Random Forest method in combination with simplex descriptors [161].

### 3.1.4. PASS

PASS (Prediction of Activity Spectra for Substances) is a tool for evaluation of biological responses of chemical compounds. PASS predicts about 2500 kinds of biological activities based on the structural formula of chemical compound. Biological types include main and side pharmacological effects (e.g. antihypertensive, hepatoprotective, sedative, etc.), mechanisms of action (5-hydroxytryptamine agonist, acetylcholinesterase inhibitor, etc.), specific toxicities (mutagenicity, carcinogenicity, teratogenicity, etc.) and names of metabolizing enzymes (CYP1A substrate, CYP3A4 substrate, etc.) [162]. The PASS approach is based on the analysis of structure-activity relationships (SAR) for the training set currently including about 60,000 drugs, drug-candidates, leads and toxic compounds whose biological activity was determined experimentally [163]. These SARs are obtained during the training procedure and they are stored in the knowledge base. Results of prediction are a list of activity types, with the probability of presence ( $P_a$ ) and absence ( $P_i$ ) for each particular activity. By definition the probabilities  $P_a$  and  $P_i$  can be also interpreted as the measures of belonging to fuzzy subsets of “active” and “inactive” compounds, or as the probabilities of the 1<sup>st</sup> and 2<sup>nd</sup> kinds of errors of prediction. By default,  $P_a > P_i$  value is used as a threshold that provides the mean accuracy of prediction about 90 % in leave-one-out cross-validation for training set [162].



### 3.2. Molecular shape-similarity

A set of different ligands that bind to the same receptor site, given rise to a similar pharmacological response, are expected to possess not only similar molecular fields but similar molecular shapes also [164]. There are two types of basic algorithms for analysis the molecular shape: superposition-free and superposition-based. The former calculates shape similarity based on rotation-and-translation invariants that are computed from different representations of molecular objects, and thus, it does not depend on the orientation or alignment of the two molecular objects. Zauhar's shape signatures[165] and the more recent USR method [166-167] belong to this category. The second calculates shape-matching quality after an optimal superposition of the two objects [168]. The popular ROCS program (Rapid Overlay of Chemical Structures) [169] of the OpenEye software belongs to the latter category.

#### 3.2.1. ROCS. Rapid Overlay of Chemical Structures

ROCS performs shape-based overlays of conformers of a candidate molecule to a query molecule. There are several possible alternatives of query creation. This approach is dependent on the conformation of the query molecule. If the structure of protein-ligand complex is available, then the information about the conformation of the reference ligand can be taken from there. When the "bioactive" conformation of the query molecule is unknown the lowest energy conformer, which is generated by Omega tool, is used as a reference structure. If there are several known active ligands, it would be preferable to use a query which is an alignment of active ligands in their "bioactive" or low energy conformations. ROCS also allows colored shape match, which is identical to field matching.

The overlays can be performed very quickly based on a representation of the molecules as atom-centered Gaussian functions. The algorithm searches and assesses the maximal overlap of the volumes of two molecules. Therefore, the query and search molecules are virtually independent of the atom types and bonding patterns present in them [170]. The shape Tanimoto (see eq. 7) is used to measure the overlap of the shapes of two molecules:

$$T_{A,B} = \frac{O_{A,B}}{(O_{A,A} + O_{B,B} - O_{A,B})}, \quad (7)$$

where  $O_{A,B}$  – overlap between molecule A and B;  $O_{A,A}$ ,  $O_{B,B}$  – self-volume overlaps of each molecule.

Shape Tanimoto is 1.0 if two shapes are identical, and 0.0 if completely different. Two shapes are never completely different, i.e., have zero overlap, but shapes may be identical for different molecules [170].

In addition to shape-alignments ROCS optionally considers chemical features alignment (“color” alignment). The chemical features include hydrogen-bond donors, hydrogen-bond acceptors, hydrophobic centers, anions, cations, and aromatic rings, which are represented as Gaussians and displayed as colored spheres. The color force field, which can be represented as an Implicit Mills Dean or Explicit Mills Dean force fields[171], is used to measure chemical similarity between the query and the database molecule and to refine shape-based overlays. The chemical similarity (Color Tanimoto) is measured in the same way as for the shapes (eq. 7). Both similarity metrics for shapes and for chemical features can be used as separately as well as in combination. Their sum gives new metrics called TanimotoCombo. [172]

The validation of the generated molecular shape model is carried out in the same way as for the molecular field model.

### 3.3. Molecular field similarity

It is known that the compounds of quite different molecular structures may bind to the same protein site and produce an identical biological response. As the chemical structures of these active compounds are different, this suggests that using an atom and bond connectivity depiction of molecules may not provide the best insight into predicting activity across different chemotypes. The atoms and bonds describe the skeleton of a molecule, but give little information on the surface and electrostatic properties that govern the interaction of that molecule with other molecules. It is assumed that molecular field calculations can provide a clearer view on molecular similarity because the molecular fields depict the outside surface of a molecule and better represent the ligand as it is ‘seen’ by a protein active site[173].

Molecular fields can be calculated for any molecule of known 3D structure. They describe the spatial variation of the interaction energy between a molecular target and a chosen probe. The target may be a macromolecule or a low molecular weight compound or a molecular complex. The probe may be a molecule or a fragment of a molecule. At its simplest, the probe is a unit positive charge representing a proton. Most probes are spheres parameterized to represent a specific atom or ion type (e.g. a carbonyl carbon atom, a methyl CH<sub>3</sub> group). Non-spherical probes containing more than one non-hydrogen atom may also be used (e.g. a carboxylate group) [174].

Most commonly used representations of molecular fields are based on grids (GRID, CoMFA, Volsurf), but they produce too many data points for fast processing, or they are restricted in their accuracy by low grid resolution. Grids also suffer from "gauge variance": if a molecule is rotated marginally within the grid, the values of those grid points change depending

on the resolution of the grid and the degree of rotation [173]. Describing fields in terms of Gaussians [175] (CoMSIA) is elegant and quicker than grids, but works best when describing fields which can be approximated by overlapping spheres (such as molecular volumes) and is less appropriate for probe-interaction energy fields. Another field similarity approach was implemented by Cresset. This approach is faster than earlier described methods and thus it is more appropriate for virtual screening.

### 3.3.1. Cresset FieldAlign approach

Here, molecular fields are represented as a set of molecular field extrema [176-178] to overcome problems of other molecular fields modeling methods. The molecular field extrema are local minimum and maximum field points [179], which are derived from calculating the interaction energy of a “probe” with the target molecule [173]. One hundred twenty points are generated at regular intervals over a slightly diminished solvent accessible surface of each atom of the molecule. A probe atom is placed at each point and its interaction energy with the molecule is optimized with a simplex method that finds a true extremal point and avoids grid techniques. The probe is given the van der Waals parameters of oxygen, and its charge is adjusted according to the field type. Extrema with very small values of the interaction energy are insignificant and filtered out [180]. The generated “field points” are represented as color-coded spheres, whose diameters are proportional to the magnitude of the extrema. There are four molecular fields implemented in Cresset software[179]:

- Positive electrostatic (coloured red);
- Negative electrostatic (coloured blue);
- Van der Waals attractive, that is “steric” (coloured yellow);
- Hydrophobic (coloured orange).

In field similarity-driven virtual screening, the reference molecule is chosen and all screened molecules are compared to it, in terms of field pattern matches (see further on). As a consequence obtained results of the comparison are sensitive to the conformation of the reference compound. Thus, the same principles as for the ROCS approach can be applied to choose the correct (“bioactive”) conformation of the active reference molecule. Alignments to a single reference molecule generally work well. If two or more active compounds are known then it might be better to align them and derive their common molecular field, which will be used as a reference in the screening.

For screened molecules, conformers are generated by means of XED force field and one conformer with the best alignment to the reference compound is chosen for each molecule. A

Dice field similarity metric is used to score alignments [180]. To determine the similarity of a pair of conformations of two molecules, A and B, the field points of A ( $fp_A$ ) are used as sampling points for the entire potentials of B:

$$E_{A \rightarrow B} = \sum_{fp_A} \sqrt{\text{size}(fp_A) \cdot F_B(\text{position}[fp_A])}, \quad (8)$$

where  $E_{A \rightarrow B}$  is a similarity of a pair of conformations of two molecules, A and B;  $fp_A, fp_B$  are field points for molecule A and B, respectively;  $F_B(x)$  is the value of the appropriate field on B at *position* x, and the sum is over all field points on A; the *size* of field point is determined from the depth of the potential energy.

This score is asymmetric, so the calculation is repeated for the field points of B sampling into the actual field of A and the two are averaged to give a symmetric score  $E_{AB}$ . The score can then be normalized to give a Dice field similarity metric and maximizing this metric between two conformations A and B gives both the best overlay (in terms of field similarity) and a single field similarity value for the two conformations.

$$S_{AB} = \frac{2E_{AB}}{E_{AA} + E_{BB}}, \quad (9)$$

where  $S_{AB}$  is the field similarity over all possible relative orientations;  $E_{AB}$  is average value between  $E_{A \rightarrow B}$  and  $E_{B \rightarrow A}$ ;  $E_{AA}$  is the energy of molecule A fitted to itself;  $E_{BB}$  is the energy of B fitted to itself.

As the energies are analytically recalculated, the entire ‘true’ field is used in the calculation, and the potential well widths are implicitly included. However, only a few field values need to be calculated at any given orientation so the technique is fast enough to be applied to large structures and many conformations in reasonable computing time. The fields of each molecule are sampled at only a few places, but the use of the field extrema of the other molecule as the sampling points ensures that the fields are sampled at biologically relevant points. It is also worth noting that this calculation is gauge-invariant and hence avoids many of the issues involved in grid-based similarity metrics [180].

One of the advantages of the Cresset software is a possibility to design new compounds. Since different molecular fragments can produce the similar fields, much greater levels of structural diversity can be inserted into design process of new compounds. Once the new compound is designed based on the reference one, it can be seen how the modifications affects the molecular fields, and activity of new compounds can be interpreted in terms of field similarity.

The other advantage of the Cresset’s molecular fields over pharmacophore approach is that there is no need to identify specific pharmacophore groups. For example, usage of pre-

defined set of groups can miss CH group, which can act as an H-bond donor due to the electronic influence of adjacent groups, as it has been identified in some kinase inhibitors [173].

For the validation of the generated molecular fields the dataset of ligands with known activity are collected. The validation dataset runs against the created query and AUC value is calculated to estimate how well the query discriminates the active and inactive compounds. The validated query then can be used in the virtual screening.

### 3.4. Pharmacophore modeling

A pharmacophore model is an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response [181]. The “classical” pharmacophore features are H-bond acceptors and donors, charged or ionisable groups, hydrophobic residue and aromatic rings. The use of such features is a natural extension of the concept of *bioisosterism*, which recognizes that certain functional groups have similar biological, chemical and physical properties.

The spatial relationships between the features in a 3D pharmacophore model can be specified as distances or distance ranges or by defining the (*xyz*) locations of the features together with some distance tolerance (typically as a spherical tolerance region). [132]

There are different possibilities to derive pharmacophore models: based on the available three-dimensional structure of a ligand-protein complex (structure-based modeling) or based on the structural information of active compounds only (ligand-based modeling) [182].

Catalyst [183], MOE [184], Phase [185] and LigandScout [186] are popular pharmacophore modeling tools applied for lead discovery. Although their pharmacophore-based screening algorithms differ in detail, the common rationale of all these screening tools is the use of a pharmacophore model as query for virtual screening of databases consisting of 3D structures of small molecules [187]. The single database molecules are usually represented by a set of conformers that include the bioactive geometry adopted during the interaction with the target protein. All conformers of screened compounds that match a user-specified number of model features are stored in a hit list [188]. Depending on the selectivity of the pharmacophore model, such a virtual screening of chemical databases consisting of millions of small molecules can result in tens to thousands of hits [189]. For ranking the molecules and to determine the quality of matching between the pharmacophore model and each molecule of the virtual screening hit list a score is calculated. [190]

### 3.4.1. LigandScout

LigandScout was used in the current work. LigandScout significantly differs from established methods in translating the 3D conformer into a set of pharmacophore features instead of directly using atom positions to align the compound to the model. This concept allows an intuitive comparison of molecules in terms of bioisosterism and expands the range of possible applications for this algorithm. Thus, the algorithm can be used not only for pharmacophore-based virtual screening but for common-feature elucidation and 3D-QSAR studies [191-192]. Besides ligand-based models, the automatic generation of the structure-based models featuring exclusion volumes is implemented in LigandScout [186].

The main difference of LigandScout from other packages (Catalyst, MOE and Phase) is the fast alignment algorithm due to the efficiency of the implementation and the advanced geometric similarity measure for the chemical features. In this algorithm, the first step concerns the generation of the 3D pharmacophore features identified for each database conformer. Then, the algorithm creates for each feature type a set of inter-feature distances. The distance sets created for the pharmacophore model and for the conformer pharmacophore features are then compared in a pairwise manner. In order to perform a pair assignment, the so-called Hungarian matching algorithm is executed. Finally, the feature distances between model and conformer are minimized using Kabsch alignment algorithm [186, 192].

For estimation of alignment quality, four different scoring functions are implemented in LigandScout. The pharmacophore fit score is a simple geometric scoring function, which favors solutions with a high number of geometric matched feature pairs and penalizes those with higher RMS deviations among those feature pairs. The overlap of atom van der Waals spheres is characterized by the atom sphere overlap score, whereas an overlap of Gaussian function representation of molecular volume is measured to obtain Gaussian shape similarity score. The last one is a combo score of the first two scores and named pharmacophore fit and atom overlap score [193]. In this study the default scoring function pharmacophore fit score had been used (see eqs. 10 and 11).

$$S_{RMS} = 9 - 3 \cdot \min(RMS_{FP}, 3); \quad (10)$$

$$S_{FCR} = c \cdot N_{MFP} + S_{RMS}; \quad (11)$$

where  $S_{RMS}$  is the matched feature pair RMSD score in the range [0,9];

$RMS_{FP}$  is the RMSD of the matched feature pair distances;

$S_{FCR}$  is the feature count/RMS distance score;

$c$  is a weighting factor for the number of matched feature pairs (currently 10.0)

$N_{MFP}$  is the number of geometrically matched feature pairs.

Two types of ligand-based model can be derived due to different generation process. The model generation is a pair-wise process. In each step one pharmacophore for two molecules is created. It can be done by including to the model only common features of a ligand set (*Shared feature pharmacophore*) or by taking all features of a ligand set (*Merged feature pharmacophore*). In the latter case each feature is scored and those are removed that do not match all input molecules.

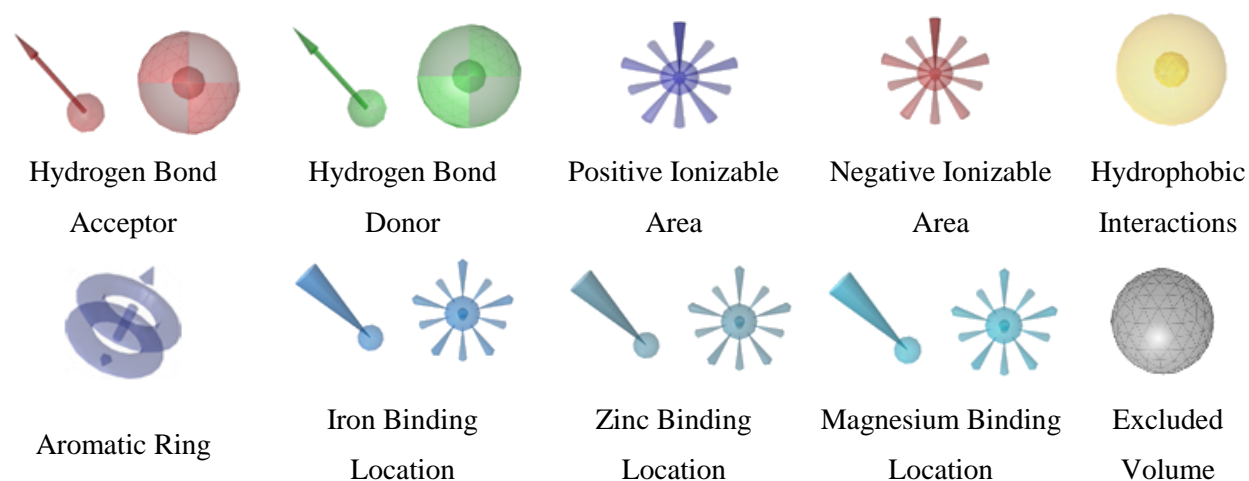
In ligand-based mode of LigandScout there is an option to cluster the active compounds that helps to find those having similar patterns of interactions with a target protein. The conformers are generated for each compound. Afterwards the compounds are clustered according to the RMSD values calculated between centers of corresponding pharmacophore features of selected conformers of each compound.

To perform consensus virtual screening on several pharmacophore models the boolean expressions can be specified to combine output of these models, e.g. “(3<sup>rd</sup> or 2<sup>nd</sup>) and 1<sup>st</sup>”.

For the validation of the generated pharmacophore models the dataset of ligands with known activity can be used. The validation dataset runs against the created pharmacophore model and statistical indices, (see eqs. 3 and 4) showing how well the model discriminates the active and inactive compound, is calculated. The validated model then can be used in the virtual screening.

In the LigandScout the "classical" pharmacophore features implemented are shown in Figure 18. Each feature of the generated pharmacophore can be marked as *optional*, what means that it won't be mandatory to find a valid alignment. During the screening process, a molecule is still valid hit when the feature marked as optional is not matched.

In order to adjust the specificity of the pharmacophore model one can increase or decrease the tolerance of the selected feature, in other words change the volume of sphere of the selected feature.



**Figure 18.** Pharmacophore features definition in LigandScout.

### 3.5. Omega

In the current study, the Omega tool of OpenEye software was used for generation of stereoisomers and conformers while using LigandScout and vROCS programs. It is a widely used rule-based generator which rapidly produces conformational ensembles of small molecules.

Omega contains a precomputed fragment database which contains one or more 3D conformations for every entry. This database is generated by fragmenting a very large collection of commercially available compounds into contiguous ring systems and small linear linkers. One or more 3D conformations for each fragment are generated by the following procedure: a distance bounds matrix is generated based on the connection table of the fragment; the distance bounds are augmented by volume constraints for chirality and planarity; the coordinates of each atom are randomly embedded in a Cartesian space and optimized to fulfill the bounds and constraints; the fragment is further refined with a modified version of the Merck molecular force field (MMFF94) [194], in which the electrostatic and attractive van der Waals terms are removed [195]. Each studied molecule is fragmented in the same manner as the fragment database and fragment conformations are drawn directly from the fragment database. If a molecular fragment was not found in the fragment database, the coordinates for the missing fragment are generated on-the-fly by a distance geometry method with a modified version of the MMFF94 force field [194].

In the subsequent step, torsion driving on the 3D structure assembled from these fragments is performed to generate a large initial “raw” conformer ensemble. This torsion driving is governed by a set of rules for torsion angles contained in a torsion library. In the final stage, this raw conformer ensemble produced from torsion driving is pruned by geometric diversity and conformational strain energy (calculated using the same modification to MMFF94 used to calculate fragment geometries) to produce the output conformer ensemble [196]. The generated conformers are ordered using a simple scoring function that eliminates conformers with internal clashes. Next, beginning with the lowest scoring conformer, all higher scoring conformers that are less than a user-defined RMSD to the lower scoring conformer are eliminated [195].

Hawkins *et al.* analyzed the failures of conformer generation with Omega tool. Their close analysis of the failures revealed that most of them were due to under-sampling the large low-energy conformational space accessible to molecules with eight or more rotatable bonds, and simple adjustment of the maximum number of conformations generated for these more flexible ligands prevented most of the failures[196].



### 3.6. Docking

Molecular docking is a widely-used computational tool for studying of ligand-receptor recognition. Its aims are prediction of a binding pose of a ligand and “binding free energy” of a ligand to a receptor. Determination of the binding pose and “binding free energy” is crucial for understanding the important ligand-receptor interactions and mechanism of action, thus valuable in the design of new drugs. In molecular docking thousands of possible poses of a ligand inside a receptor cavity are tried and evaluated; the pose with the lowest energy score is predicted as the “best match”, i.e. the binding pose [197].

A typical protein-ligand docking program consists of two essential components: sampling and scoring. Sampling refers to the generation of putative ligand binding orientations/conformations into a binding site of a protein and can be further divided into two aspects: ligand sampling and protein flexibility [197]. Identification of the putative orientations/conformations of the ligand in an active site of a receptor is challenging, as even relatively simple organic molecules can contain many conformational degrees of freedom. Sampling these degrees of freedom must be performed with sufficient accuracy to identify the conformation that best matches the receptor structure, and must be fast enough to perform the evaluation of thousands of compounds [198-199]. In reference to the second aspect of the sampling, ligand binding commonly induces protein conformational changes, which range from local rearrangements of side-chains to large domain motions. Due to the large size and many degrees of freedom of proteins, modeling of protein flexibility may be the most challenging issue in molecular docking [200-205].

Scoring is the prediction of the binding strength for individual ligand orientations/conformations with a physical or empirical energy function. Speed and accuracy are two important characteristics of a scoring function. Scoring functions can be classified into three main categories [199]: empirical scoring functions, knowledge-based potentials and force-field methods.

Empirical scoring functions include several terms describing properties known to be important in drug binding to construct a master equation for predicting binding affinity. These terms generally describe polar and apolar interactions, loss of ligand flexibility (entropy) and eventually also desolvation effects. The major disadvantage of the empirical scoring functions is the need for a training set to derive the weight factors of the individual energy terms. Force-field scoring functions are based on the non-bonded terms of a classical molecular mechanics force field (e.g. AMBER, CHARMM, etc.). The main drawback of force-field calculations is the omission of the entropic component of the binding free energy. Knowledge-based scoring

functions encode structural information gathered from protein-ligand X-ray coordinates into Helmholtz free interaction energies of protein-ligand atom pairs. The score is defined as the sum over all interatomic interactions of the protein-ligand complex [206]. The disadvantage is that their derivation is essentially based on information implicitly encoded in limited sets of protein-ligand complex structures [199].

In addition to mentioned problems, there is a problem of preparation of binding sites for docking simulation. The difficulties are caused by low resolution of crystallographic structures of proteins and their complexes, position of N and O atoms in side chains of asparagine and glutamine, correct tautomeric state of some amino acids, the appropriate position and orientation of water molecules, which can participate in protein–ligand interactions.[199, 207]

### **3.6.1. MOE, FlexX, PLANTS**

Three programs were used for ligand to protein docking in the framework of this study: MOE [208], FlexX [199] and PLANTS [209]. All of them have different sampling algorithms and scoring functions. In all three approaches the ligand is flexible and the protein is kept rigid, but in contrast to other two programs, PLANTS includes the flexibility of some protein's amino acid side chains [209-210]. Moreover they use different procedures of binding site preparation.

A prepared binding site in FlexX includes all protein atoms within a distance of up to 6.5 Å to 8 Å from any atom of the co-crystallized ligand. This distance is set by visual inspection such that the binding site is completely enclosed. The protonation states and orientation of mobile hydrogen atoms of side chains of amino acids in the protein-binding site are defined manually. Water molecules and metal ions which are known to play a critical role in ligand binding are kept. All these steps require some expert supervision.

FlexX is a fragment-based sampling tool using a systematic search algorithm [199]. The key step of the algorithm is the choice of the base fragment of the ligand, which should be responsible for principal interactions with target protein. The ligand is fragmented into components by severing at all acyclic single bonds [211]. Then FlexX automatically forms a set of alternative base fragments by selecting single components or their combinations. During the selection of the base fragment some parameters are taken into account: number of centers of tight interactions, the total number of the interaction centers and number of conformations of the fragment [212]. After identification of the base fragments, they are placed into the active site of the receptor. The complete ligand is constructed by adding the remaining components one after another. At each step of reconstruction a specified number of optimal partial solutions are selected for the next extension step [211].

For the ranking of the generated solutions FlexX uses an empirical modified Böhm scoring function, which estimates the free energy  $\Delta G$  of the protein-ligand complex (eq. 12). It includes such terms as entropic, hydrogen bonding, ionic, aromatic and lipophilic [213].

$$\Delta G = \Delta G_0 + \Delta G_{rot} \cdot N_{rot} + \Delta G_{hb} \sum_{neutralH-bond} f(\Delta R, \Delta \alpha) + \Delta G_{io} \sum_{ionicint} .f(\Delta R, \Delta \alpha) + \quad (12)$$

$$+ \Delta G_{aro} \sum_{aroint} .f(\Delta R, \Delta \alpha) + \Delta G_{lipo} \sum_{lipo.cont} .f^*(\Delta R),$$

where  $\Delta G$  designates different energetic terms,  $\Delta G_0$  – the energetic contribution which accounts for the reduction in translation and rotational entropy;  $\Delta G_{rot}$  – the energetic contribution which describes the loss of binding energy due to freezing of internal degrees of freedom in the ligand;  $N_{rot}$  – number of free rotatable bonds;  $\Delta G_{hb}$  – the energetic contribution from an ideal hydrogen bond;  $f(\Delta R, \Delta \alpha)$  – scaling function penalizing deviations from the ideal geometry;  $\Delta G_{io}$  – the energetic contribution from an unperturbed ionic interaction;  $\Delta G_{aro}$  – the energetic contribution from interactions of aromatic groups;  $\Delta G_{lipo}$  – the energetic contribution from lipophilic interactions;  $f^*(\Delta R)$  – scaling factor which accounts for contacts with “ideal” distance but penalize for close contacts.

The various contributions to the function are the following:

- a fixed term ( $\Delta G_0$ ),
- an entropic term  $\Delta G_{rot} \cdot N_{rot}$  taking into account the loss of entropy during ligand binding.
- contributions for matched and charged hydrogen bonds, salt bridges, aromatic and hydrophobic interactions. Each of these terms consists of a fixed contribution per interaction ( $\Delta G_{hb}$ , for example) multiplied by a penalty function  $f(\Delta R, \Delta \alpha)$ . The penalty functions are piecewise linear functions scaling the contribution of an interaction with respect to its optimal geometry [212], and implicitly penalizing clashes.

A binding site in PLANTS includes by default all protein atoms within 12Å distance from the geometric center of a ligand. A Structure Protonation and REcognition System (SPORES) is used for normalization of raw PDB structures of complexes and determination of tautomeric and protonation states of binding site amino acids. Water molecules and metal ions which are known to play a critical role in ligand binding are kept after manual inspection.

The PLANTS docking algorithm is based on one of stochastic optimization algorithms called *ant colony optimization* (ACO). ACO is inspired by the behavior of ants finding a shortest path between their nest and food source [210]. While walking, ants deposit *pheromones* to mark paths they follow. When they have to choose between several paths to follow, paths with higher pheromone concentration are chosen with higher probability. ACO algorithms use *virtual pheromones* to imitate this behavior. They are represented as numerical values that are associated with each possible solution component [209].

Two empirical scoring functions are implemented in PLANTS:  $PLANTS_{PLP}$  and  $PLANTS_{CHEMPLP}$  (eqs. 13 and 14) [209]. They are based on parts of already published scoring functions and force fields. The piecewise linear potential (PLP) scoring function is used in both cases to model steric complementarity of the protein and the ligand. In  $PLANTS_{CHEMPLP}$ , terms of GOLD's Chemscore implementation are used to introduce angle-dependent terms for hydrogen bonding and metal binding. The torsional potential from the Tripos force field together with a heavy-atom clash term is employed to account for intra-ligand interactions.

$$f_{PLANTS_{PLP}} = f_{plp} + f_{tors} + f_{clash} + C_{site} \quad (13)$$

$$f_{PLANTS_{CHEMPLP}} = f_{plp} + f_{hb} + f_{hb-ch} + f_{hb-CHO} + f_{met} + f_{met-coord} + f_{met-ch} + f_{met-coord-ch} + f_{clash} + f_{tors} + C_{site}, \quad (14)$$

where  $f_{plp}$  – the piecewise linear potential;  $f_{tors}$  – the torsional potential;  $f_{clash}$  – the empirical heavy-atom potential;  $C_{site}$  – quadratic potential;  $f_{hb}$  – potential for the donor-acceptor pairs;  $f_{hb-ch}$  – potential for the charged donor and charged acceptor pairs;  $f_{hb-CHO}$  – potential for the hydrogen bonding pairs containing an oxygen-acceptor;  $f_{met}$  – distance- and angle-dependent potential;  $f_{met-coord}$  – coordination polyhedra;  $f_{met-ch}$  – potential for charged acceptor atom, involved in a metal interaction;  $f_{met-coord-ch}$  – potential for charged acceptor atom, involved in a metal interaction.

The preparation of the binding site in MOE includes several steps. The protonation of the whole complex is performed followed by energy minimization with MMFF94 force field. As in the previous cases water molecules and metal ions which are known to play a critical role in ligand binding are kept in the binding site.

In MOE it is possible to use two different protocols: simulated annealing and tabu search to optimize the spatial contacts and the electrostatic interactions between the molecules [208]. Simulated annealing is based on the Monte Carlo method [214]. The following parameters control the search: number of cycles per run, iteration limit, and initial temperature. Tabu search performs a stochastic search maintaining a list of previously visited conformations that are forbidden to future moves. The new conformations are compared to the visited conformations using the root mean square deviation. The parameters that control the search process are number of steps per run, number of moves per step, and tabu list length [208].

Ligands poses are ranked by one of the implemented scoring functions based on different geometrical and energetic terms. For instance, the London  $\Delta G$  scoring function (eq. 15) estimates the free energy of binding of the ligand in a given pose by taking into account entropy, desolvation, hydrogen bonding, metal ligation and ligand flexibility [215].

$$\Delta G = c + E_{flex} + \sum_{h-bonds} C_{HB} f_{HB} + \sum_{m-lig} C_M f_M + \sum_{atoms\ i} \Delta D_i \quad (15)$$

where  $c$  - the average gain/loss of rotational and translational entropy;  $E_{flex}$  - the energy due to the loss of flexibility of the ligand (calculated from ligand topology only);  $f_{HB}$  - geometric imperfections of hydrogen bonds;  $C_{HB}$  - the energy of an ideal hydrogen bond;  $f_M$  - geometric imperfections of metal ligations;  $C_M$  - the energy of an ideal metal ligation;  $D_i$  - the desolvation energy of atom  $i$ .

The quality of the docking can be benchmarked by the RMSD between poses of crystallized and docked ligand obtained in re-docking and cross-docking procedures. Re-docking consists in docking of the ligand extracted from the X-ray structure of a ligand/protein complex back into this binding site. Cross-docking consists in docking of the ligand extracted from the X-ray structure of a ligand/protein complex into a slightly different shaped protein binding site of another complex or of the ligand-free protein. RMSD values lower than 2Å are considered satisfactory.

Also for the validation purpose of the docking study library containing active and inactive compounds is screened. The compounds are docked into a protein binding site and are ranked according to the docking score, with which ROC curve and AUC value are evaluated.

*Chapter 3.*

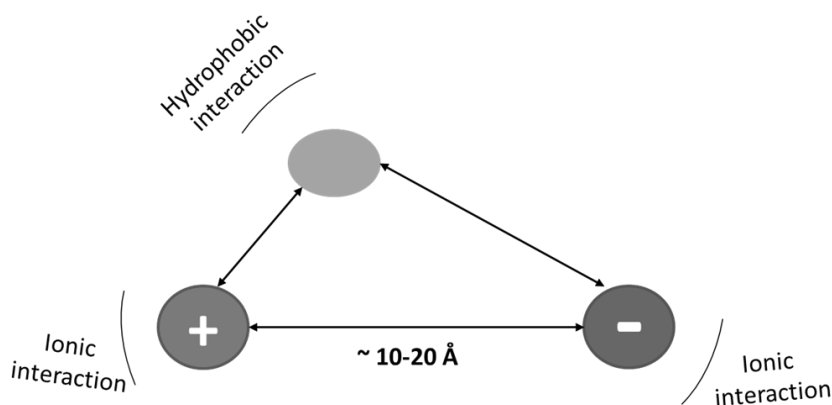
**RGD-peptidomimetics – antagonists of integrin  
 $\alpha_{IIb}\beta_3$  receptor**

Here, we give some information about previously performed modeling studies on RGD-peptidomimetics (Section 4) followed by the results obtained in this work (Section 5).

## 4. Previously reported models

### 4.1. Pharmacophore models

In the early works the pharmacophore models of fibrinogen antagonists were represented as a simple two-point model which consists of positive and negative charged groups with distance between them is within 10-20Å (approximating the distance between central carbons of the positively and negatively charged groups) [126, 216]. The known antithrombotic drug tirofiban was found using such two-point pharmacophore model. In further works other authors have added the hydrophobic feature to the existed two-point model for a more accurate model definition (Figure 19) [217-218].



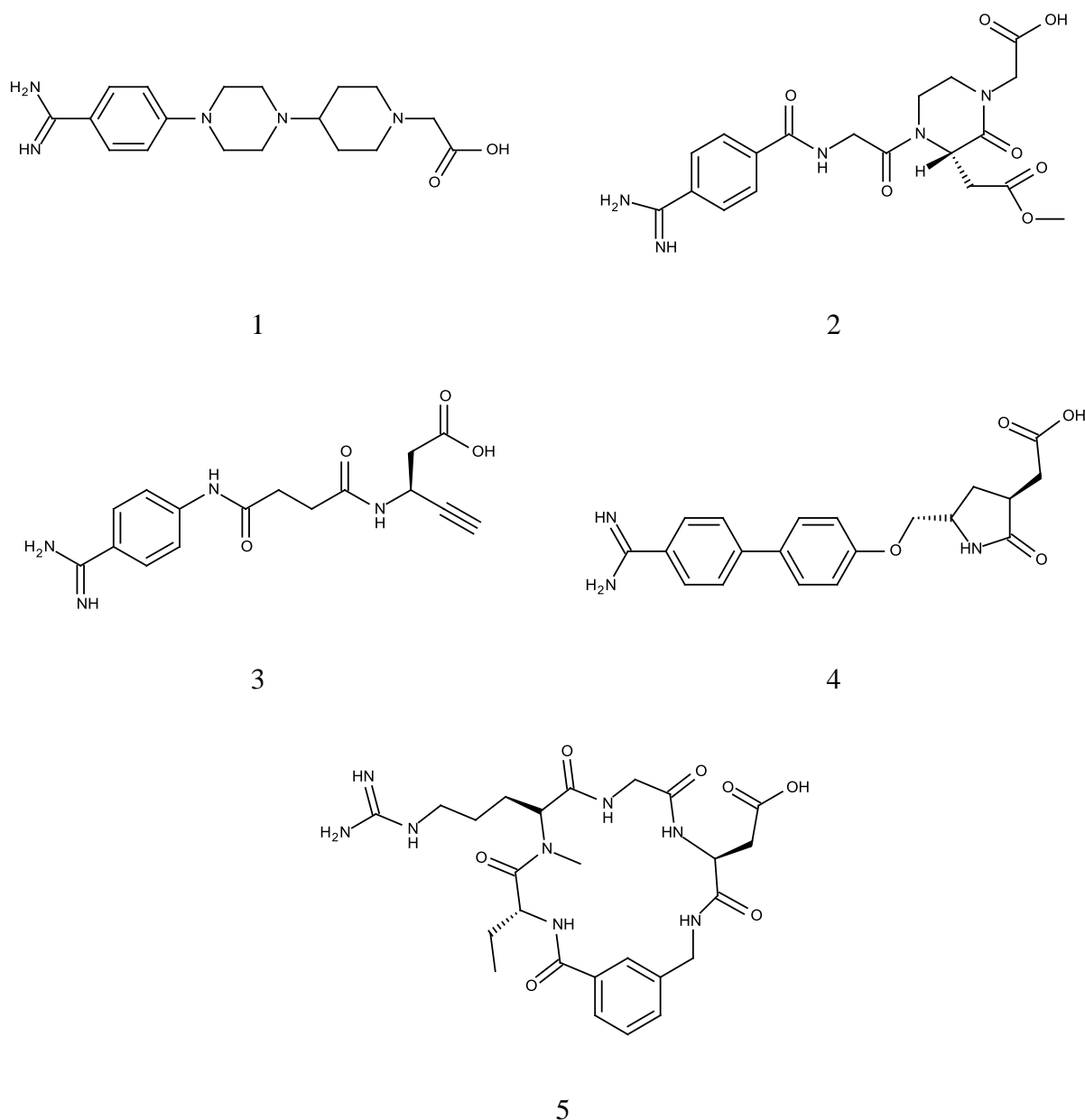
**Figure 19.** Simplified pharmacophore model for antithrombotic agents from references [217-218].

In later works the authors used alignment tools to generate 3D pharmacophore models which helped to extend the already known simple two-point model [219-222]. Thus, Okumura et al. [221] based on the set of known fibrinogen antagonists concluded that

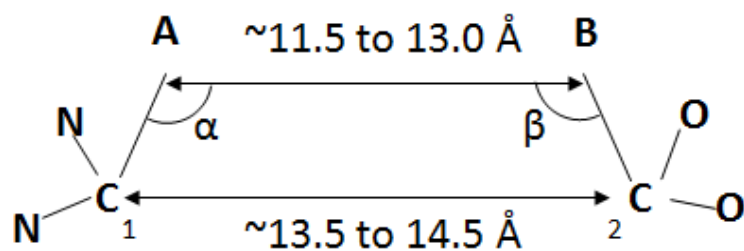
- (i) two functional groups (amidino or guanidino group and carboxyl group) are essential for the activity, and
- (ii) a spacer should be properly selected in order to correctly position these two functional groups.

To search the spatial position and orientation of the two functional groups vital for the activity, alignment of five compounds was performed using the DISCO module of SYBYL 6.3. The cyclic structure (5) obtained from the Cambridge Structure Database was used as a template (Figure 20). The distances between pharmacophore features were extracted from the alignments of the compounds used for model generation (Figure 21). The distance between the amidino/guanidino and carboxylate groups carbons ranged from 13.5 to 14.5 Å. The distance

between the atoms carrying the amidino and carboxylate groups respectively ranged from 11.5 to 13.0 Å. The “ $\alpha$ ” angle  $C^1$ -A-B ranged from  $150^\circ$  to  $170^\circ$ , and “ $\beta$ ” defined  $C^2$ -B-A ranged from  $85^\circ$  to  $110^\circ$ , respectively [221].



**Figure 20.** The compounds used for the alignment



**Figure 21.** Schematic representation of 3D pharmacophore model from reference [221]



The work of Dixon *et al.* [185] in which a 3D pharmacophore model has also been developed using PHASE, describes the fibrinogen antagonists more widely. The studied dataset contained 49  $\alpha_{IIb}\beta_3$  antagonists (benzodiazepine series and 1,3,4-trisubstituted 2-oxopiperazine derivatives) for which affinity values were available. The dataset was divided in training (23 compounds) and test sets (26 compounds) for the generation and validation of 3D pharmacophore models. Nine training set compounds with affinity values lower than 100 nM were considered as active, whereas five compounds with affinity values greater 1 $\mu$ M as inactive. A maximum of 500 conformations were generated for each molecule using MacroModel torsional sampling with OPLS\_2005-based post-processing. The developed pharmacophore model based on nine actives contained four features: positive and negative ionic centers at a distance of 14Å and two H-bond acceptors (Figure 22). Unfortunately, the authors didn't give any other spatial information about the pharmacophore model. The model was scored according to superposition of pharmacophore site points, alignment of vector characteristics, overlap of molecular volumes, and penalization of matches to inactive training set molecules.



**Figure 22.** 3D pharmacophore model developed by Dixon et al [185]. The positive and negative ionic features are indicated by cyan and red spheres, while the hydrogen bond acceptor features are rendered as pink spheres with lone pair vectors.

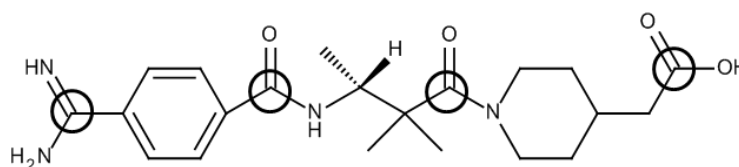
For purposes of QSAR development, van der Waals models of the aligned training set molecules were placed in a regular grid of cubes, with each cube allotted zero or more “bits” to account for the different types of atoms in the training set that occupy the cube. This representation gives rise to binary-valued occupation patterns that can be used as independent variables to create PLS QSAR models. The model was validated using the test set. The obtained correlation coefficient ( $Q^2$ ) was 0.40, the root mean-squared error (RMSE) = 0.89 and Pearson-R=0.67.

The developed pharmacophore model was used to carry out a search of the ASINEX and ACB databases of 226000 molecules, preliminarily seeded with eight high-affinity  $\alpha_{IIb}\beta_3$  antagonists belonging to five distinct chemical classes apart from those used in the training and test sets. This search recovered five of the eight known actives and only one false-positive that corresponds to an enrichment of  $(5/6) \div (8/226000) \approx 23500$ .

## 4.2. QSAR modeling

There are several three-dimensional QSAR studies that examined antagonists of the  $\alpha_{IIb}\beta_3$  receptor. The group of Miyashita et al. [223] examined a series of RGD-mimetics using Comparative Molecular Field Analysis (CoMFA), Hydrophobic Interaction (HINT), and Comparative Molecular Similarity Indices Analysis (CoMSIA) approaches. All of 63 compounds in this series contained an amidino- or guanidinophenyl groups as Arg-mimetic and piperidine-4-acetic acid or piperidine 4-carboxylic acid as Asp-mimetic. For all compounds the values of  $IC_{50}$  (the 50% inhibitory concentration in mol/l for platelet aggregation induced by collagen) were given. The 3D structures of the compounds in this study were generated by the molecular software package SYBYL 6.5.33. Systematic search was used to obtain the low-energy conformations and the structures were minimized using molecular mechanics calculation performed with the Tripos force field. During the systematic search and energy minimization, the distance between the carbon atoms of the amidino and carboxylate groups was constrained to be greater than 12Å since cationic and anionic groups are strongly attractive. The structure minimized by molecular mechanics was fully optimized by the semiempirical molecular orbital method, (PM3). For the purpose of alignment the compound which has relatively high potency and a fairly fixed conformation was used as template (Figure 23). The carbon atoms of the amidino and carboxylate groups and the carbon atoms adjacent to both the benzene and piperidine rings were taken as the key atoms for alignment.

The steric and electrostatic field energies were calculated by CoMFA. The obtained data was analyzed by the partial least squares (PLS) method. The obtained 3D QSAR model was validated by leave-one-out cross-validation (Table 2): the correlation coefficient was 0.714 and the standard error - 0.774.



**Figure 23.** Template compound used for the alignment in CoMFA study [223]. Key atoms for alignment are marked by circles.

Using HINT, hydrophobic fields were incorporated into CoMFA to evaluate hydrophobic interactions as a component of 3D QSAR model, the obtained leave-one-out cross-validation correlation coefficient (Table 2) was 0.72 and the standard error equaled 0.77. Also the analysis of steric, electrostatic and hydrophobic fields was performed with CoMSIA module of SYBYL. The obtained CoMSIA model which included steric and electrostatic fields was characterized by

the leave-one-out cross-validation correlation coefficient 0.66 and the standard error 0.85. The CoMSIA model incorporated all three field types (steric, electrostatic and hydrophobic) had the correlation coefficient 0.70 and the standard error 0.84. In both cases the quality of CoMFA models was better than CoMSIA models in terms of the leave-one-out cross-validation procedure.

Table 2. Statistics for developed 3D-QSAR models by Miyashita et al.

Model	Methods	Conventional <sup>a</sup>		Cross-validated <sup>b</sup>		RC <sup>c</sup>			Eq no.
		S	R <sup>2</sup>	S <sub>press</sub>	Q <sup>2</sup>	St	El	Hp	
1	CoMFA	0.49	0.89	0.77	0.71	80.2	19.8	-	16
2	CoMFA+HINT	0.48	0.89	0.77	0.72	44.2	10.4	45.3	17
3	CoMSIA(St+El)	0.51	0.88	0.85	0.66	34.2	65.8	-	18
4	CoMSIA(St+El+Hp)	0.17	0.99	0.84	0.70	15.3	30.6	54.1	19

<sup>a</sup> Calculated using training set; <sup>b</sup> Calculated using the leave-one-out cross-validation; <sup>c</sup> Relative contribution (%) of St: steric, El: electrostatic, Hp: hydrophobic effects. S - the standard error of estimate; S<sub>press</sub> - the standard error calculated using the leave-one-out cross-validation procedure.

The authors supposed that CoMFA better characterized the studied dataset than CoMSIA, since compounds in the dataset were diverse structurally rather than electrostatically and CoMFA emphasized steric interactions. Although, the statistics of model (2) was better than for model (1), the found hydrophobic regions failed to satisfactorily explain the variations in inhibitory activity. Thus, the CoMFA model (1) was chosen as best one.

The generated CoMFA field map showed that the positive electrostatic regions around the ligand which are favorable for high activity surrounded the amidino group and negative electrostatic regions appeared near the carboxylate group. These electrostatic fields indicate the importance of cationic and anionic functional groups in the RGD mimetics for high activity. The optimal distance between these groups was close to 15.3 Å. Also they have found that negative electrostatic regions above carbonyl groups that are placed close to the cationic and anionic functional groups suggested that electrostatic interactions or hydrogen bonds between these oxygen atoms and the receptor contributes to an increase in activity [223].

A sterically favorable region for higher activity is found around  $\alpha$  and  $\beta$  carbon atoms of  $\beta$ -alanine fragment, and presence of steric forbidden regions near basic amidino-NH group has a negative effect.[223]

Later on, the group of Yan et al. [224] has investigated  $\alpha_V\beta_3/\alpha_{IIb}\beta_3$  dual antagonists that could be potential drugs for treating of acute ischemic diseases. The studied dataset was

composed of 101 tricyclic piperazine/piperidine-containing dual  $\alpha_V\beta_3/\alpha_{IIb}\beta_3$  antagonists with associated affinity values ( $IC_{50}$ ). For model development the authors also used CoMFA and CoMSIA approaches.

The whole dataset was divided into training (81 compounds) and test (20 compounds) sets, respectively. The test set molecules were selected manually such that low, moderate, and high activity compounds were present in approximately equal proportions in both sets, so that training and test sets cover the same potency range. The 3D-QSAR studies were performed using the SYBYL 6.9 molecular modeling software package. Energy minimization and conformational search were performed using Tripos molecular mechanics force field.

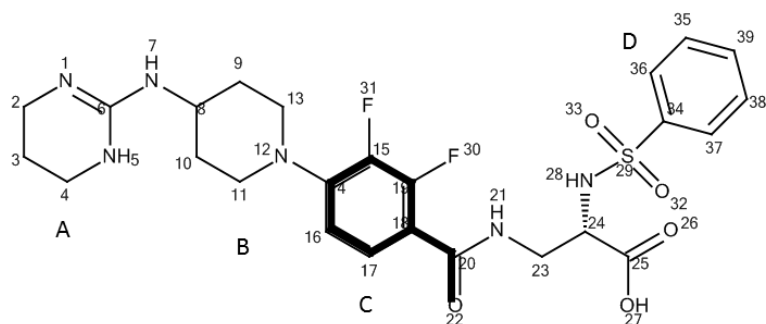
The most potent dual antagonist was chosen as a template (Figure 24) to fit the remaining compounds of training and test sets. To evaluate the reliability of the obtained PLS models, cross-validation analysis was accomplished with the leave-one-out methodology. The predictive ability of the best generated models was evaluated on the external validation set of 20 compounds. The obtained correlation coefficients ( $R^2$ ) between observed and predicted affinity values for  $\alpha_{IIb}\beta_3$  receptor of test set compounds were shown in Table 3.

Table 3. Statistics for developed 3D-QSAR models by Yan et al.

Methods	Training set			Training set		Contribution <sup>a</sup>				
	$Q^2$	$R^2$	SEE	$R^2_{pred}$	SEP	St	El	Hp	HBD	HBA
CoMFA	0.50	0.85	0.52	0.72	0.65	58	42	-	-	-
CoMSIA	0.52	0.88	0.45	0.47	0.62	17	16	20	29	18

$Q^2$  – cross-validated correlation coefficient after the leave-one-out procedure;  $R^2$  – Pearson correlation coefficient; SEE – standard error of estimate;  $R^2_{pred}$  – predicted correlation coefficient for the test set of compounds; SEP – standard error of prediction. <sup>a</sup>Contribution (%) of St: steric, El: electrostatic, Hp: hydrophobic effects, HBD: hydrogen-bond donor; HBA: hydrogen-bond acceptor.

In CoMFA, the steric and electrostatic fields were calculated, whereas in CoMSIA, the steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor descriptors were used (Table 3). The analysis of calculated steric CoMFA field showed that bulkier groups near the benzene ring D are favorable and vice versa the bulkier substituents near the 4-aminopiperidine ring B decrease the activity. Also the calculated electrostatic contour map confirmed the importance of positive and negative charged groups for antagonistic activity of ligands of  $\alpha_{IIb}\beta_3$  receptor.

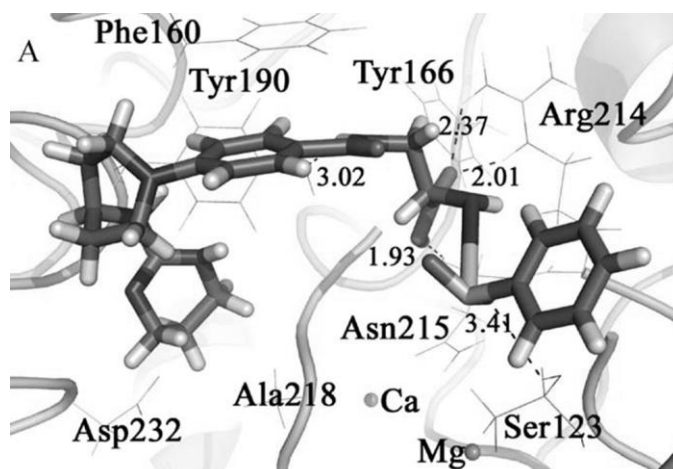


**Figure 24.** The template compound chosen for CoMFA/CoMSIA studies [224]. The fragment used as a reference during alignment is marked in bold.

Analysis of CoMSIA results showed that the following factors may be relevant:

- i) the substitution at positions 15, 19 and 24 by hydrophilic groups,
- ii) H-bond acceptors near benzene ring C ,
- iii) strong H-bond acceptor over position 24 and 25 is detrimental to the biological activity.

Additionally the docking study with Surflex from SYBYL was carried out into the X-ray structure of  $\alpha_{IIb}\beta_3$  receptor (PDB: 2VDM). All studied compounds were docked into the prepared binding site. During the docking process, the protein was considered as a rigid one and the antagonists molecules as flexible ones.



**Figure 25.** Docked template compound in the binding site of  $\alpha_{IIb}\beta_3$  receptor from reference [224]

According to the authors, the docking results concur with the ones of CoMFA and CoMSIA studies. Several hydrogen bonds were observed for the docked template in same area as it was seen in CoMSIA maps (from 20 to 25 positions). Two hydrogen bonds were formed by Arg214 with the  $-C=O$  group at position 26 in the template compound, Asn215 – with  $-OH$  group at position 27, Tyr190 – with  $-NH-$  group at position 21 and the sulfonamide group – with Ser123. Also near the ring D the amino acid residues in the docked model were absent, what, similar to CoMFA and CoMSIA maps, shows that steric interaction is favorable there. In

addition, hydrophilic amino acids Arg214, Asn215 appear around sulfonamide substituent, which is similar to the CoMSIA hydrophobic contour maps at position 24 of template compound.

There are some drawbacks in the described QSAR studies. In first one, the external validation is lacking, failing to assess the real predictive ability of the developed models. Among the problems with the second study one may cite:

- 1) the usage of the neutral forms of compounds for the molecular fields and docking studies (as it was shown by pharmacophore models the two charged features of the ligands are essential for compounds binding),
- 2) the distorted conformation of the sulfonamide group of template ligand, that could lead to the misrepresentation of the generated molecular fields and wrong docking poses and scoring.

Despite the mentioned drawbacks, the described studies bring useful information about  $\alpha_{IIb}\beta_3$  antagonists. Both studies found that the steric effects make much more contribution to the variation of the activity of the antagonists than electrostatic ones, what can be explained by the high similarity of the structures in the used training sets. It can be supposed that the electrostatic effects don't change significantly from one molecule to another, that's why the electrostatic effects have less contribution to the variation of the activity.

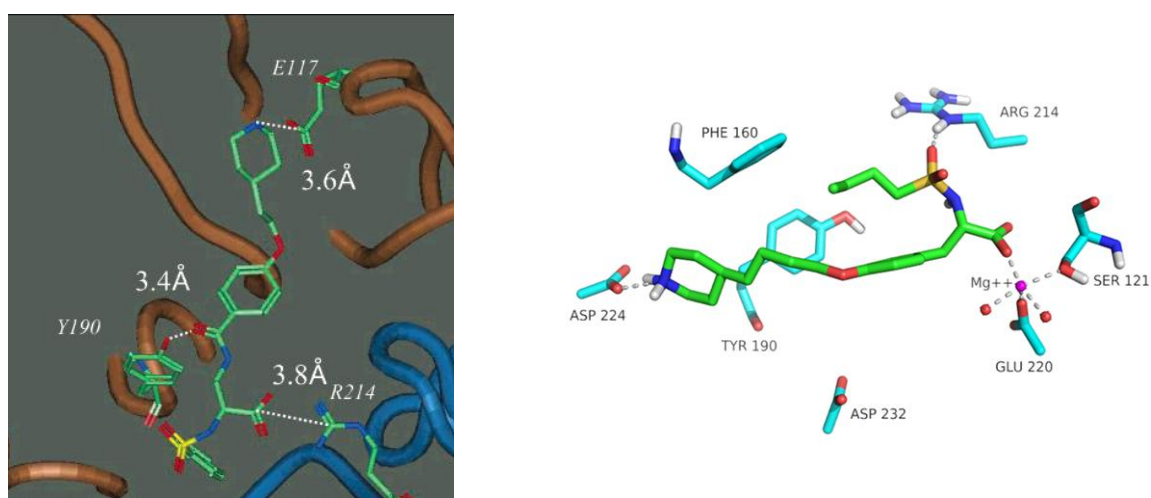
The presence of cationic and anionic centers is considered as preferable for activity, as well as presence of hydrogen bonds. In one of the work the optimal distance between cationic and anionic center was measured and was 15.3 Å, which is also in agreement with pharmacophore models.

### 4.3. Molecular docking

Since the X-ray structure of  $\alpha_{IIb}\beta_3$  receptor was determined recently, a modeled receptor has been used in the earlier docking studies of fibrinogen antagonists. Usually, for modeling of  $\alpha_{IIb}\beta_3$  the authors have used  $\alpha_v\beta_3$  integrin, a member of the same superfamily of cell surface glycoproteins. Integrins are highly homologous (38% identity and 54% homology for  $\alpha_v$  and  $\alpha_{IIb}$ ) [225-227]. Both receptors,  $\alpha_{IIb}\beta_3$  and  $\alpha_v\beta_3$ , have some similar characteristics, they are known to bind an Arg-Gly-Asp (RGD) segment of such peptides as fibrinogen and vitronectin [228]. Also, their binding sites include a MIDAS motif ( $Mg^{2+}$  ion in the  $\beta_3$  metal ion-dependent adhesion site), observed after crystal structures were determined [229-230].

The molecular model of the  $\alpha_{IIb}\beta_3$  receptor has been developed based on the crystal structure of  $\alpha_v\beta_3$  receptor by Feuston et al. and used for docking studies [226]. The studied dataset consisted of nine RGD-peptidomimetics, three of them belonged to sulfonamide class of

compounds and were specific antagonists of  $\alpha_{IIb}\beta_3$ , one – also a sulfonamide, was a dual antagonist. Other four compounds were selective antagonists of  $\alpha_V\beta_3$  receptor and the remaining one was ineffective against either integrin. To define the active binding site two compounds were docked. The first one is RGD-peptidomimetic, a specific antagonist of  $\alpha_{IIb}\beta_3$  receptor, while the second one was selective inhibitor of  $\alpha_V\beta_3$  receptor used as the counter example for the modeled receptor. For docking purpose, the ligands were formally charged. While the docking procedure for the first ligand resulted in a reasonable binding mode, it failed for the second one, as expected. The important interactions for ligand binding were defined for  $\alpha_{IIb}$ : Glu117, Tyr190, and  $\beta_3$ : Arg214; with the critical distance between the charged residues (Glu117 – Arg214) in 15.3Å (Figure 26). After the active binding site was defined, the remaining compounds were docked. [226]

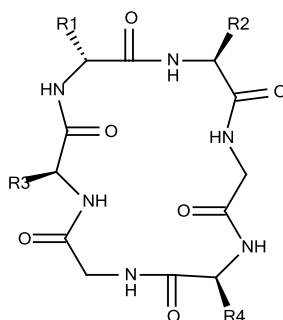


**Figure 26.** The binding mode of one of the specific antagonists of  $\alpha_{IIb}\beta_3$  from reference [226] (left) and the binding mode of Tirofiban from the X-ray structure (right).

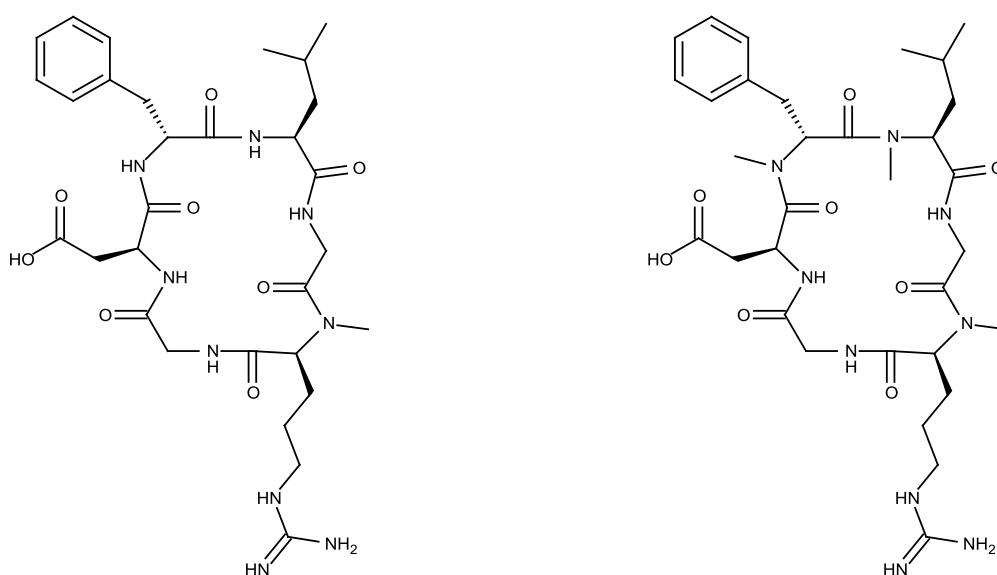
Comparing the modeled receptor with the X-ray structure (PDB: 2VDM), it can be seen that there are some differences which can influence the binding mode. The  $Mg^{2+}$  ion, which, as was seen latter, is responsible for interaction with Asp-mimicking fragment of RGD-mimetics [231], is absent in the modeled receptor. Residue  $\alpha_{IIb}$  Glu117, which according to the docking study is one of the key hot spots, is turned out of the binding site in the X-ray structure, by contrast to the modeled receptor.

In the latter work of Chatterjee *et al.* [232] a docking study was carried out to define the binding mode of the designed compounds. For the purpose of design, the cyclic hexapeptide cyclo(-G<sup>1</sup>R<sup>2</sup>G<sup>3</sup>D<sup>4</sup>f<sup>5</sup>L<sup>6</sup>-) (Figure 27) was used. It was designed a small library of seven derivatives in which only the externally oriented (solvent exposed) amide protons were N-methylated. N-methylation was used for enhancement of selectivity of  $\alpha_{IIb}\beta_3$  receptor antagonists by reduction

in the flexibility of arginine and glycine residues that presented the peptide backbone in an extended orientation.



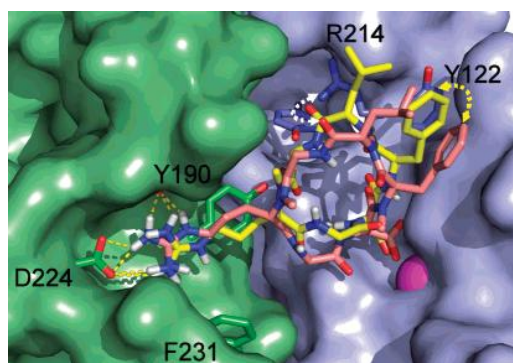
**Figure 27.** Template compound used for design in reference [232].



6

7

**Figure 28.** The compounds used for docking studies in reference [232].



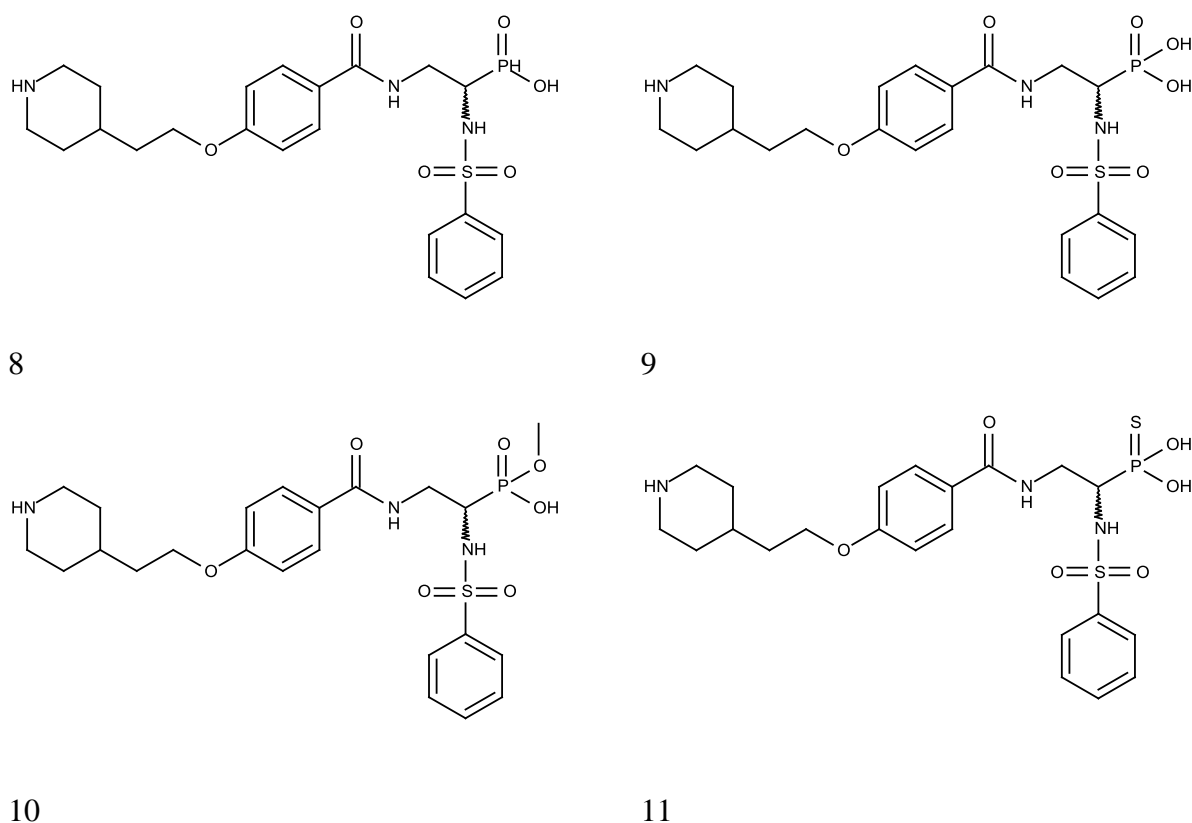
**Figure 29.** Docked **6** (yellow) and **7** (pink) in the  $\alpha_{11b}\beta_3$  integrin. The  $\alpha_{11b}$  subunit of the receptor is represented by the green surface, while the  $\beta_3$  subunit is represented by the violet surface. In both subunits, important side chains are highlighted as sticks. The metal ion in the MIDAS region is represented by a magenta sphere. The loss of  $\pi$ - $\pi$  interaction of D-Phe residue of **7** with Tyr122 is shown by the yellow arrow, and the improper orientation of Leu CO of **7** to form a hydrogen bond with the Arg214 side chain is shown by the white arrow.



For two compounds (Figure 28) from the designed library, the docking studies (Figure 29) based on the complex of  $\alpha_{\text{IIb}}\beta_3$  and Eptifibatide (PDB: 1TY6) were performed using Autodock. In both peptides, the carboxylate group of Asp was found to coordinate the metal ion at the MIDAS region, whereas the Arg side chain extended into the deep  $\beta$ -propeller pocket, forming a hydrogen bond to the  $\alpha_{\text{IIb}}$  Asp224 carboxylate group. There are some differences in the predicted binding modes of the compounds 6 and 7. Due to multiple N-methylation in 7, and especially due to the N-methylation of D-Phe residue, 7, compared to 6, seems to lower its  $\pi$ - $\pi$  interaction with  $\beta_3$  Tyr122 and didn't properly orient the Leu carbonyl group to hydrogen bond with the  $\beta_3$  Arg214 side chain.

In conclusion, the authors considered that the selectivity of their library arises predominantly due to the reduced flexibility of the peptide. Thus, multiple N-methylation of peptides could be straightforward and simplistic to obtain highly potent and selective ligands [232].

Recently, Bollinger et al. [233] reported a docking study and drug design of  $\alpha_{\text{IIb}}\beta_3$  antagonists. They investigated new phosphorus-containing integrin ligands with the aim of unraveling the steric and electrostatic requirements of the MIDAS region. The essential carboxyl group of tirofiban analogs was replaced by different phosphorus-containing groups and four compounds were designed (Figure 30). Docking with AutoDock targeted the active site of the complex of  $\alpha_{\text{IIb}}\beta_3$  and tirofiban (PDB: 2VDM). The reliability of this approach was assessed through re-docking.



**Figure 30.** Designed phosphoryl-containing analogues of tirofiban (Figure 11) tested in docking study by Bollinger et al. [233].

As a result of docking, the predicted binding mode of all designed compounds is very similar to that of tirofiban. All compounds in (R)-configuration bound to the MIDAS similar to tirofiban and additionally form some H-bonds with  $\beta_3$ Asn215 and  $\beta_3$ Tyr122. All other interactions are similar to those formed by tirofiban. In order to understand how the different chirality of the carbon bearing phosphinic/phosphonic group influences the inhibitory activity, docking calculations on the corresponding (S)-enantiomers of the most potent compounds 8-10 were performed. The results showed that in the (S)-configuration, all the three compounds are able to coordinate the metal – with some differences in the binding mode with respect to the corresponding (R)-enantiomers. While in the  $\alpha_{IIb}$  subunits both interactions with Asp224 and Ser225 are conserved, in the  $\beta_3$  subunit a loss of the hydrogen bond between the sulfonamide group and  $\beta_3$ Asn215 and the benzenesulfonamide moiety is not located in the  $\alpha_{IIb}\beta_3$  aromatic pocket is observed. [233]

The authors assert that phosphinates as well as phosphonate monomethyl esters are suitable isosteres for the carboxyl group in integrin ligands, while phosphate and thiophosphate groups are not: highly negatively charged metal-coordinating groups are not well tolerated in the  $\alpha_{IIb}\beta_3$  MIDAS [233].

The docking studies described above, except the one of Feuston *et al.*, confirm the interactions which can be seen in the X-ray structure. There are some useful conclusions that

emerge: 1) reduction in the flexibility of ligand can enhance the selectivity; 2) highly negatively charged metal-coordinating groups are not well tolerated in the  $\alpha_{\text{IIb}}\beta_3$  MIDAS.

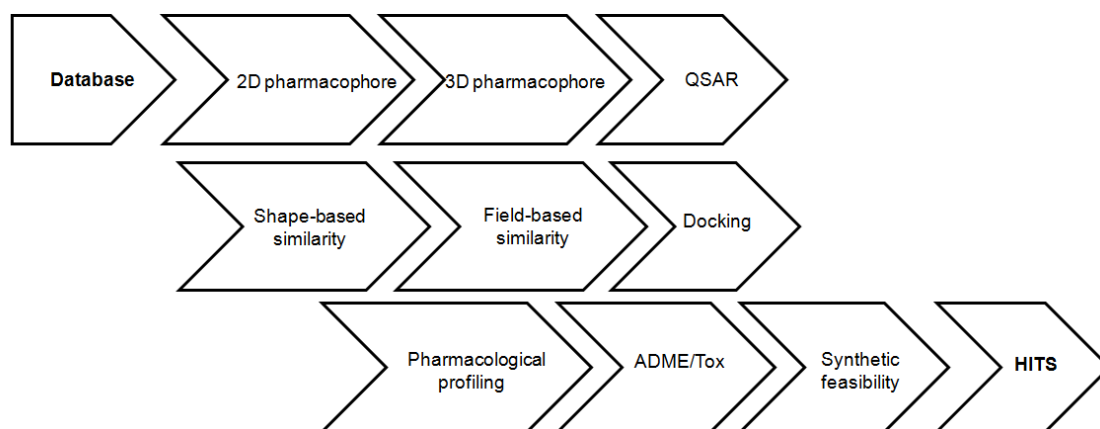
#### 4.4. Conclusions

The analysis of the studies described above showed that all of them are consistent about the importance of cationic and anionic centers. The pharmacophore and 3D QSAR models determine the distance  $\sim 15$  Å between these centers as optimal. The described docking studies revealed that reduction in the flexibility of ligand can enhance the selectivity and highly negatively charged metal-coordinating groups are not well tolerated in the  $\alpha_{\text{IIb}}\beta_3$  MIDAS.

Some results of the described studies are questionable and can lead to wrong conclusions. Improperly charged ligands could negatively influence on the estimation of electrostatic interactions in molecular docking and CoMFA modeling. The using of compounds with high similarity as training sets in 3D QSAR modelling could result in controversial judgment that the steric effects contribute more to the variation of activity in contrast to the electrostatic ones. Since N- and C-terminal groups of ligands varied a little across the training sets one can expect that the electrostatic fields are very similar for all compounds and thus electrostatic effects could not be estimated properly. High similarity of training set compounds may substantially decrease the number of possible binding modes described by ligand-based pharmacophore models. Also it should be noted that some of the models were not validated on external datasets.

## 5. Computer-aided design of new RGD-peptidomimetics

In this section we will describe the virtual screening of RGD peptidomimetics. It includes structure-based as well as ligand-based approaches.



**Figure 31.** Virtual screening workflow used to discover new RGD-peptidomimetics.

### 5.1. Dataset description

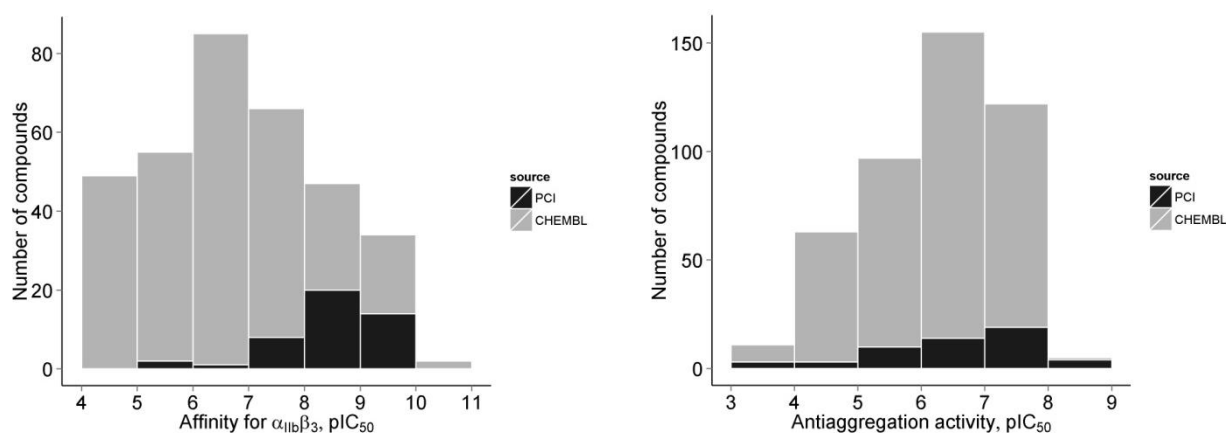
Two datasets of RGD-peptidomimetics which report either affinity for  $\alpha_{\text{IIb}}\beta_3$  or anti-aggregation activity have been provided by PCI. All of those compounds have been synthesized and tested for affinity for  $\alpha_{\text{IIb}}\beta_3$  and anti-aggregation activity at the Medicinal Chemistry Department of A.V. Bogatsky Physical-Chemical Institute (see Appendix 1). Anti-aggregation activity of compounds was measured by Born's method on human platelet rich plasma [234]. Affinity for  $\alpha_{\text{IIb}}\beta_3$  was measured as inhibition of fluorescein isothiocyanate-labeled fibrinogen binding to activated human platelets by tested compounds [235]. These two datasets contained relatively small numbers of compounds (45 compounds with reported affinity values and 53 – with reported anti-aggregation activity) and they were significantly imbalanced because they contained mostly active compounds (Figure 32). Due to this fact these datasets have been expanded by data from the other source – ChEMBL database (version 7) [236], which is a publicly available collection of organic compounds with associated bioactivity data taken from the leading medicinal chemistry journals. The compounds from ChEMBL have been selected taking into account similarity of the used bioassays with respect to used in PCI tests. This was the crucial step, because activity values for the same compound obtained in different assays may differ more than by one order of magnitude [237].

Curation of the compounds structures of both datasets has been performed by using ChemAxon *Standardizer* tool [238]: (i) mixtures and inorganics were removed, (ii) normalization of specific chemotypes (aromaticity and nitro groups were checked) were

performed, (iii) explicit hydrogen atoms were added. Entries featured both achiral and chiral compounds (single stereoisomers and racemic mixtures). Therefore, separate stereoisomers were excluded if corresponding racemic compounds were present in the dataset. Duplicates in the datasets have been removed using ChemAxon Instant JChem [239]. All values of affinity for  $\alpha_{IIb}\beta_3$  and anti-aggregation activity of compounds were converted to  $pIC_{50}$  ( $-\lg IC_{50}$ ,  $IC_{50}$  in mol/l). Thus, two modeling sets - *Affinity* and *Anti-aggregation* - were created, see Table 4 and Figure 32.

Table 4. Size and composition of the modeling sets.

	PCI data	Literature data	Overall size of the dataset
<i>Affinity</i> set	45	293	338
<i>Anti-aggregation</i> set	53	400	453



**Figure 32.** Distribution of experimental  $pIC_{50}$  ( $-\lg IC_{50}$ ,  $IC_{50}$  in mol/l) values for the *Affinity* (left) and *Anti-aggregation* (right) datasets. See Table 4 for the details.

Each RGD peptidomimetic could formally be represented by three parts: Arg-mimicking fragments (positively charged), linkers or Gly-mimicking fragments and Asp-mimicking fragments (negatively charged). In the collected datasets, the Arg-mimicking fragment is represented by primary, secondary and aromatic amines, which may be both linear (e.g. alkyl guanidine) or cyclic fragments (e.g. piperidine, piperazine, pyridine, benzamidine, tetrahydroisoquinoline etc.). Asp-mimicking fragments are  $\alpha$ ,  $\beta$ -substituted, hydroxyl, keto- and amino acids (Val, Phe,  $\beta$ -amino acids, etc). Linkers are linear or branched chains, as well as cyclic fragments, such as benzene, piperidine, piperazine, indole, diazepam, pyrrolidine and thiophene (see Figure 34).

It should be noted that for the actives from the *Affinity* set (for which  $pIC_{50} > 6$ ) the topological distance between positively charged nitrogen atom in *Arg*-mimetic and negatively charged COO group in *Asp*-mimetic is larger or equal to 13 (see Figure 33).

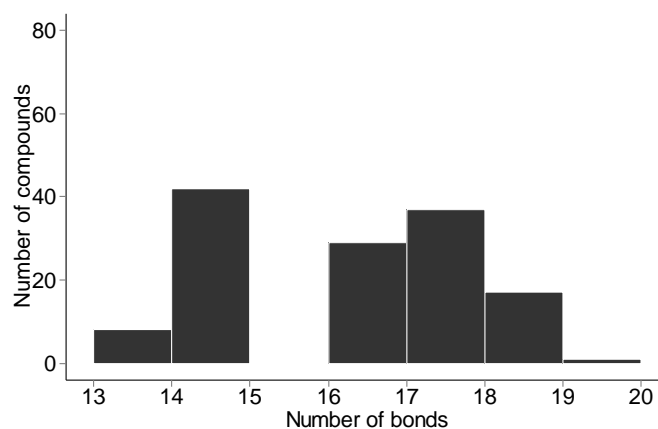


Figure 33. The topological distance between positively charged nitrogen atom in *Arg*-mimetic and negatively charged COO group in *Asp*-mimetic for the compounds with  $pIC_{50} > 6$ .

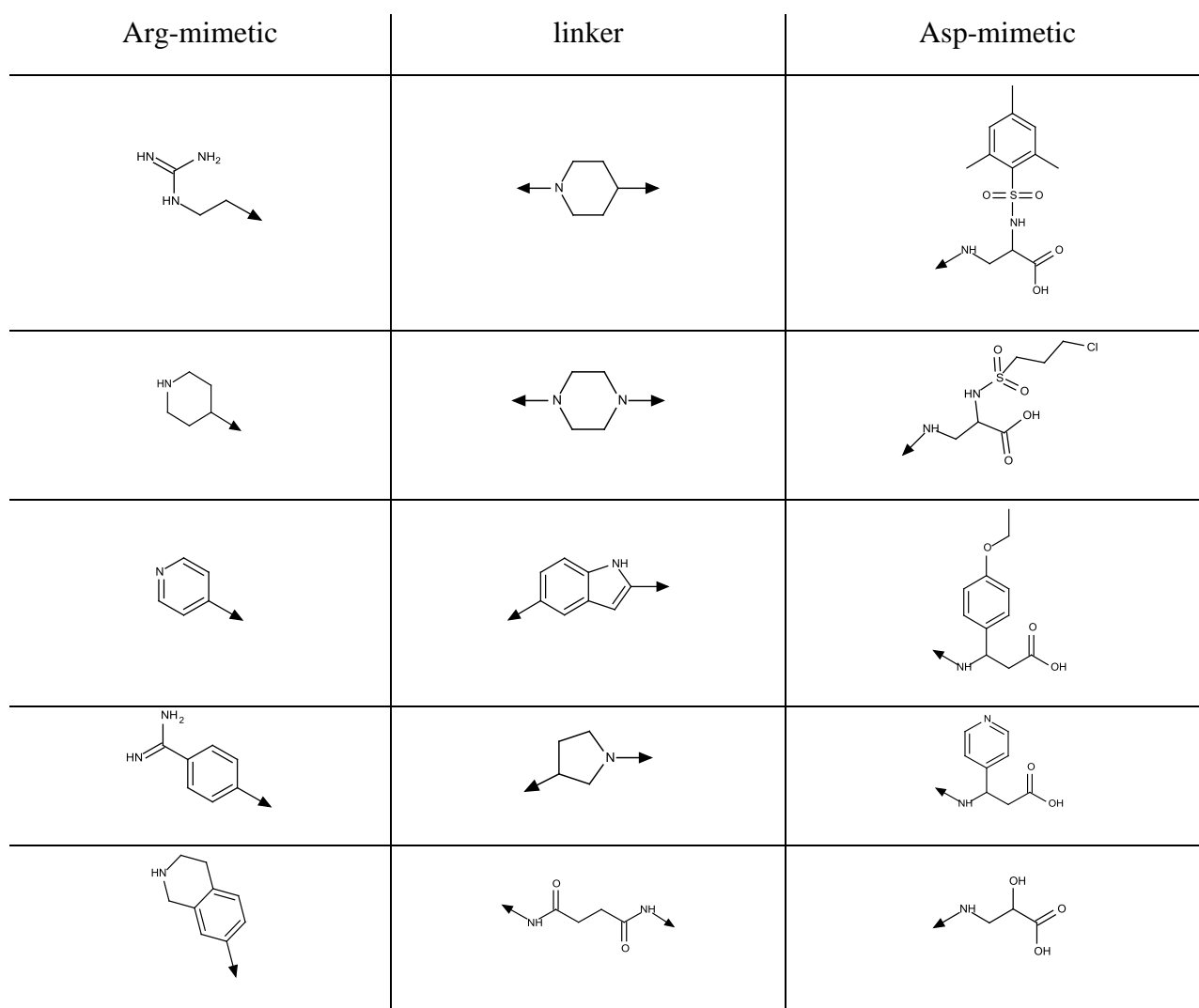


Figure 34. Examples of *Arg*-, *Asp*-mimetics and linkers encountered in the collected datasets.

## 5.2. QSAR modeling

Bounded and unbounded two-dimensional simplex descriptors were generated using Hierarchical QSAR Technology (HiT QSAR) software [137]. Besides, atom type also other physico-chemical characteristics of an atom, such as partial charge, lipophilicity, electronegativity, refraction, and the ability for atom to be a donor or acceptor in hydrogen bond formation were calculated for atom differentiation in SiRMS. The SMF descriptors were generated by ISIDA/QSPR software. As types of molecular subgraphs sequences of atoms and bonds and “terminal” groups were calculated. In affinity models, the length of bonds sequences were selected in the range from 2 to 8, whilst longer sequences (from 2 to 15) were included for modeling the anti-aggregation activity. FPT descriptors were also generated, with default parameters. The edge length of basis triangles, i.e., the number of bonds between two pharmacophore types was in range from 2 to 12.

In order to clarify the complementarity of the two sets of compounds collected from different sources – PCI and ChEMBL, Kohonen maps of the distribution of compounds in space of descriptors were obtained. Three types of descriptors were used (SiRMS, SMF and FPT), with tuning parameters mentioned above.

The maps were built using the *Kohonen* package implemented in the R package. The Kohonen map clusters the input compounds in such a way that compounds which are similar in chemical space are mapped in the same node. Euclidian distance was used as a similarity measure. The size of the circles represents the portion of compounds in comparison with the whole.

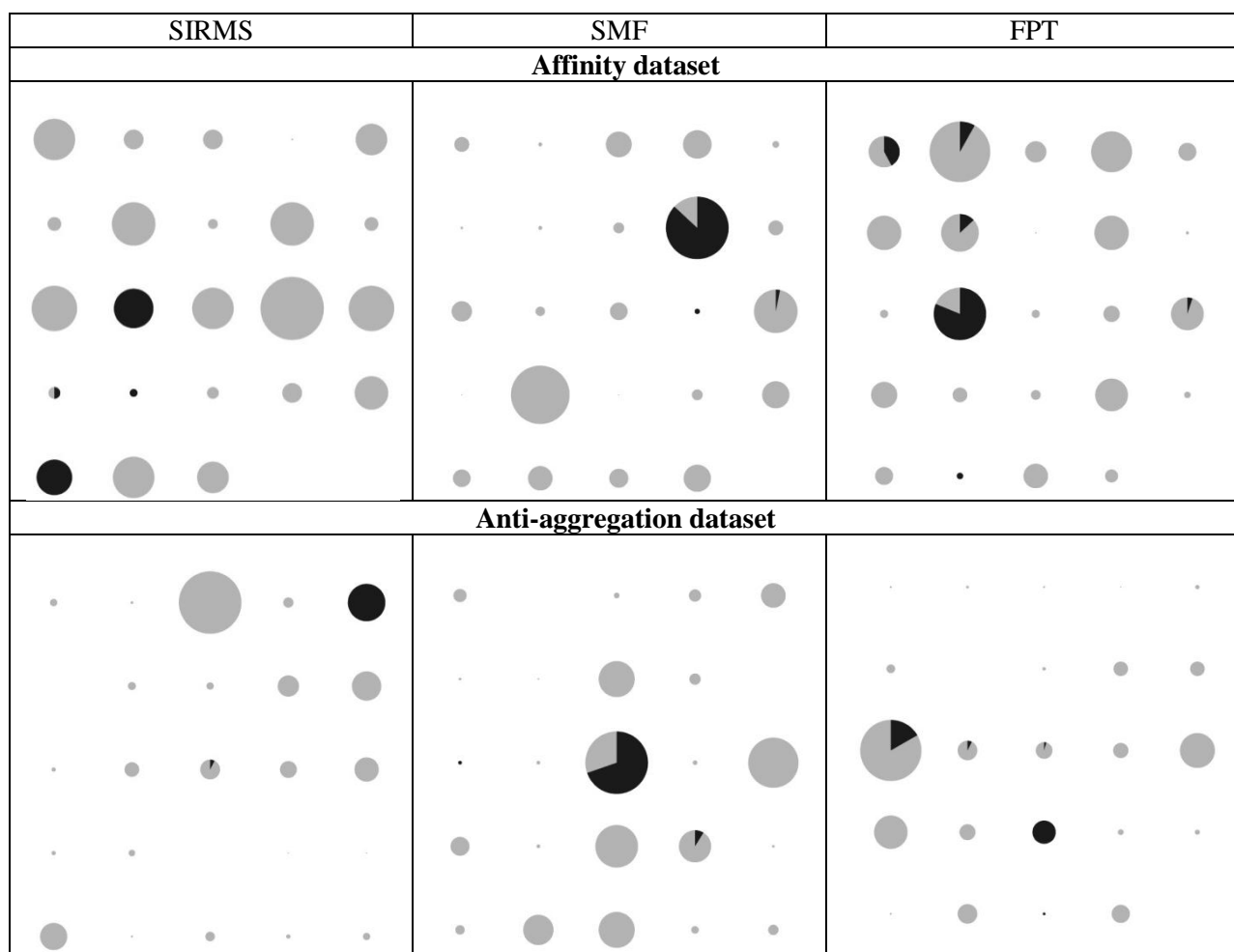


Figure 35. Kohonen maps of the distribution of compounds in space of descriptors. ChEMBL compounds are shown in grey, PCI in black.

Additionally, average inter-set PCI vs. ChEMBL Tanimoto scores were estimated and are shown in Table 5. The calculations were performed with SIRMS, SMF and FPT descriptors.

Table 5. Average value of Tanimoto coefficient for pairwise similarities of the compounds from PCI and ChEMBL sets.

SIRMS	SMF	FPT
<b>Affinity dataset</b>		
0.54	0.56	0.34
<b>Anti-aggregation dataset</b>		
0.65	0.57	0.36

The mapping of both sets of compounds in three types of descriptors spaces and Tanimoto coefficients shows ChEMBL and PCI collections are largely non-overlapping in



chemical space. Thus, it can be expected that one subset will not be able to reliably predict another one, which was confirmed by QSAR analysis.

A priori, two learning scenarios can be envisaged: 1) training on ChEMBL, validation against PCI compounds; 2) training on PCI, validation against ChEMBL compounds.

Table 6. Statistics for 2D QSAR models of affinity for  $\alpha_{IIb}\beta_3$

	Training set		Test set	
	$R_{oob}^2$	RMSE <sub>oob</sub>	$R^2$	RMSE
Descriptors	ChEMBL		PCI	
SIRMS	0.73	0.76	-0.64	2.48
FPT	0.77	0.67	-0.24	2.16
SMF	0.73	0.76	-0.44	2.33
	PCI		ChEMBL	
SIRMS	0.17	0.75	0.43	1.72
FPT	0.07	0.80	0.3	1.9
SMF	0.06	0.80	0.29	1.92

Table 7. Statistics for 2D QSAR models of anti-aggregation activity

	Training set		Test set	
	$R_{oob}^2$	RMSE <sub>oob</sub>	$R^2$	RMSE
Descriptors	ChEMBL		PCI	
SIRMS	0.58	0.70	-0.08	1.38
FPT	0.61	0.68	-0.17	1.44
SMF	0.56	0.72	-0.03	1.35
	PCI		ChEMBL	
SIRMS	0.52	0.89	-0.16	1.22
FPT	0.50	0.91	-0.18	1.24
SMF	0.46	0.94	-0.53	1.41

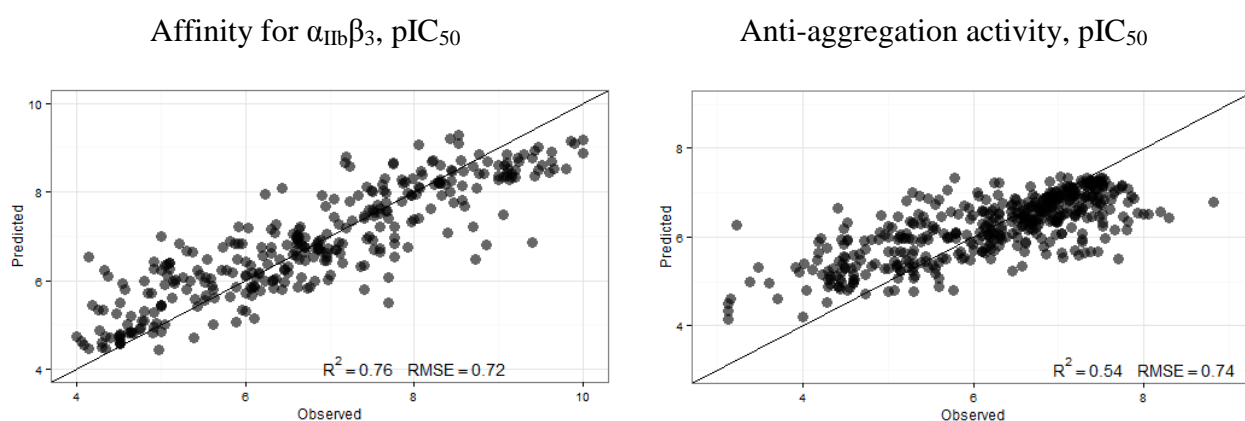
As expected, QSAR models based on the joint usage of both subsets should fare better, due to data better coverage of chemical space. Three individual 2D models for each joint dataset were developed using Random Forest method in combination with three types of descriptors, such as SiRMS, FPT and SMF described above. Predictive ability of the obtained models was estimated by 5-fold external cross-validation procedure (Table 8). The consensus QSAR models were developed by averaging predictions of the corresponding individual models (see Figure 36).

To estimate the applicability domain for the consensus models the standard deviation of the prediction, obtained from the ensemble of models, was used. As a threshold, 0.5 of the standard deviation, which is close to the experimental error (see Appendix 1), was taken. For the models of affinity for  $\alpha_{\text{Iib}}\beta_3$  and anti-aggregation activity - 10 compounds and 6 compounds, respectively, were out of applicability domain. The reason of unreliable predictions for some of these compounds can be easily explained. It was found that the biological experiments which were used to test some of the compounds differ. This obviously influences on the accuracy of their activity predictions. Since we assess AD of the consensus model by concordance between predictions of different models, and their predictions were inaccurate and differ a lot, so these compounds were out of the AD.

Table 8. 5-Fold external cross-validation statistics for 2D QSAR models of affinity for  $\alpha_{\text{Iib}}\beta_3$  and anti-aggregation activity performed on combined ChEMBL+PCI sets.

Descriptors	$R^2$	RMSE	$R^2_{\text{AD}}^{\text{a}}$	$\text{RMSE}_{\text{AD}}^{\text{a}}$	AD coverage
Affinity for $\alpha_{\text{Iib}}\beta_3$					
SiRMS	0.73	0.79	0.79	0.70	0.89
SMF	0.74	0.77	0.78	0.71	0.91
FPT	0.75	0.76	0.81	0.67	0.86
<b>consensus</b>	<b>0.75</b>	<b>0.76</b>	<b>0.76</b>	<b>0.72</b>	<b>0.97</b>
Anti-aggregation activity					
SiRMS	0.52	0.76	0.64	0.67	0.86
SMF	0.50	0.78	0.66	0.65	0.88
FPT	0.53	0.76	0.62	0.68	0.91
<b>consensus</b>	<b>0.52</b>	<b>0.77</b>	<b>0.54</b>	<b>0.74</b>	<b>0.99</b>

<sup>a</sup> results within applicability domain of the model



**Figure 36.** Observed versus predicted values of affinity for  $\alpha_{\text{Iib}}\beta_3$  and anti-aggregation activity of 5-fold external cross-validation procedure of consensus models.

Relatively big error (RMSE  $\approx$  0.65-0.74 log units) in predicted affinity for  $\alpha_{IIb}\beta_3$  and anti-aggregation activity can be explained by various reasons: difference of experimental conditions and high inter-individual variability of  $\alpha_{IIb}\beta_3$  population on platelets (see Appendix 1).

Predictive performance of the consensus model is similar to that of individual models (SiRMS, SMF, FPT) for both *Affinity* and *Anti-aggregation* sets (Table 7). However, the coverage of consensus models is much higher and, therefore, they were selected for virtual screening.

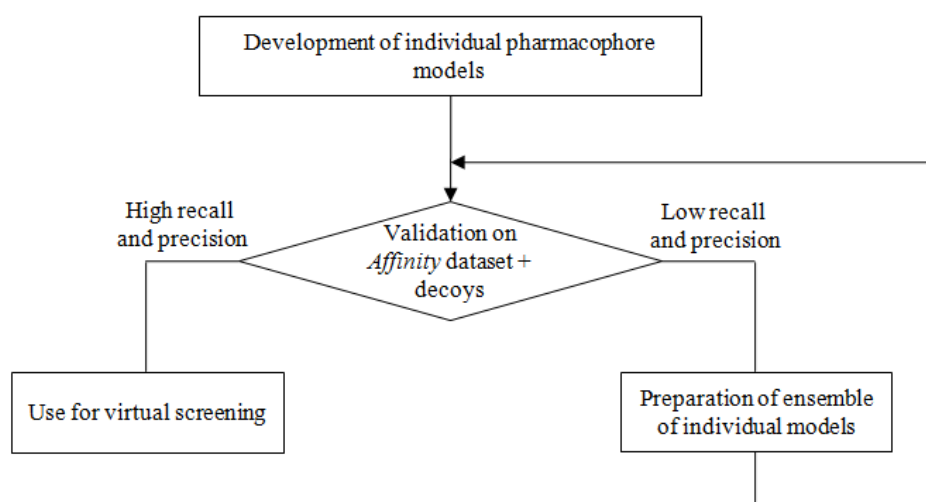
### 5.3. Pharmacophore models

In this chapter we will describe the development of the pharmacophore models for the antagonists of  $\alpha_{IIb}\beta_3$  receptor. Both structure-based and ligand-based models were developed. The models performance has been assessed in virtual screening on the validation set composed from the *Affinity* set (from which molecules used for the training were excluded) and a subset of decoys.

In our calculations, we consider as active molecules with  $pIC_{50} \geq 6$ . With this threshold, the total number of actives in the *Affinity* set is 234. The whole *Affinity* set was used in validation of structure-based models. In case of ligand-based pharmacophore, *Affinity* set were used with exclusion of training set molecules (the number of training molecules varies from 2 to 50)

A subset of 1518 decoys was selected from the ChEMBL database in such a way that some physic-chemical parameters (the number of H-donors and H-acceptors, logD, molecular weight, the number of rotatable bonds, topological polar surface area and partial charge) were similar to the compounds from the *Affinity* set. Thus, the number of inactives in the validation set was 1622 both for structure- and ligand-based models.

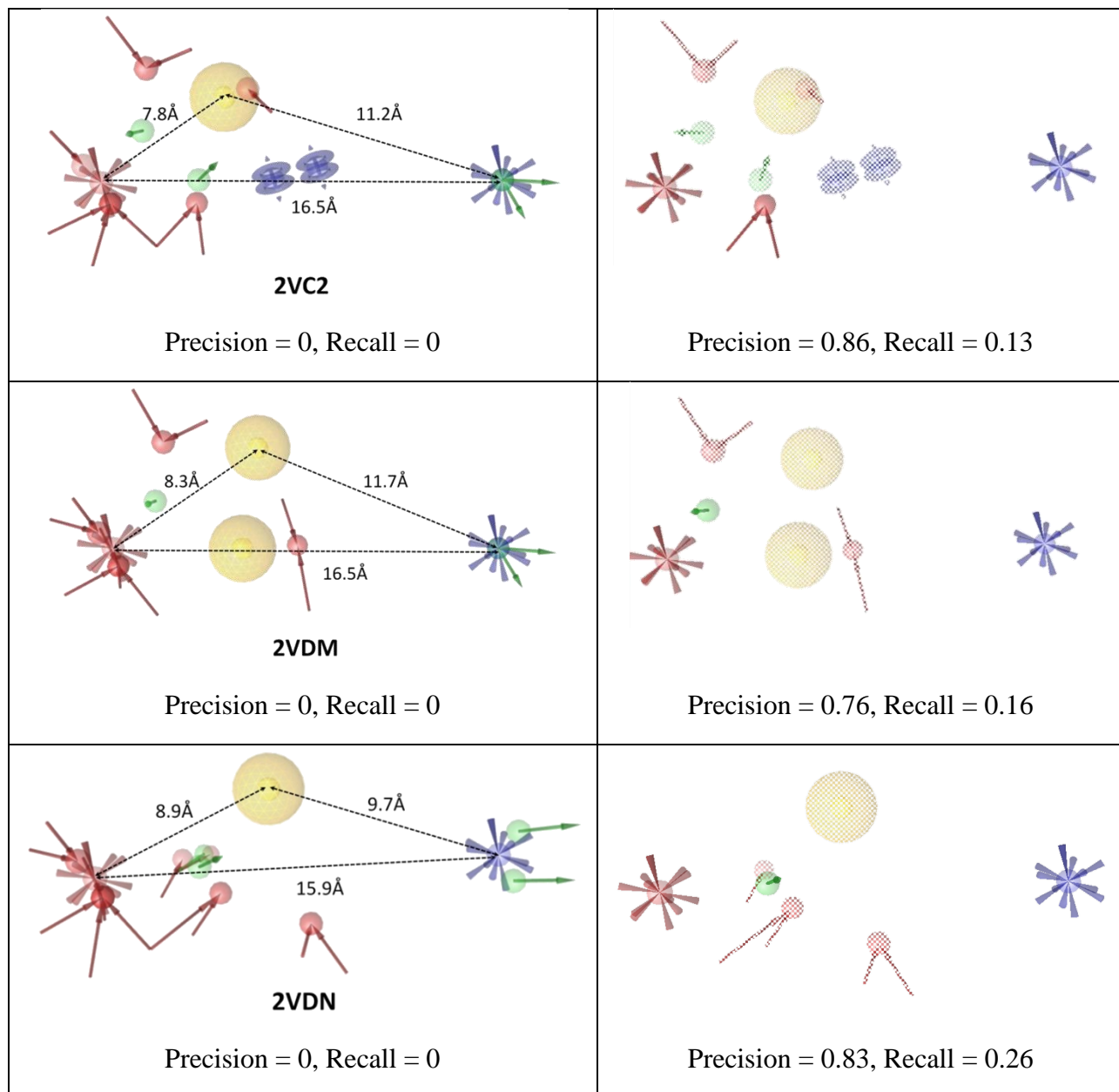
The general scheme for pharmacophore model obtaining and optimization is shown on Figure 37.



**Figure 37.** Workflow for pharmacophore model development

### 5.3.1. Structure-based models

Three structure-based pharmacophore models (Figure 38, left) were built with LigandScout, using 3 available X-ray structures of  $\alpha_{IIb}\beta_3$  complexes with its antagonists L-739,758, tirofiban, eptifibatide (PDB codes 2VC2, 2VDM, 2VDN, respectively). For this purpose, all ligands in complexes were ionized and minimized using the MMFF94 force field. All models were generated with exclusion volumes.



**Figure 38.** Structure-based pharmacophore models for  $\alpha_{IIb}\beta_3$  complexes. Two kinds of models are presented for each protein-ligand complex: automatically developed by LigandScout (left) and manually optimized (right). The color-coding of the pharmacophore features is shown on Figure 18.

The three models (Figure 38, left) have several main common features: positive and negative charges separated by 15.9-16.5 Å and one or two hydrophobic features situated between

them; near the positive and negative centers there are several hydrogen bond donors and acceptors, respectively.

The models were validated against the *Affinity* dataset. For the purpose of pharmacophore models validation the dataset was prepared in several steps. (i) All of the compounds were ionized using *Filter* tool from OpenEye software. (ii) For each compound with unspecified stereocenters, all possible stereoisomers were generated using *Omega* from OpenEye. (iii) At most 200 conformers within 10 kcal/mol energy window were generated for each stereoisomer using also *Omega*.

A validation set compound was predicted as active if at least one of its conformers (issued from either of its stereoisomers, if stereochemistry was ambiguous) fit the pharmacophore model. The screening results were ranked with respect to the pharmacophore fit scoring function, which takes into account only chemical features overlap.

Obtained structure-based models were not selective in discrimination of active and inactive compounds. All three models 2VC2, 2VDN and 2VDM returned no hits.

Since automatically developed pharmacophore models displayed little efficacy in virtual screening, an effort has been to modify them manually. Two essential features, the centers of positive and negative charges, which are common in all models, were always kept enabled whereas other features present in the automated models were alternatively toggled on or off. The best models having high precision and recall values are presented on Figure 38 (right). Joint application of these three tuned models resulted in *precision*= 0.81 and *recall*= 0.36. In these experiments, a compound was predicted as active if it fit at least one model of three.

To summarize, the structure-based models are not efficient enough to recover all actives in the database. This could be explained by the fact that they include only 3 or 4 features which probably is not sufficient to fully describe protein-ligand binding.

### 5.3.2. Ligand-based models

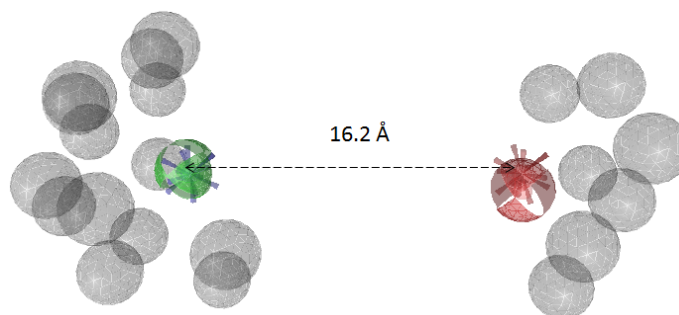
Ligand-based models were built using structures of active compounds from the *Affinity* dataset. A pharmacophore extraction subset was used, and a high activity value  $pIC_{50} \geq 8$  was chosen as a threshold to divide active and inactive compounds. 83 selected active compounds were ionized with *Filter*. For each compound having unspecified stereocenters, all possible stereoisomers were generated and at most 200 conformers within 10 kcal/mol energy window were produced for each stereoisomer using *Omega*.

Because of high flexibility of the ligands, the sets of conformers contained many folded and bend structures, which could seriously distort the obtaining models (Figure 39). To avoid

this problem, generated conformers were pre-filtered based on a simple rule: the distance between positively and negatively charged atoms should be within the interval 16 - 18 Å. An in-house Python script based on OpenEye OEChem [240] has been written for this purpose. Discarding useless conformers resulted in a database containing 1756 3D structures corresponding to 61 compounds. The model built on these structures performs reasonably well in virtual screening (Figure 40). However, this model is extremely simple because it contains only few features and, therefore, it doesn't sufficiently describe the complexity of ligand interactions with the integrin binding site. Indeed, one may distinguish two types of interactions: (i) strong electrostatic interactions via the positively and negatively charged sites conserved in all ligands and, (ii) van der Waals and/or H-bond interactions which may vary from one type of ligands to another. Therefore, we decided to perform a clustering of the dataset, expecting each cluster to correspond to a particular binding mode.

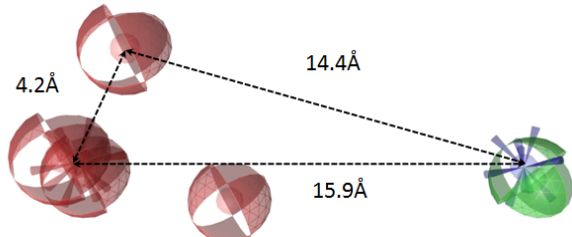
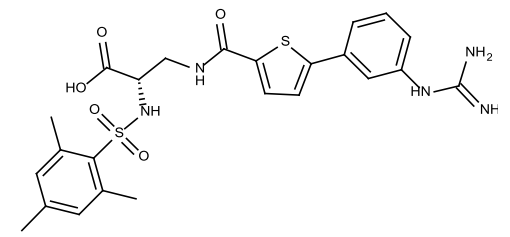
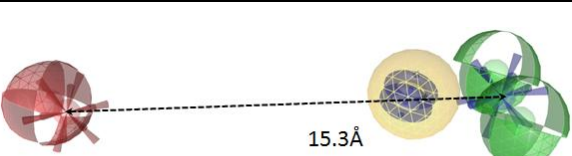
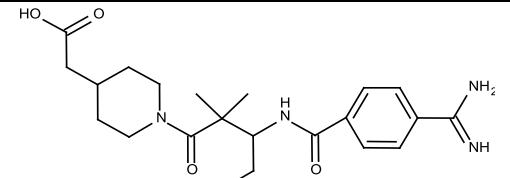
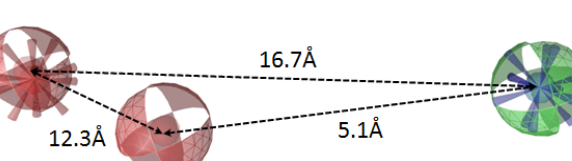
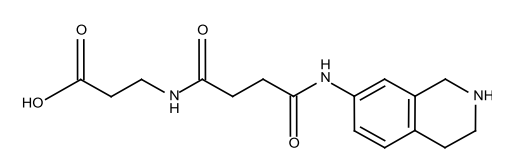
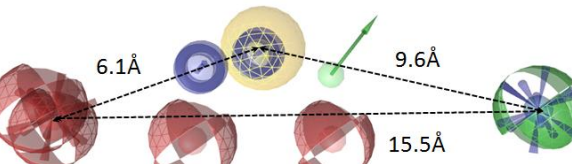
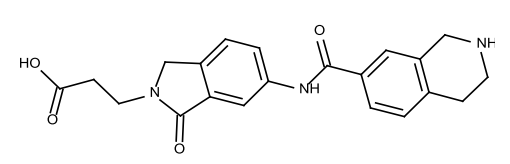
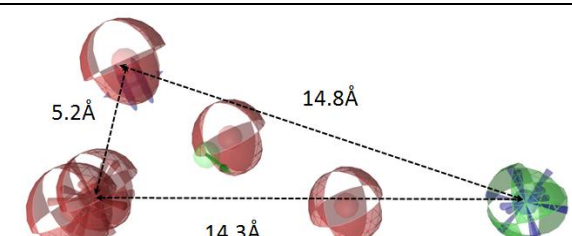
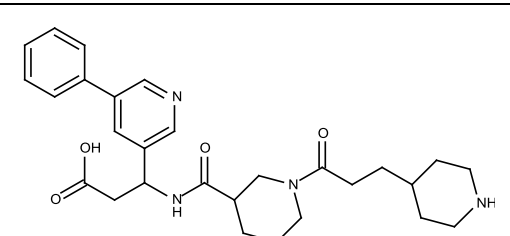
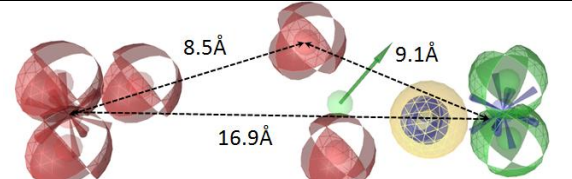
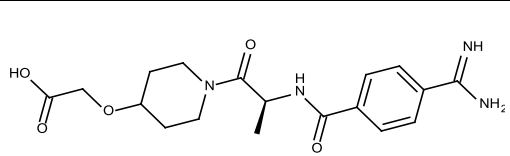
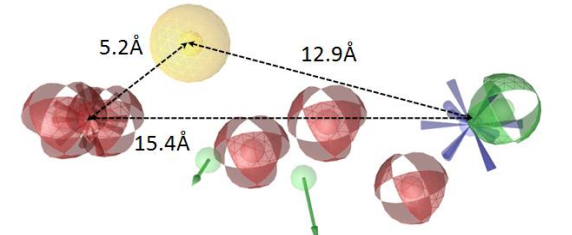
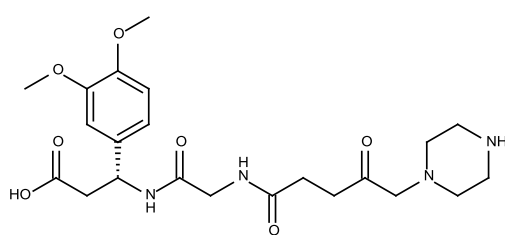


**Figure 39.** Pharmacophore model based on compounds in bend conformations. One can see that the distance between positive and negative features is too small for efficient binding with the receptor.



**Figure 40.** Pharmacophore model built on database containing 1756 3D structures corresponding to 61 compounds. Virtual screening on the remaining part of the *Affinity* set (277 compounds) lead to *Precision* = 0.70 and *Recall* = 0.99.

The clustering performed with LigandScout (default setting) resulted in 7 clusters and 5 singletons. All 7 clusters were taken to prepare shared pharmacophore models with exclusion volumes. Virtual screening with these models resulted to very low recall values (see Figure 41).

№	Pharmacophore models	The most representative compounds
1	 <p>Precision = 0.77, Recall = 0.27</p>	
2	 <p>Precision = 0.67, Recall = 0.29</p>	
3	 <p>Precision = 0.72, Recall = 0.77</p>	
4	 <p>Precision = 0, Recall = 0</p>	
5	 <p>Precision = 0.9, Recall = 0.04</p>	
6	 <p>Precision = 1, Recall = 0.01</p>	
7	 <p>Precision = 1, Recall = 0.01</p>	

**Figure 41.** Shared pharmacophore models (left) and the most representative compounds from each cluster (right). See Figure 18 for the details of feature color-coding.

There are some similarities and differences between generated structure-based and ligand-based models. Both types of models contain positive and negative charges separated by 15.9 - 16.5 Å (in structure-based) or by 14.3-16.9 Å (in ligand-based) which fits experimentally detected distances [126, 216], and one or two hydrophobic features.

Since the performance of the individual models didn't lead to success, we decided to use the ensemble of generated ligand-based models (see Figure 41), which resulted in reasonable *precision* = 0.67 and *recall* = 0.93. Attempt to combine structure-based and ligand-based models didn't lead to any variation of these parameters. Therefore, only ligand-based models issued from Clusters 1-7 have been further used in virtual screening. We believe that this is more reasonable than using a simple model described on Figure 40 because of the possibility to distinguish binding modes of different sub-families of the screened compounds.

Notice, that pharmacophores developed in this work have some characteristics in common with earlier reported models. It concerns the presence of cationic and anionic groups and the hydrophobic feature detected in the models by Cheng [211] and by Ojima [217-218] (Figure 19). Two hydrogen bond acceptors in the Dixon's model (Figure 22) [185], one near the anionic center and another one in the center of the ligand, were also observed in the generated ligand-based models. However, our models developed on larger datasets contain more features, the distances are better tuned and, hence, they are more selective compared to earlier reported pharmacophores.

## 5.4. Docking

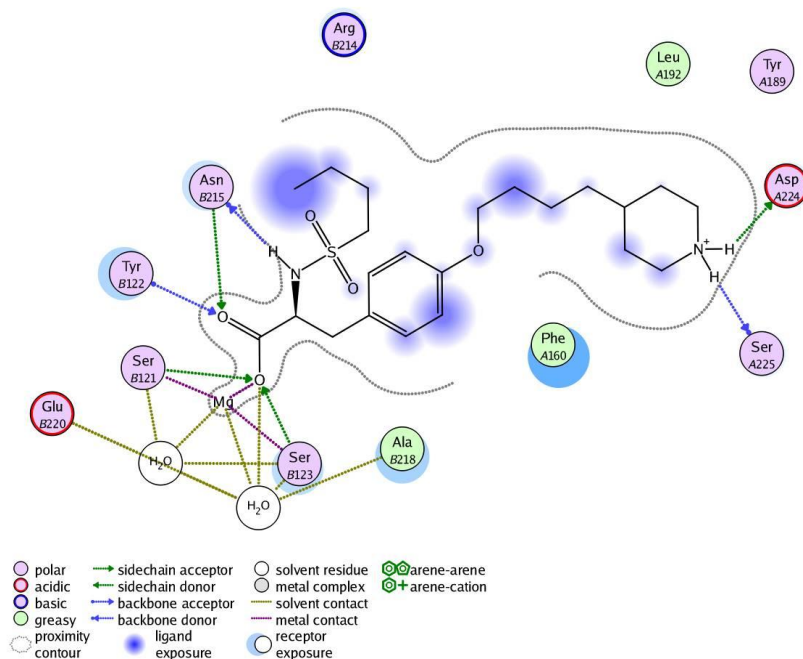
This section describes the molecular docking study. At the first step, we focused on two main tasks: (1) selection and preparation of the binding site of  $\alpha_{\text{IIb}}\beta_3$  receptor and (2) the choice of the appropriate docking program. At the second step, we used this docking setup to study a series of new RGD peptidomimetics recently synthesized and studied in PCI.

### 5.4.1. Choice of the docking software and binding site preparation.

Three complexes (PDB codes 2VC2, 2VDM, 2VDN) of the  $\alpha_{\text{IIb}}\beta_3$  headpiece in its open form, with three different ligands (L-739,758, tirofiban, eptifibatide respectively) were used. The binding pockets in these complexes are quite different. Thus, their pairwise comparison resulted in RMSD (all atoms) = 2.9Å for 2VDM/2VC2, 3.2Å for 2VDM/2VDN and 4.9Å for 2VC2/2VDN. On the other hand, in each complex roughly the same ensemble of amino acid residues is responsible for interactions with the ligand.



Antagonists mainly interact with  $\alpha_{IIb}\beta_3$ Asp224 and  $\alpha_{IIb}\beta_3$ Ser225 residues,  $Mg^{2+}$  metal ion-dependent adhesion site (MIDAS) and  $\beta_3$ Ser121,  $\beta_3$ Asn215 residues [231] (Figure 42). In all complexes, the MIDAS  $Mg^{2+}$  ion has two coordinated water molecules which seem important for binding and correct orientation of the ligand in the protein pocket. Re-docking and cross-docking experiments were performed in order to choose appropriate binding pocket and docking program which provide with an “optimal” ligand pose.



**Figure 42.** Schematic representation of interactions between tirofiban and  $\alpha_{IIb}\beta_3$  receptor in the 2VDM complex

Three programs have been used to carry out the docking studies: PLANTS, FlexX, MOE. For each program the complexes have been prepared individually. Each program uses its own scoring function: CHEMPLP (PLANTS), London  $\Delta G$  (MOE) and empirical (FlexX).

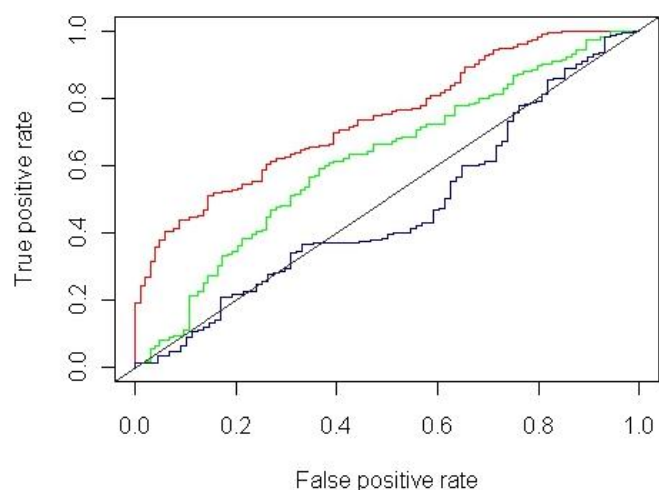
Since there is no evidence about importance of two water molecules coordinating  $Mg^{2+}$ , two docking experiments were performed: with and without these molecules. Although, in both cases we obtained similar RMSD values, results for tirofiban and L-739,758, (more similar to the ligands from our datasets than eptifibatide), indicate that the presence of these water molecules is preferable because it results in more accurate poses with  $RMSD \leq 2\text{\AA}$ . It has also been found that MOE provides better results for tirofiban and L-739,758 ( $RMSD = 1.15\text{-}2.14\text{\AA}$ ) than FlexX ( $1.85\text{-}3.55\text{\AA}$ ) and PLANTS ( $2.07\text{-}8.87\text{\AA}$ ).

Table 9. RMSD values in re-docking and cross-docking studies of complexes 2VC2, 2VDM, 2VDN with MOE, FlexX and PLANTS programs.

Binding site	with water			without water		
	Tirofiban	L-739758	Eptifibatide	Tirofiban	L-739758	Eptifibatide
MOE						
2VDM	<b>1.15</b>	2.14	3.20	<b>0.98</b>	2.16	3.20
2VC2	<b>1.37</b>	<b>1.26</b>	8.70	<b>1.31</b>	6.58	8.13
2VDN	<b>1.72</b>	<b>1.88</b>	3.36	<b>1.69</b>	<b>1.84</b>	8.15
FlexX						
2VDM	3.55	3.39	5.81	3.61	6.13	9.39
2VC2	<b>1.85</b>	2.25	5.15	<b>1.88</b>	9.10	2.25
2VDN	2.52	2.15	8.54	2.53	6.91	2.16
PLANTS						
2VDM	2.07	4.46	5.74	5.89	5.85	2.17
2VC2	8.87	3.49	8.43	2.70	4.06	<b>1.31</b>
2VDN	4.46	4.62	8.90	3.37	4.59	<b>1.23</b>

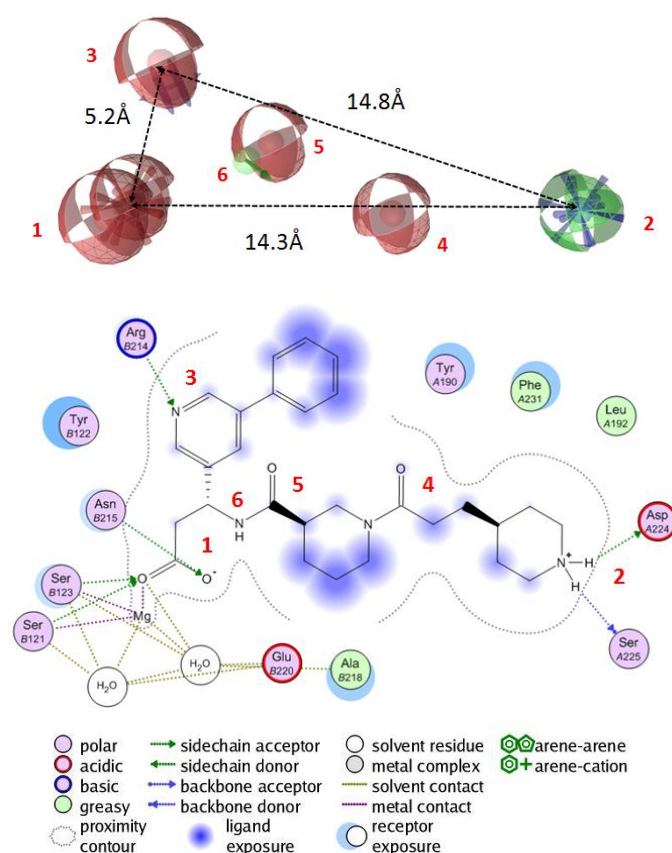
Additional tests were performed in virtual screening the *Affinity* dataset on 2VDM binding pocket by all three programs. For the analysis of the docking results, the scores of stereoisomers were averaged. In order to obtain ROC curves (Figure 43), the compounds from *Affinity* dataset were divided on actives and inactives using  $pIC_{50}=6$  as a threshold. The following ROC AUC values have been obtained: 0.47 (FlexX), 0.62 (PLANTS) and 0.74 (MOE). The enrichment ratio for the 10% subset was 1.44, 1.03 and 1.03 for MOE, FlexX and PLANTS, respectively.

Ensemble of performed tests clearly shows that MOE outperforms two other programs. Thus, 2VDM binding pocket containing the two water molecules at the MIDAS site and the MOE program were chosen for further screening experiments (see section 3.2.3.2. and our publication in *Bioorganic & Medicinal Chemistry Letters* [241]).



**Figure 43.** ROC curves corresponding to screening of the *Affinity* dataset on 2VDM binding pocket by FlexX (blue curve), PLANTS (green) and MOE (red).

Comparing pharmacophore and docking approaches, one can conclude that the protein-ligand interaction patterns described by ligand-based pharmacophore models are consistent with those observed in docking. This is demonstrated on Figure 44 where the shared ligand-based pharmacophore issued from cluster 5 is compared with the docking pose of one of the compound from the same cluster.



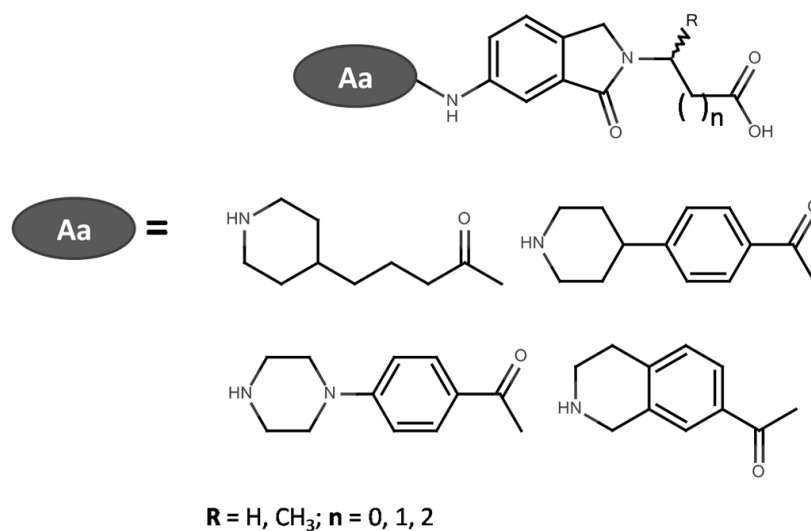
**Figure 44.** Pharmacophore model (top) obtained on cluster 5 and schematic 2D representation of the docking pose of the ligand belonging to the cluster 5. One can see that the pharmacophore model is consistent with most of protein-ligand interactions observed in the docking experiments.

One can clearly see that : (1) the two hydrogen bond acceptors and associated with them negative charged feature correspond to two oxygens of the carboxyl group interacting with  $\beta_3$ Asn215,  $\beta_3$ Ser121,  $\beta_3$ Ser123 and  $Mg^{2+}$ ; (2) the hydrogen bond donor and associated positive charged feature correspond to piperidine ring which interact with  $\alpha_{IIb}$ Asp224,  $\alpha_{IIb}$ Ser225; (3) the hydrogen bond acceptor in the pharmacophore corresponds to hydrogen bond between pyridine and  $\beta_3$ Arg214. For two hydrogen bond acceptor (3, 4) and donor (5), no related interactions were observed in docking. One may suppose that they interact with water molecules which were removed at the stage of binding site preparation.

#### 5.4.2. Docking experiments on selected subsets of RGD peptidomimetics.

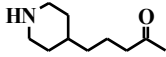
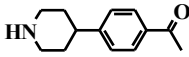
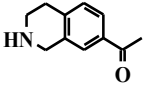
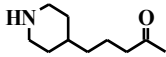
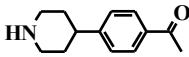
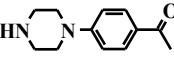
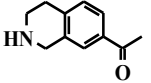
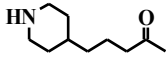
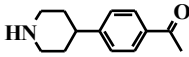
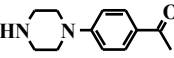
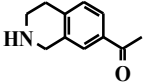
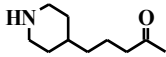
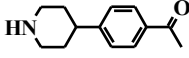
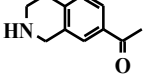
In this section we report docking studies on two series of RGD peptidomimetics recently synthesized and tested on affinity for  $\alpha_{IIb}\beta_3$  and anti-aggregation activity in the Medicinal department of PCI. Our goal was to give a microscopic insight into experimentally observed structure-property relationships.

The first series consisted of 14 RGD peptidomimetics containing a phthalimidine fragment is represented in Table 10 (see Figure 45) [241].



**Figure 45.** RGD peptidomimetics containing a phthalimidine fragment

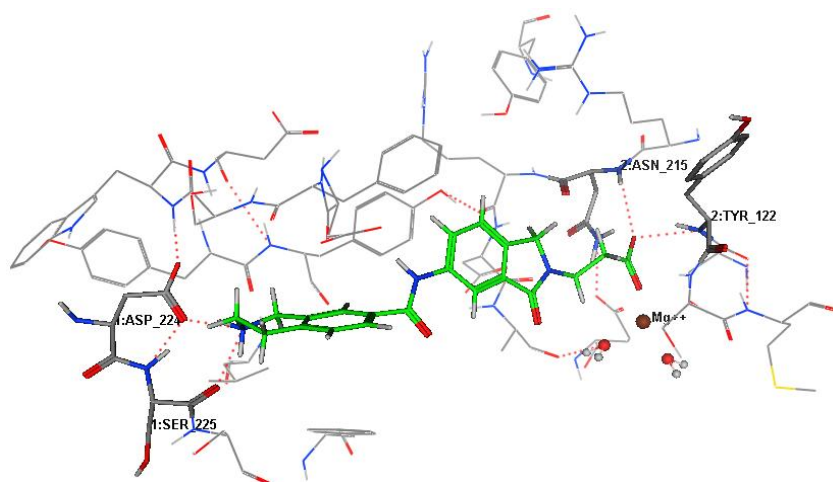
Table 10. Biological properties of first series of RGD peptidomimetics, RGDS peptide and Tirofiban

Compounds	Aa	n	R	IC <sub>50</sub> , μM (PRP) <sup>a</sup>	IC <sub>50</sub> , μM (FITC-Fg/ $\alpha_{IIb}\beta_3$ ) <sup>b</sup>
12a		0	H	66.0±9.0	-
12b		0	H	24.0±3.0	0.27±0.06
12c		0	H	120.0±20.0	1.2±0.1
12d		1	H	5.9±0.6	0.0055±0.009
12e		1	H	9.6±1.9	0.0068±0.0012
12f		1	H	0.54±0.06	-
12g		1	H	1.1±0.1	0.0065±0.0005
12h		1	CH <sub>3</sub>	5.4±1.0	0.35±0.03
12i		1	CH <sub>3</sub>	6.2±1.2	-
12j		1	CH <sub>3</sub>	3.74±0.51	0.037±0.08
12k		1	CH <sub>3</sub>	0.086±0.007	0.0065±0.0012
12l		2	H	51.0±30.0	-
12m		2	H	410.0±60.0	-
12n		2	H	330.0±50.0	-
	<b>RGDS</b>			31.0±2.0	13.0±1.6
	<b>Tirofiban</b>			0.032	

<sup>a</sup>Concentration required to reduce ADP-induced human platelet aggregation response by 50%. The IC<sub>50</sub> values are expressed as the average of at least two determinations. The average error for the IC<sub>50</sub> determinations was 15%.

<sup>b</sup>Concentration required to reduce binding of FITC-Fg to  $\alpha_{IIb}\beta_3$  on the suspension of washed human platelets by 50%. The IC<sub>50</sub> values are expressed as the average of at least two determinations. The average error for the IC<sub>50</sub> determinations was 15%.

The docking study was carried out using setup established in section 3.2.3.1: the protein binding site from the 2VDM complex in combination with MOE. The docking experiments of the synthesized compounds reveal several common binding patterns observed for the most of molecules in the studied dataset. In particular, the nitrogen atom of Arg-isosteres (the fragments of 4-piperidine-4-yl-butyric, 4-piperidine-4-yl-benzoic, 4-piperazine-4-yl-benzoic and 1,2,3,4-tetrahydroisoquinoline-7-carboxylic acids) interacts with two amino acid residues of  $\alpha_{IIb}$  chain of  $\alpha_{IIb}\beta_3$  receptor, namely with carboxyl group of Asp224 side chain and Ser225 amide bond. The carboxyl group of peptidomimetics is involved in coordination sphere of  $Mg^{2+}$ , and also interacts with Tyr122 amide bond and amide group of Asn215 side chain incorporated in the  $\beta_3$  chain, see Figure 46 as an example.



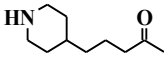
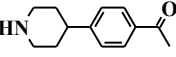
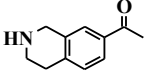
**Figure 46.** Binding of molecule **12g** to the  $\alpha_{IIb}\beta_3$  receptor observed in docking experiments.

Our studies show that S-enantiomers of compounds **12i** and **12j** (see Table 10) bind to the receptor stronger than the corresponding R-enantiomers. On the other hand, no significant differences in values of scoring function for S and R-enantiomers of the compounds **12h** and **12k** were observed.

The second series presents 21 RGD peptidomimetics which are shown in Table 11 [242]. Since the affinity value is available only for a small subset thereof, the anti-aggregation activity was used for investigation of correlation, which may still be apparent if docking predicts affinities well, and affinity is in itself the main parameter controlling anti-aggregation propensity. The docking study performed with MOE didn't show, however, correlation between the docking score and anti-aggregation activity ( $R_{\text{Spearman}} = -0.06$ ). Therefore docking study was repeated with FlexX. The correlation was still poor ( $R_{\text{Spearman}} = 0.35$ ). However, in the case of FlexX exclusion of one outlier significantly improves the correlation ( $R_{\text{Spearman}} = 0.52$ , Figure 47).

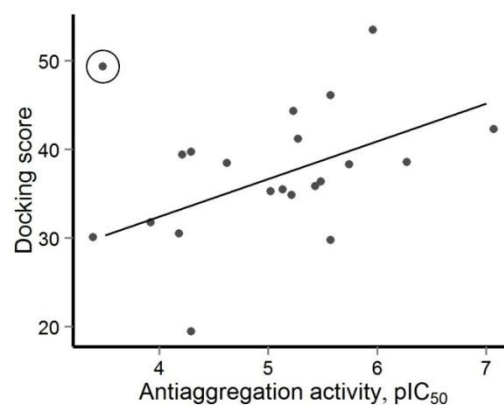
Table 11. Biological properties of the second series of RGD peptidomimetics, RGDS peptide and Tirofiban

Compounds	Aa	N	R	IC <sub>50</sub> , μM (PRP) <sup>a</sup>	IC <sub>50</sub> , μM (FITC-Fg/α <sub>IIb</sub> β <sub>3</sub> ) <sup>b</sup>
13a		0	H	51	
13b		0	H	66.0±9.0	-
13c		0	H	24.0±3.0	0.27±0.06
13d		0	H	3.3	
13e		0	H	120.0±20.0	1.2±0.1
13f		1	H	7.4	
13g		1	H	5.9±0.6	0.0055±0.009
13h		1	H	9.6±1.9	0.0068±0.0012
13i		1	H	0.54±0.06	-
13j		1	H	1.1±0.1	0.0065±0.0005
13k		1	H	2.7	
13l		1	CH <sub>3</sub>	2.7	
13m		1	CH <sub>3</sub>	5.4±1.0	0.35±0.03
13n		1	CH <sub>3</sub>	6.2±1.2	-
13o		1	CH <sub>3</sub>	3.74±0.51	0.037±0.08
13p		1	CH <sub>3</sub>	0.086±0.007	0.0065±0.0012
13q		1	CH <sub>3</sub>		
13r		2	H	1.8	

Compounds	Aa	N	R	IC <sub>50</sub> , μM (PRP) <sup>a</sup>	IC <sub>50</sub> , μM (FITC-Fg/α <sub>IIb</sub> β <sub>3</sub> ) <sup>b</sup>
13s		2	H	51.0±30.0	-
13t		2	H	410.0±60.0	-
13u		2	H	330.0±50.0	-
	<b>RGDS</b>			31.0±2.0	13.0±1.6
	<b>Tirofiban</b>			0.032	

<sup>a</sup>Concentration required to reduce ADP-induced human platelet aggregation response by 50%. The IC<sub>50</sub> values are expressed as the average of at least two determinations. The average error for the IC<sub>50</sub> determinations was 15%.

<sup>b</sup>Concentration required to reduce binding of FITC-Fg to α<sub>IIb</sub>β<sub>3</sub> on the suspension of washed human platelets by 50%. The IC<sub>50</sub> values are expressed as the average of at least two determinations. The average error for the IC<sub>50</sub> determinations was 15%.



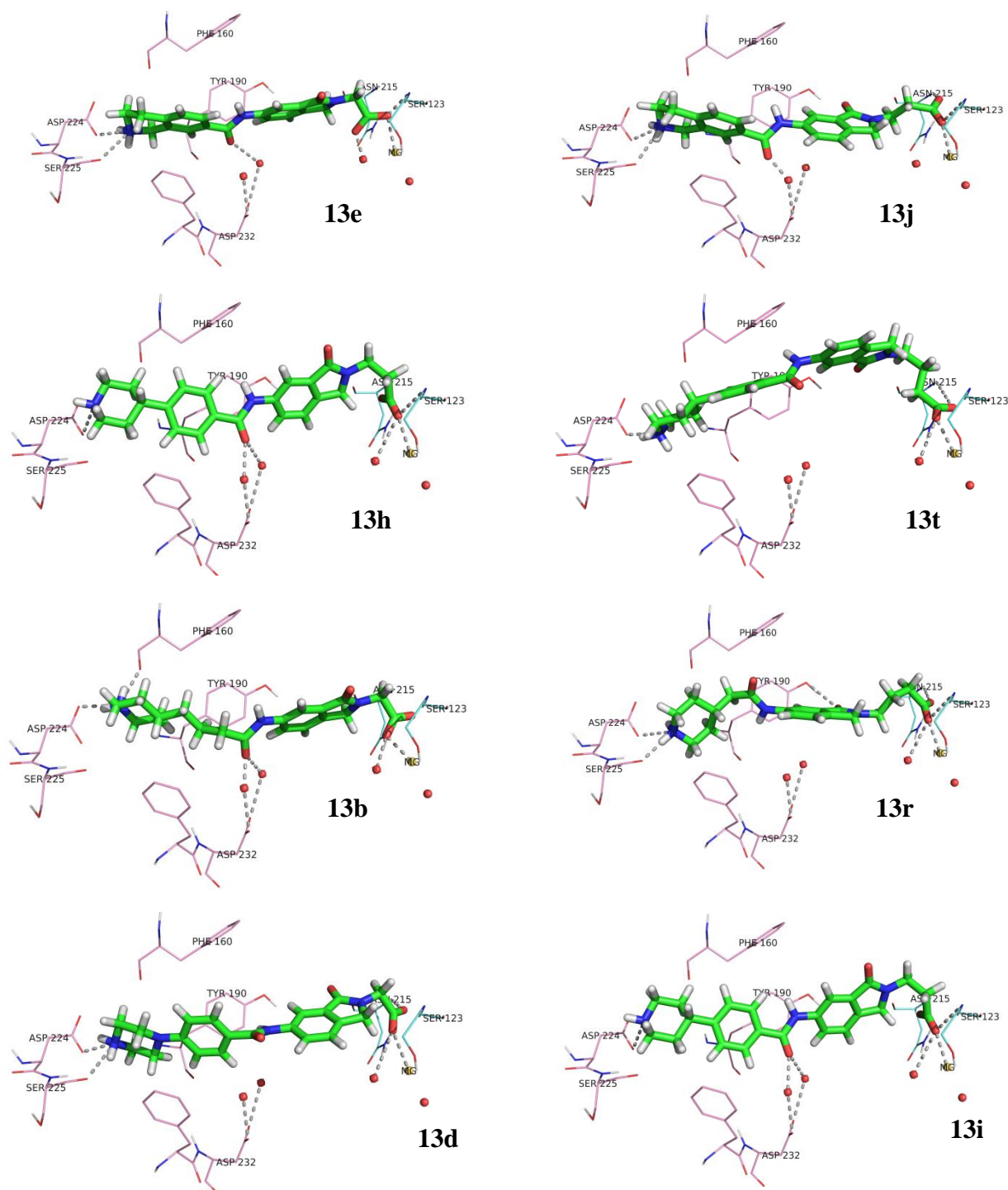
**Figure 47.** Plots of anti-aggregative activity of synthesized compounds against docking score obtained by FlexX. Excluded outlier is shown inside the circle.

Observed on Figure 47 correlation suggests that the docking score obtained by FlexX can be used for the assessment of anti-aggregative activity of the compounds. Thus, higher anti-aggregative potency of **13f-j** compared to **13a-e** could be explained by the fact that the former form additional H-bonds to OH-group of α<sub>IIb</sub>Tyr190 or to water molecules connected with α<sub>IIb</sub>Asp232, whereas the latter can't form hydrogen bonds to OH-group of α<sub>IIb</sub>Tyr190 due to the steric reasons (Figure 48, compounds **13e** and **13j**).

Compounds **13s,t** resulted from the increase of the length of the Asp-mimetic part of **13g,h** are too big compared to the cavity and, therefore they don't interact efficiently with the protein residues (Figure 48, compounds **13h** and **13t**). The difference between activity of the compound **13b** and **13r**, which have the same topological length between Asp- and Arg-mimetic parts, can be explained by different protein-ligand interaction patterns. The carboxylic acid group of **13r** is buried deeper in the receptor pocket and interacts with β<sub>3</sub>Asn215 as well as with



MIDAS, and carbonyl group of 2,3-dihydroisoindol – one moiety forms H-bond with OH-group of  $\alpha_{IIb}$ Tyr190. On the other hand, carboxylic acid group of **13b** interacts only with MIDAS, whereas carbonyl group of 4-piperidin-4-yl-butanoyl forms H-bond with water molecules coordinated to  $\alpha_{IIb}$ Asp232 (Figure 48, compounds **13b** and **13r**). Comparison of the docking pose of **13i** with that of **13d** shows that **13i** is characterized by a bent conformation whereas **13d** adopts an almost linear form. Furthermore, **13i** is characterized by higher docking score value compared to **13d** (38.6 vs 36.4).



**Figure 48.** Docking poses and interactions of compounds **13b,e,h,j,r,t,d,i** inside the  $\alpha_{IIb}\beta_3$  receptor cavity obtained by FlexX.

## 5.5. Molecular shape-based comparison

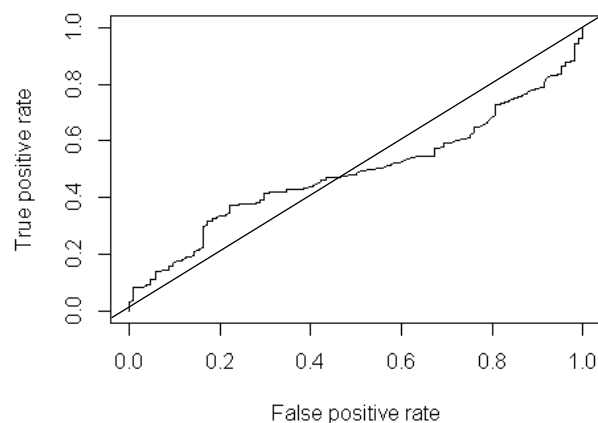
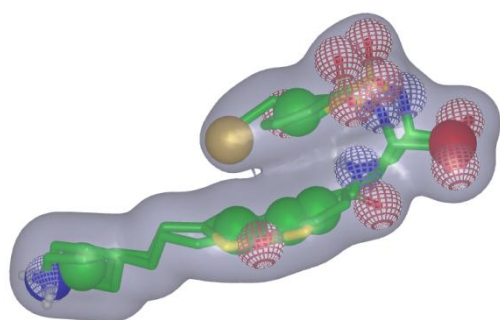
The shape-based superposition method implemented in ROCS package of OpenEye software was used for generation of molecular shape models for the purpose of virtual screening. The developed models were validated using *Affinity* dataset, which contained 234 active and 104 inactive molecules. The validation set has been prepared in the same way as for the pharmacophore models described above. The compound of the validation set was predicted as active if at least one of its stereoisomers fit the shape-based model.

Goodness-of-fit measure of screened compounds has been estimated using TanimotoCombo score which has been averaged among different stereoisomers of the same compound from the validation set. Model quality has been estimated by AUC values calculated from ROC curves.

ROCS is a 3D approach and it requires “bioactive” conformers for model generation. For this reason, we used the ligands from the same complexes as for the docking. But in this case only two of the ligands were chosen for model generation (tirofiban and L-739,758), since there are none representatives in the training set of the third one, eptifibatide, which is a cyclic peptide. Tirofiban and L-739,758 were extracted from their complexes (2VDM and 2VC2), which were preliminary minimized using MMFF94 force field implemented in MOE. Both ligands were ionized using *Filter*.

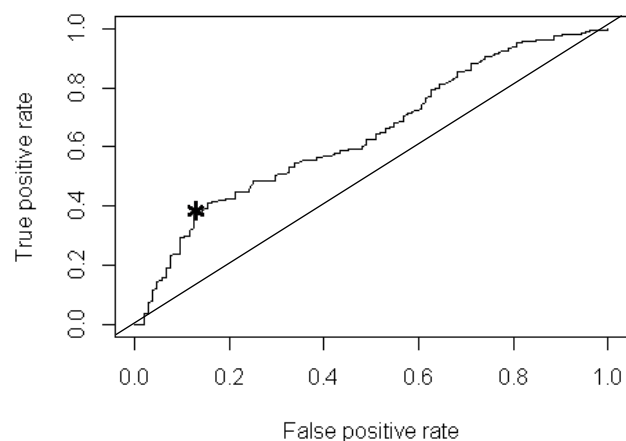
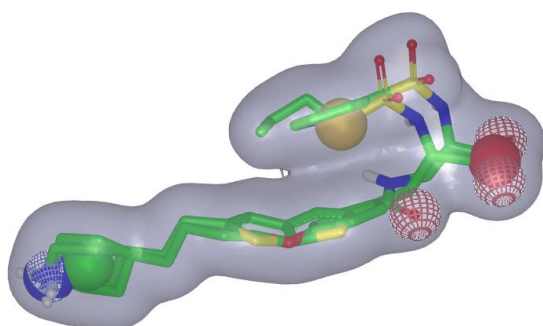
An attempt to obtain individual shape-based models for tirofiban and for L-739,758 was not successful: the ROC AUC value didn't exceed 0.42. Any modifications didn't help to improve the results.

Since the individual models were unsuccessful it was decided to build a new one based on both ligands. For this purpose the ligands have been pre-aligned by the rigid alignment method implemented in MOE. The developed model based on both compounds had also low AUC value. Since actives do not have to comply with the shape constraints extrapolated from the two ligands, the method performed poorly in terms of retrieval of actives – presumably using, in part, different anchoring points to bind to the site (Figure 49). The model is too detailed and therefore too selective: it contains too many chemical features.



**Figure 49.** The initial shape-based model and corresponding ROC curve (AUC = 0.49). The chemical features are color-coded: negative ionizable feature (red sphere), positive ionizable feature (blue sphere), hydrophobic feature (beige sphere), H-bond acceptor (red grid sphere), H-bond donor (blue grid sphere), aromatic ring (green sphere).

A manual “pruning” of this automated model was thus undertaken, in order to keep only the essential anchoring groups (positive and negative charge, hydrogen bond acceptor and hydrophobic center seen in pharmacophore & docking models).



**Figure 50.** The modified shape-based model and corresponding ROC curve (AUC = 0.64). The color-coded see Figure 49. The threshold (TanimotoCombo score = 0.64) for virtual screening corresponds to the point (\*) where the gradient of ROC decreases.

Taking into account this information, several combinations of the most important chemical features were tried. The model which showed the best AUC value (Figure 50) was chosen for further virtual screening. Virtual screening with the TanimotoCombo score = 0.64 as threshold lead to *precision* = 0.85 and *recall* = 0.41.

The selected shape-based model contained such chemical features as positive and negative ionizable features, three hydrogen bond acceptors and one donor, hydrophobic feature and aromatic ring.

## 5.6. Molecular fields similarity

The FieldAlign tool of Cresset was used for comparison of molecular fields of  $\alpha_{1\text{Ib}}\beta_3$  receptor antagonists. The *Affinity* dataset was used as a validation set. For each compound with unspecified stereocenter all possible stereoisomers were generated using *Omega* tool from the OpenEye software. 100 conformations for each compound were generated using the Normal settings of FieldAlign software. The compound of the validation set was predicted as active if at least one of its stereoisomer fit the field-based model.

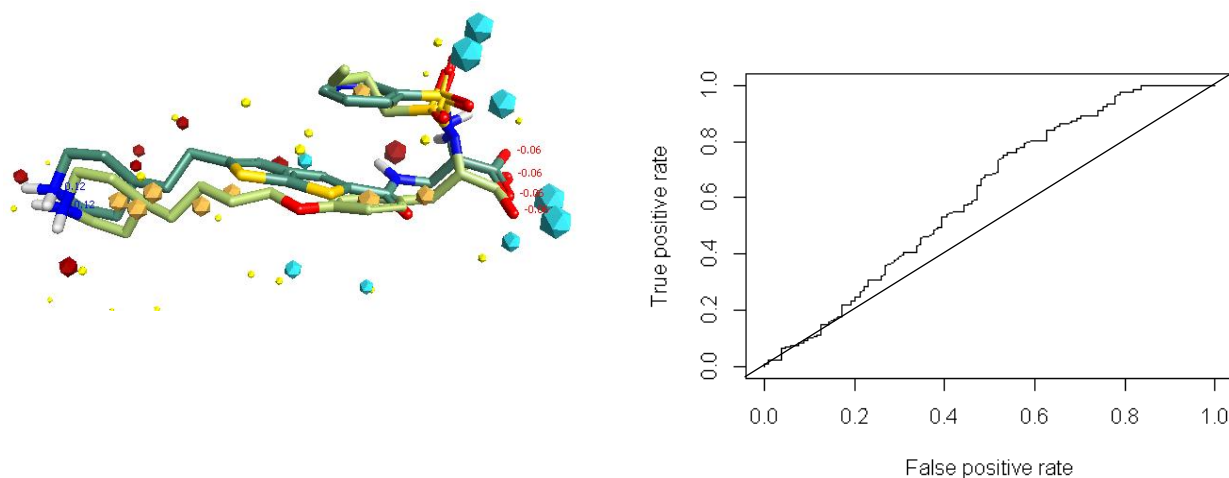
As well as for the ROCS package the proper “bioactive” conformation is required for field generation, so tirofiban and L-739,758 ligands were used for models generation as well as for the shape-based models.

The obtained field similarity value was used to compare the screened compounds with the model and was averaged over different stereoisomers of the same compound from the validation set. Models performance was estimated by ROC AUC values. For this purpose, as in the previous case, the compounds from *Affinity* dataset were divided on two classes using  $\text{pIC}_{50}=6$  as a threshold.

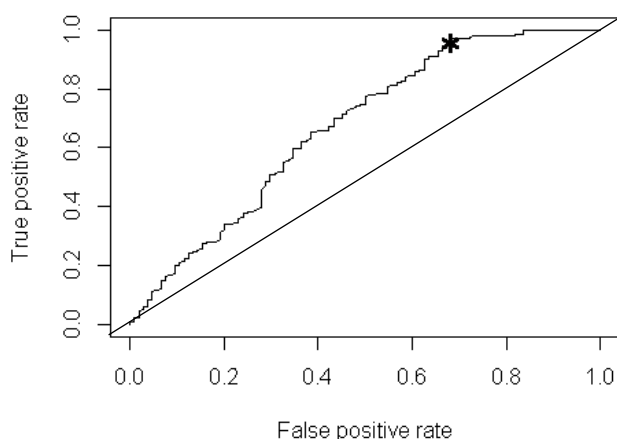
Similarly to shape-based models, the individual models didn't perform well. Any modifications of both models didn't improve their performance. A joint model has been built using information about the ligands from both complexes. The information about protein binding site was used to assign excluded volume features. Screening of the obtained model on validation dataset lead to  $\text{AUC} = 0.61$ .

The FieldAlign software gives user an opportunity to vary (increase or decrease) the value of generated molecular fields. Any molecule with field points which doesn't match the constrained field points will have a penalty applied to its field similarity score. This option was used to tune the initial model. The initial values of fields (constrains) were modified according to previously made observations. Thus, we took into account the importance of the anionic and cationic groups, the hydrogen bond acceptor, which is able to interact with  $\beta_3\text{Asn215}$  amino acid residue, and another hydrogen bond acceptor able to bind to  $\alpha_{1\text{Ib}}\text{Asp232}$ . The weights of the negative, hydrogen bond acceptors and positive field points were increased whereas the weights of the hydrophobic field points were decreased. These modifications lead to acceptable AUC

value (AUC=0.68, Figure 52). The model was chosen for virtual screening. Virtual screening with the similarity value = 0.12 as threshold lead to *precision* = 0.77 and *recall* = 0.89.



**Figure 51.** The initial field-based model and corresponding ROC curve (AUC = 0.61). The molecular fields are represented as points: negative field (blue points), positive field (red), van der Waals surface field (yellow), hydrophobic field (beige).



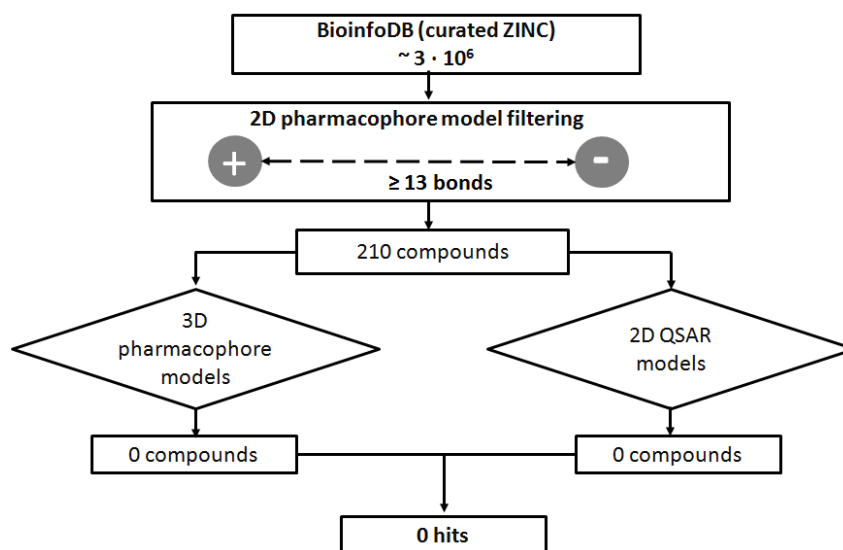
**Figure 52.** The ROC curve for the modified field-based model (AUC = 0.68). The threshold (Similarity value = 0.12) for virtual screening corresponds to the point (\*) where the gradient of ROC decreases.

## 5.7. Virtual screening of databases of commercially available compounds

Here, we will describe virtual screening of BioInfoDB database applying previously developed models.

BioInfoDB [243], containing ~3 million commercially available chemical compounds from different manufacturers, was screened for novel antagonists of  $\alpha_{Ib}\beta_3$  receptors using a multistage procedure comprising all developed models. Compounds were preliminary

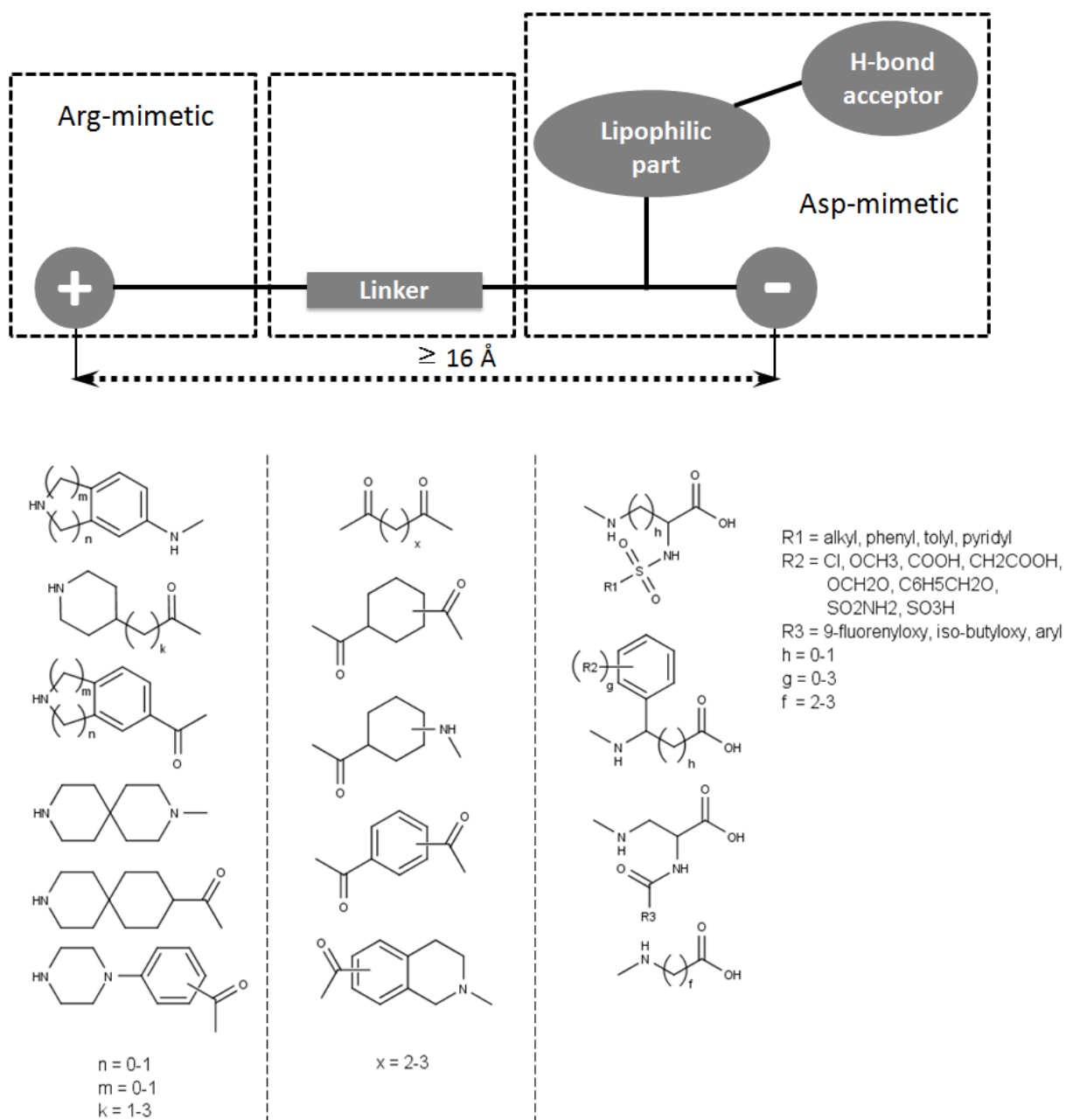
standardized with Chemaxon *Standardizer* and charged with *Filter*. The two-feature 2D pharmacophore model, derived from 3D structure-based pharmacophore model has been used for quickly filtering of standardized compounds. The model takes into account the most relevant features: centers of positive and negative charges separated by not less than 13 bonds (see section 5.1). The latter roughly corresponds to 16Å distance between the centers in the 3D pharmacophore models. 210 compounds fit the 2D model, they were screened by two 3D ligand-based pharmacophore models and 2D QSAR models (see Figure 53). This screening resulted in no hits.



**Figure 53.** Discovering of novel  $\alpha_{IIB}\beta_3$  antagonists.

## 5.8. Design and screening of $\alpha_{IIB}\beta_3$ -focused virtual compound library

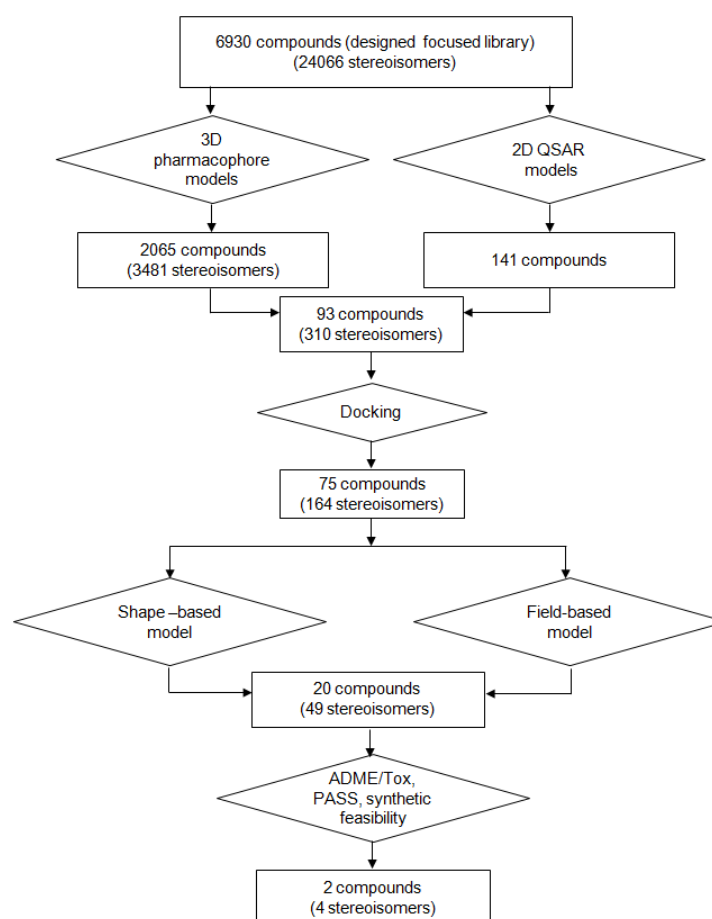
Since screening of BioInfoDB was not successful, a  $\alpha_{IIB}\beta_3$ -focused virtual compound library has been generated. Design of new ligands of  $\alpha_{IIB}\beta_3$  was based mostly on obtained pharmacophore models and docking studies. The following rules for ligand generation were derived: (i) positively and negatively charged groups should be present in the structure of a ligand on a distance at least 16Å; (ii) lipophilic fragments near the basic and acidic part of a molecule are favorable for its higher affinity for  $\alpha_{IIB}\beta_3$ ; and (iii) lipophilic fragments near the acidic part of a molecule should be connected to an acceptor of H-bond (in order to form an H-bond with Arg214 residue of  $\alpha_{IIB}\beta_3$ ). According to these rules various Arg- and Asp-Phe-mimetic fragments and different linker groups (see Figure 54 ) were used to generate combinatorial virtual library with in-house computer program. After exclusion of synthetically irrelevant structures, 6930 structures have been retained for the screening.



**Figure 54.** Examples of fragments used for  $\alpha_{\text{IIB}}\beta_3$  focused library generation.

Since some information issued from the pharmacophore models was implicitly used for library generation, significant number of structures (2065) fit the 3D ligand-based pharmacophore models (Figure 55). Screening with consensus QSAR models resulted in 141 compounds corresponding to the specified threshold values: for affinity  $\text{pIC}_{50} \geq 8.0$ , for anti-aggregation activity  $\text{pIC}_{50} \geq 7.0$ . 310 individual stereoisomers of 93 compounds were docked into the  $\alpha_{\text{IIB}}\beta_3$  receptor pocket (2VDM), and 164 stereoisomers of 75 compounds with the highest values of scoring function were selected for further consideration. The selected 164 stereoisomers were screened with previously obtained shape- and field-based models (Figure 50

and Figure 52), the screening resulted in 49 stereoisomers of 20 compounds. For these structures some ADME/Tox properties (water solubility [160-161], mutagenicity in Ames test [155]) have been predicted using earlier reported QSAR models whereas their polypharmacological profiles were estimated with the PASS program. None of these compounds were predicted as mutagen, all compounds had moderate predicted values of solubility in water  $10^{-2}$ - $10^{-5}$  mol/l, which are much higher than predicted  $pIC_{50}$  values of anti-aggregation activity. The PASS predictions were used to exclude compounds with high probability of other biological responses. The compounds were filtered out if the difference between  $P_a$  and  $P_i$  of possible adverse effects or toxicity exceeded 0.6. Two hits which were represented by S enantiomers (molecules I and II, Table 12) have been finally chosen for their synthesis and biological evaluation considering their synthetic feasibility.



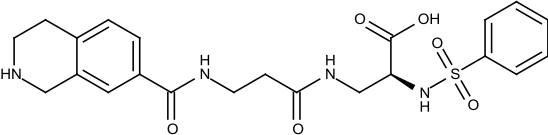
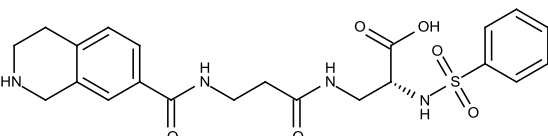
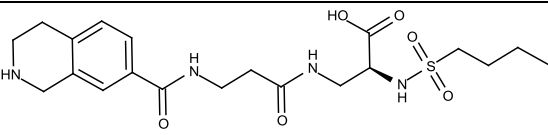
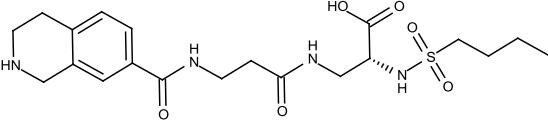
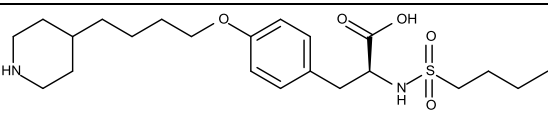
**Figure 55.** Workflow of the virtual screening of the designed  $\alpha_{IIB}\beta_3$ -focused library.



## 5.9. Experimental validation

Both S- and R-enantiomers of suggested theoretically molecules I and II have been synthesized and tested at PCI. The stereoisomers of compound I display a high affinity for  $\alpha_{IIb}\beta_3$  and anti-aggregation activity, even higher than predicted values. Docking correctly differentiated the stereoisomers: the more active compound **I-S** has higher docking score than **I-R**.

Table 12. Affinity for  $\alpha_{IIb}\beta_3$  and anti-aggregation activity of designed compounds and reference compound tirofiban.

	Compound	Affinity for $\alpha_{IIb}\beta_3$ , pIC <sub>50</sub>		Anti-aggregation activity, pIC <sub>50</sub>		Docking score
		pred	obs	pred	obs	
I-S		8.32	9.66	7.37	8.21	61.9
I-R		8.32	9.02	7.37	7.60	60.0
II-S		8.24	7.21	7.27	6.49	62.4
II-R		8.24	7.1	7.27	6.17	59.5
Tirofiban			8.62		7.48	

## 5.10. Conclusions

This section describes computer-aided design of new potent antagonists of the open form of integrin  $\alpha_{IIb}\beta_3$ . At the first stage, different types of models have been developed using experimentally available information on affinity for  $\alpha_{IIb}\beta_3$  and anti-aggregation activity, and

structure of protein-ligand complexes. This concerns: QSAR models (CF), structure-based and ligand-based 3D pharmacophore models (LigandScout), 2D pharmacophore model, shape-based model (ROCS) and molecular field-based model (FieldAlign, Cresset). Application of these models to virtual screening of BioInfoDB database resulted in no hits. Large, zwitterionic compounds are underrepresented in commercial databases – therefore, a small virtual combinatorial library of 6930 compounds has been generated from the list of preselected fragments. Its screening using the above models and ligand to protein docking (MOE) leads to several hits, 4 of which were synthesized and experimentally studied in PCI. One of these compounds is found more active than tirofiban, a commercial drug molecule.

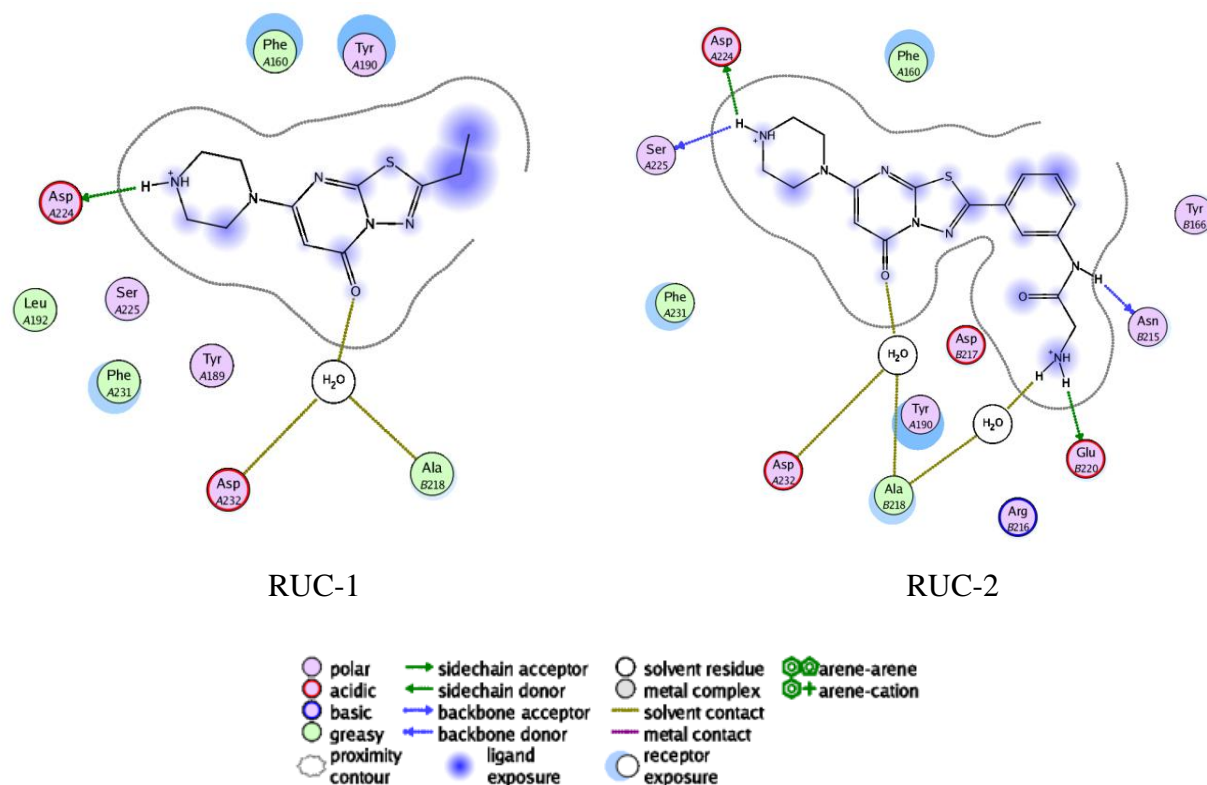
*Chapter 4*

**Development of RUC-1 and RUC-2 analogs -  
antagonists of  $\alpha_{IIb}\beta_3$  receptor**

## 6. Development of RUC-1 and RUC-2 ligands

All commercialized antagonists of  $\alpha_{IIb}\beta_3$  integrin (abciximab, eptifibatide and tirofiban) target the receptor only in its open form. Intravenous injection of these drugs, administered during percutaneous coronary intervention, reduce the risk of death and risk of myocardial infarction for the period from 30 days to six months [244]. On the other hand, it increases a risk of severe bleeding [244] and thrombocytopenia [245]. It has been suggested that these undesired side effects are associated with conformational transformations of the receptor caused by the binding of the antagonists [246-248].

Recently, RUC-1, a novel antagonist of  $\alpha_{IIb}\beta_3$  receptor, has been discovered experimentally in the screening of 33264 compounds [58]. According to mutagenesis studies RUC-1 binds only to the  $\alpha_{IIb}$  subunit of the integrin, this contrasts with tirofiban, eptifibatide and other  $\alpha_{IIb}\beta_3$  antagonists which coordinate to MIDAS  $Mg^{2+}$  ion in the  $\beta_3$  subunit. Also RUC-1 doesn't induce transition to an open headpiece form of the integrin as determined by gel filtration and dynamic light scattering studies.



**Figure 56.** Interactions of RUC-1 and RUC-2 ligands with  $\alpha_{IIb}\beta_3$  integrin in the X-ray structures of corresponding complexes 3NIF and 3T3M (pictures prepared with MOE).

Later on, these observations were confirmed by X-ray structure of the complex of RUC-1 with  $\alpha_{IIb}\beta_3$ . RUC-1 has weak inhibition potency of ADP-induced aggregation tested on human

platelet rich plasma ( $IC_{50} = 13 \pm 5 \mu M$ ) [1]. In order to explore the  $\alpha_{IIb}\beta_3$  binding pocket and to obtain additional information on the links between the binding mechanism and the induction of conformational changes in the receptor, a series of derivatives of RUC-1 guided by structural considerations have been synthesized. One of discovered compounds, RUC-2, was more than 100 times more potent than RUC-1 in inhibiting ADP-induced platelet aggregation ( $IC_{50} = 0.096 \pm 5 \mu M$ ) [2]. RUC-2 binds to Asp224 residue of the  $\alpha_{IIb}$  subunit similarly RUC-1, but additionally RUC-2 binds to Glu220 residue of the  $\beta_3$  subunit and thus it displaces  $Mg^{2+}$  ion from MIDAS (Figure 56). At the same time RUC-2 doesn't induce change in conformational state of the  $\alpha_{IIb}\beta_3$  headpiece.

The goal of this study is the development of analogs of RUC-1 and RUC-2 with improved anti-aggregation potency. This can be achieved by two ways: i) virtual screening of commercial databases; ii) design of new compounds and their virtual screening. Since only information about structures of complexes is available, docking and structure-based approaches are applied.

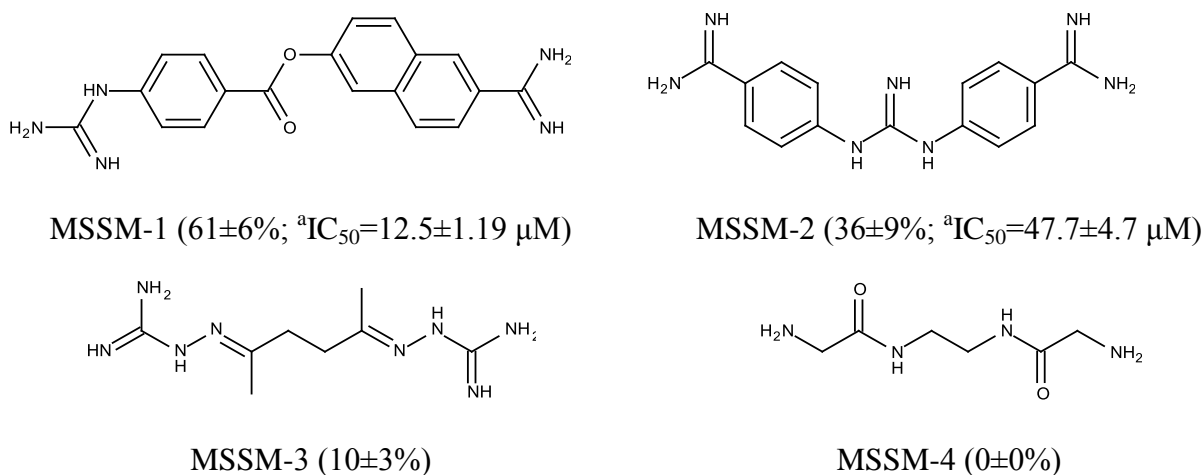
### 6. 1. Previously reported *in silico* screening study

Negri et al. [249] performed a structure-based *in silico* screening study aimed to identify  $\alpha_{IIb}\beta_3$  antagonists which have the similar binding mode with RUC-2 ligand. Since at that moment the X-ray structure of RUC-2 wasn't available, the screening was performed using the binding pocket extracted from the crystal structure of the complex of RUC-1 with  $\alpha_{IIb}\beta_3$  (PDB: 3NIF). When the crystal structure of the RUC-2- $\alpha_{IIb}\beta_3$  complex (PDB: 3T3M) became available, the authors performed an additional screening using this structure, but no additional hits have been discovered.

A subset of ZINC database containing over 2.5 million commercially available compounds was docked into the  $\alpha_{IIb}\beta_3$  pocket (PDB: 3NIF) using DOCK3.5.54 program. All non-protein atoms, except for the adjacent to MIDAS (ADMIDAS) and the synergetic metal binding site (SyMBS) metal ions, were removed from the pocket and the most probable protonation state at pH 7.4 was assigned to the ionizable protein residues. The protein was kept rigid, the ligands flexible. An energy-based score corresponding to the sum of the receptor-ligand electrostatic and van der Waals interaction energies, corrected for ligand desolvation, was used to evaluate the docking of each molecule of the ligand set into the  $\alpha_{IIb}\beta_3$  binding site.

The 500 top-scoring docking hits were visually inspected. Four small molecules (Figure 57) were extracted based on the following criteria: (1) The presence of hydrogen bond interactions between the ligand and both the  $\alpha_{IIb}$ Asp224 and the  $\beta_3$ Glu220 residues, or as an alternative, the  $\alpha_{IIb}$ Asp232 and  $\beta_3$ Glu220 residues; (2) chemotype diversity; and (3)

purchasability. Two compounds (MSSM-3 and MSSM-4) were removed since they produced small inhibition of platelet adhesion to fibrinogen. Remaining compounds (MSSM-1 and MSSM-2) were active enough in the platelet adhesion assay and therefore, they were further tested as inhibitors at the initial wave of ADP-induced platelet aggregation of citrated PRP.

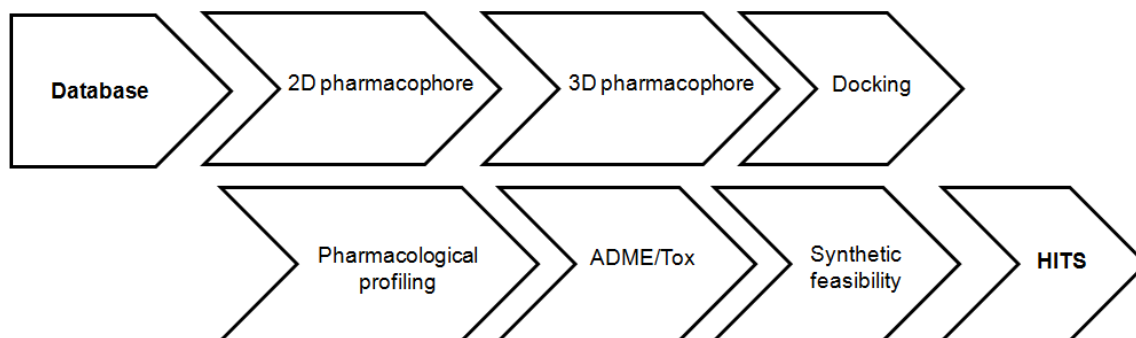


**Figure 57.** Selected compounds from ZINC database. <sup>a</sup>Ability to inhibit the initial wave of ADP-induced platelet aggregation of citrated PRP. Inhibition of adhesion is given between parenthesis.

According to docking experiments, these compounds interact with key interaction sites of RUC-2, including  $\alpha_{\text{IIb}}\text{Asp224}$ ,  $\alpha_{\text{IIb}}\text{Tyr190}$  and  $\beta_3\text{Glu220}$ . One of the compounds, MSSM-1, has been declared as a potent serine protease inhibitor [250], in 2005 it was licensed as Nafamostat in Korea, where citrate anticoagulation in CRRT is unavailable [251]. Nafamostat has been reported to inhibit platelet aggregation induced by ADP ( $\text{IC}_{50}=9.3\pm 2.8\ \mu\text{M}$ ) [252-253]. Based on the experimental data Negri et al. concluded that MSSM-1 and MSSM-2, similarly to RUC-1 and RUC-2, do not induce major conformational changes in the receptor. Therefore, they represent promising starting points for structure-based ligand optimization.

## 7. Computer-aided design of antagonists of $\alpha_{IIb}\beta_3$ receptor found in closed form

This section describes the search of novel antagonists of  $\alpha_{IIb}\beta_3$  receptor found in closed form using a workflow shown on Figure 58.

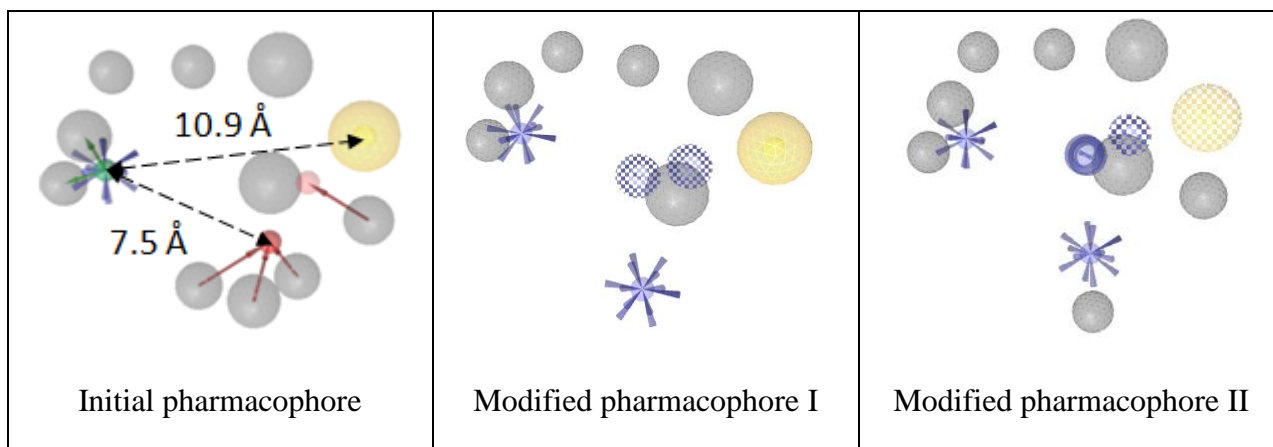


**Figure 58.** Virtual screening workflow used to discover binders of the closed form of  $\alpha_{IIb}\beta_3$  receptor.

### 7.1. Pharmacophore models

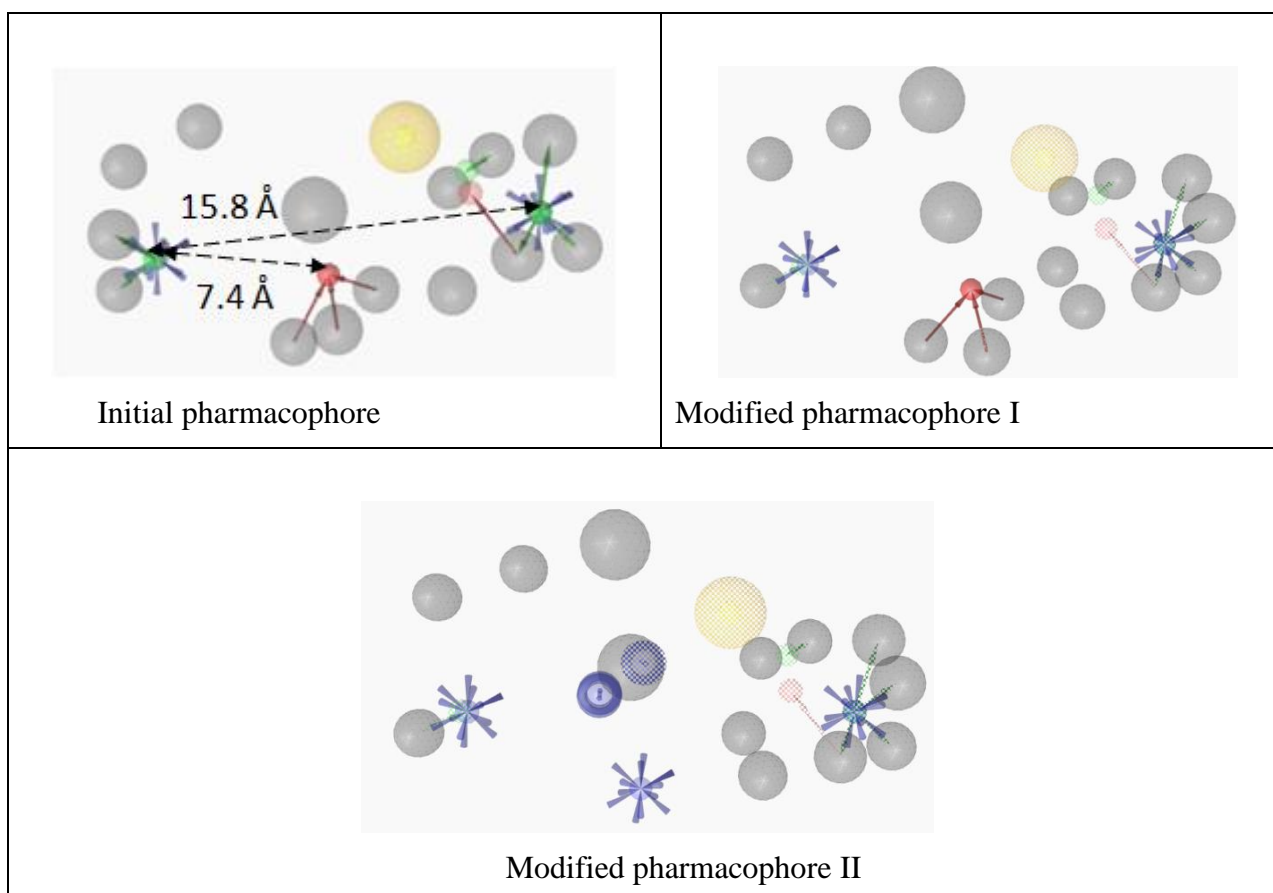
The structure-based pharmacophore models have been produced by LigandScout, based on the structures of complexes RUC-1 and RUC-2 with  $\alpha_{IIb}\beta_3$  integrin (PDB codes are 3NIF and 3T3M, respectively). The initial structure-based model for RUC-1 contained eight features (see Figure 59) : 1) the positive ionizable area associated with two H-bond donors directed to Asp224 of the  $\alpha_{IIb}$  subunit of the receptor, 2) three H-bond acceptors associated with carbonyl group of the ligand, which binds to Asp232 residue via water molecules, 3) H-bond acceptor associated with aromatic nitrogen atom, which doesn't have any specific binding point, 4) hydrophobic area.

Since RUC-1 has a weak anti-aggregation activity ( $IC_{50} = 13 \pm 5 \mu M$ ) there is no sense to develop direct analogs. That's why the initial pharmacophore model was modified. The acceptor features, representing binding to Asp232 *via* water molecules, have been deleted and the center of positive charge has been added to the model in order to replace those water molecules. All other H-bond donor and acceptor features were removed. Two aromatic rings features have been added to the pharmacophore, since a heterocyclic moiety of RUC-1 ligand may interact with the complementary Tyr190 residue. The importance of Tyr190 residue for RUC-1 binding was confirmed by point mutagenesis studies [58]. Thus, two derivatives of initial model were developed (Figure 59).



**Figure 59.** RUC-1 structure-based pharmacophore models. For details of feature color-coding see Figure 18.

The initial RUC-2 structure-based pharmacophore model contained: 1) two positive ionizable areas with approximately 15.8Å between them, 2) five H-bond donors associated with positive ionizable areas, two of them are directed to  $\alpha_{IIb}$ Asp224 amino acid and three – to  $\beta_3$ Glu220, 3) three H-bond acceptors associated with carbonyl group of the ligand, which binds to  $\alpha_{IIb}$ Asp232 residue via water molecules, 4) H-bond donor group bounded with  $\beta_3$ Asn215, 5) H-bond acceptor in C-terminal part of the molecule, 6) hydrophobic area (Figure 60).



**Figure 60.** The initial RUC-2 structure-based pharmacophore model and its derivatives. For details of feature color-coding see Figure 18.



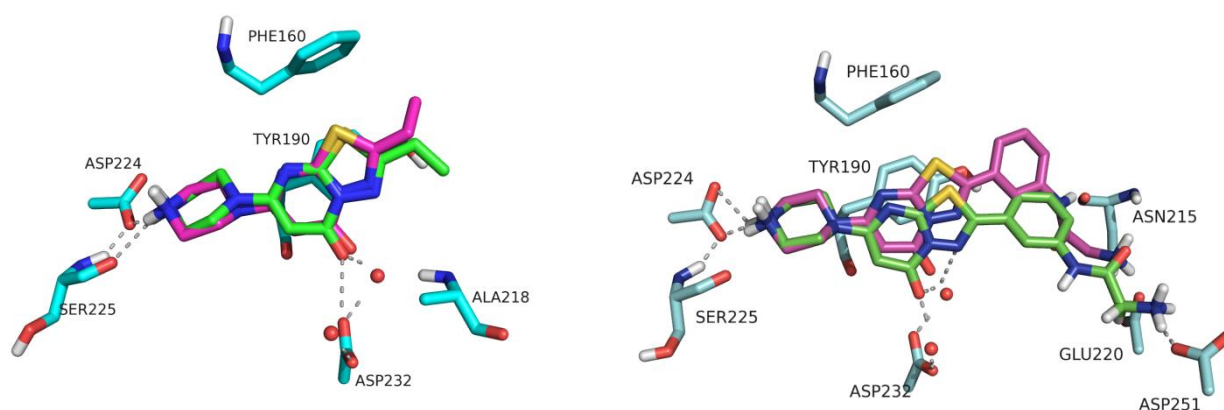
Ideas similar to those used for modification of the initial RUC-1 pharmacophore model were applied for the initial RUC-2 structure-based pharmacophore model: 1) two aromatic ring features were added, 2) a center of positive charge was added to replace water molecules bound to Asp232 residue, 3) some features which seem not so important for ligand binding were set as optional or were removed. Overall, two new models were produced according to the mentioned rules and they were used for virtual screening along with the initial RUC-2 pharmacophore model.

## 7.2. Ligand to Protein docking

This section describes ligand to protein docking. Two tasks have been solved: 1) preparation of the binding site of  $\alpha_{IIb}\beta_3$  receptor found in closed form; 2) selection of the appropriate docking program.

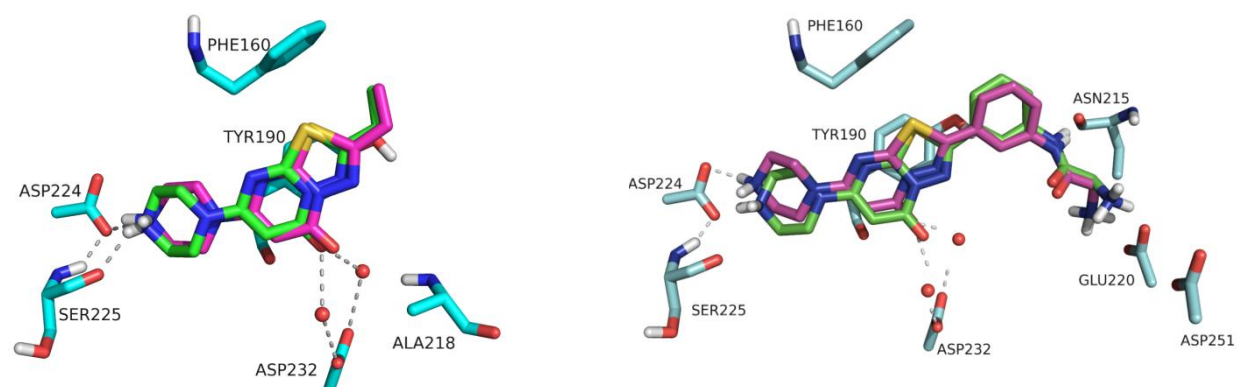
Two programs have been used to carry out the docking study: FlexX and MOE. For each program binding sites of 3NIF and 3T3M complexes have been prepared. Important water molecules that stabilize the ligand in the binding site have been remained. For the 3NIF complex with RUC-1 as well as for the 3T3M complex with RUC-2, two water molecules that make hydrogen bonds with Asp232 residue have been kept. Re-docking was performed in order to choose the appropriate binding pocket and docking program which provide with an “optimal” ligand pose.

Native ligands have been re-docked in the corresponding binding sites to reproduce their poses. The re-docking for RUC-1 ligand with MOE gave good results, the program reproduced the X-ray pose (RMSD=0.78Å). In case of re-docking of RUC-2 MOE wasn't able to reproduce the pose: RMSD value was quite high 2.2Å and no interaction with  $\beta_3$ Glu220 residue were observed.



**Figure 61.** The re-docking of RUC-1 (left) and RUC-2 (right) with MOE. The X-ray ligand is represented in purple, docked ligand – in green, the receptor amino acid is shown in turquoise.

With FlexX, the above-mentioned kept water molecules have been set as freely rotatable. The re-docking results had been successful for both ligands (RMSD between original pose of RUC-1 and RUC-2 and their docked poses was 0.65Å and 0.78Å, respectively). Since the re-docking results were better with FlexX, this tool was chosen for virtual screening.



**Figure 62.** The re-docking of RUC-1 (left) and RUC-2 (right) with FlexX. The X-ray ligand is represented in purple, docked ligand – in green, the receptor amino acid is shown in turquoise.

### 7.3. Virtual screening of commercially available compounds

Three databases have been used for virtual screening in the current study: i) advanced and HTS Enamine databases, which contain collection of 1.5M structurally diverse compounds; ii) REAL Enamine database, which contains ~17M of synthetically feasible compounds; iii) ZINC database, which comprises collections of compounds from different vendors with overall more than 17M compounds. The compounds from the selected databases were ionized using *Filter* tool from OpenEye software.

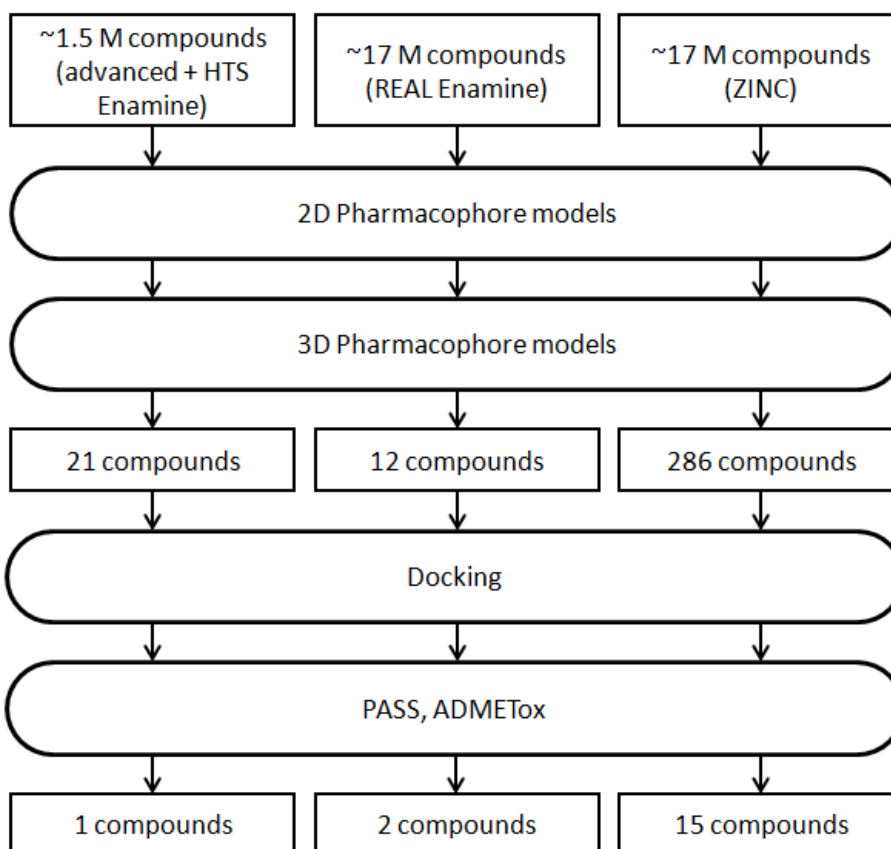
For faster screening, three prepared databases have been filtered by simplified 2D pharmacophore models developed for RUC-1 and for RUC-2 ligands. 2D models included two the most restrictive features of corresponding 3D pharmacophore models and the distance between those features translated to the minimal number of bonds, which could connect chosen features. For RUC-1, the 2D pharmacophore model included two centers of positive charge, which were not tertiary nitrogens and not aromatic atoms, at a distance of at least of 7 bonds (that should cover 8Å distance in 3D). For RUC-2 2D, two remote centers of positive charges, which were not tertiary nitrogens and not aromatic atoms, were set at more than 12 bonds apart (15.9Å in 3D).

The compounds fitting 2D pharmacophore models have been filtered by corresponding 3D pharmacophore models (Figure 59, Figure 60). For this purpose all remaining compounds were aromatized, the explicit hydrogens were added; solvents and small fragments were removed using ChemAxon *Standardizer*. For the compounds with unspecified stereochemistry all possible stereoisomers were generated using *Omega* tool from OpenEye software and at most 200 conformers for each stereoisomer with energy window 10 kcal/mol were produced by *Omega*.

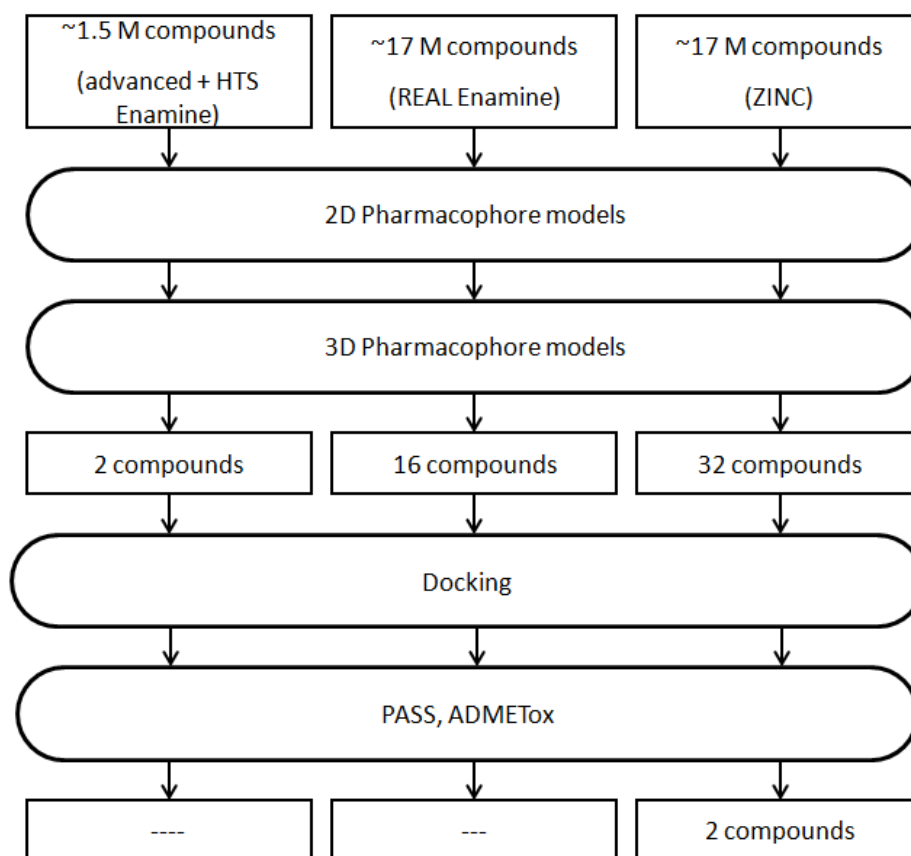
The compounds, which passed 3D pharmacophore filter, have been docked by FlexX. Since the screening was aimed to find the compounds which should interact with  $\alpha_{\text{IIB}}\text{Asp232}$  residue directly, water molecules were removed during the docking procedure.

To choose the most promising compounds out of well-docked virtual hits, the spectrum of their potential pharmacological effects has been predicted with PASS (the compounds were filtered out if the difference between  $P_a$  and  $P_i$  of possible adverse effects or toxicity exceeded 0.6) and some of their ADME/Tox properties like aqueous solubility and mutagenicity have been predicted by earlier published QSAR models.

As a result of virtual screening 18 analogs of RUC-1 and 2 analogs of RUC-2 ligands were found.



**Figure 63.** Screening workflow of RUC-1 analogs



**Figure 64.** Screening workflow of RUC-2 analogs

One of the remained compounds has been identified as already known drug – *Nafamostat*, a synthetic serine protease inhibitor [254]. This compound was also found by Negri et al. in their structure-based virtual screening [249] (section 6.1).

*Nafamostat* possesses activity to a lot of proteins, such as thrombin, urokinase, trypsin, plasmin et al.[255-258] (see Table 13) It is noteworthy that *Nafamostat* was introduced as an alternative anticoagulant in continuous renal replacement therapy (CRRT) in 1990, but its usage has been mainly limited to Japan [259-260]. In 2005, *Nafamostat* was licensed in Korea where citrate anticoagulation in CRRT is unavailable [251].

Table 13. Several known inhibitory activities of *Nafamostat* shown in reference [255]

Proteins	K <sub>i</sub> , M
urokinase	$1.92 \cdot 10^{-8} \pm 1.02 \cdot 10^{-9}$
t-PA	$1.92 \cdot 10^{-8} \pm 1.02 \cdot 10^{-9}$
kallikrein	$1.23 \cdot 10^{-8} \pm 2.11 \cdot 10^{-9}$
trypsin	$1.62 \cdot 10^{-8} \pm 2.42 \cdot 10^{-9}$
prot Ca	$1.14 \cdot 10^{-7} \pm 3.38 \cdot 10^{-8}$
F XIIa	$1.05 \cdot 10^{-7} \pm 1.25 \cdot 10^{-8}$
plasmin	$3.74 \cdot 10^{-6} \pm 3.47 \cdot 10^{-7}$
thrombin	$1.44 \cdot 10^{-5} \pm 3.90 \cdot 10^{-6}$
F Xa	$1.15 \cdot 10^{-4} \pm 6.53 \cdot 10^{-5}$

None of the remained compounds after screening procedure has been found in the catalogs of the companies for purchase, therefore it was decided to design the analogs of RUC-1 and RUC-2 taking into consideration the observation made during the modeling.

#### 7.4. Design of focused libraries

Since the virtual screening of databases with commercially available compounds was not successful, we designed two small virtual libraries of analogs of RUC-1 and RUC-2 ligands. During the design of analogs of RUC-1 and RUC-2, information obtained from mutagenesis studies, developed structure-based 3D pharmacophore models and molecular docking.

The following considerations have been used for modification of RUC-1 and RUC-2 structures and for design of their analogs (Figure 65):

i) a designed ligand should comprise positively charged group, which can interact with  $\alpha_{\text{IIb}}\text{Asp224}$  residue;

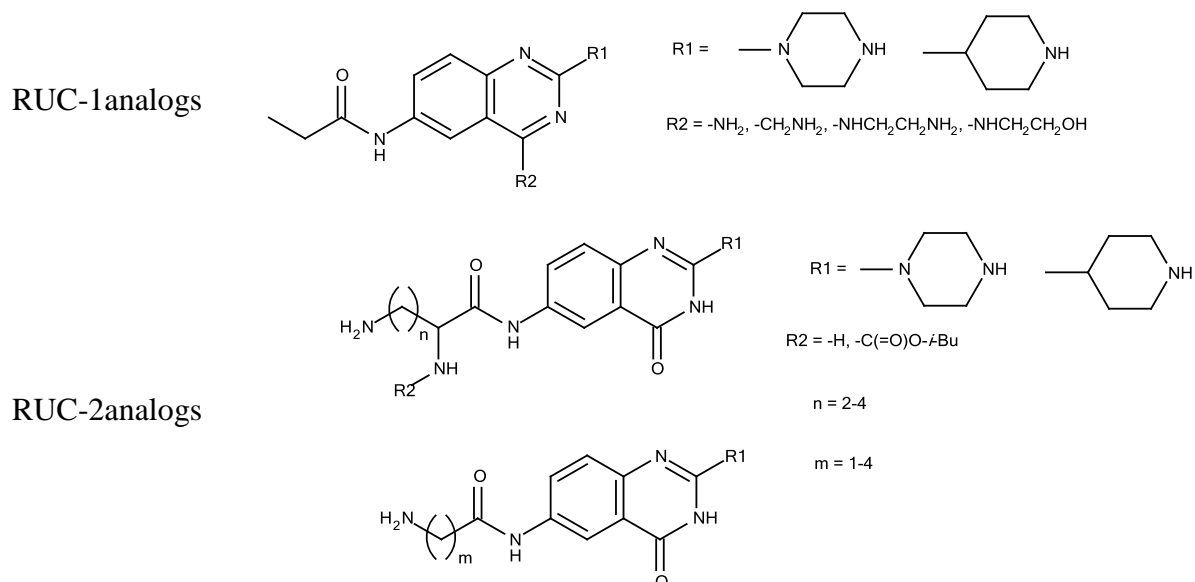
ii) a heterocyclic moiety of a ligand can be replaced with other ones but its position should be remained near the same, since it can interact with the complementary  $\alpha_{\text{IIb}}\text{Tyr190}$  residue;

iii) an acceptor group which interact with  $\alpha_{\text{IIb}}\text{Asp232}$  residue should be remained or it can be replaced with the group, which can displace water molecules surrounding  $\alpha_{\text{IIb}}\text{Asp232}$  residue and interact directly with this residue.

iv) additionally RUC-2 analogs should contain positively charged group which can interact with  $\beta_3\text{Glu220}$  residue and displace  $\text{Mg}^{2+}$  ion.

**Figure 65.** Principles of design of novel antagonists of  $\alpha_{\text{IIb}}\beta_3$  integrin - analogs of RUC-1 and RUC-2.

19 analogs of RUC-1 and 28 of RUC-2 have been designed using described considerations and taking into account synthetic feasibility of designed compounds (Figure 66).



**Figure 66.** Structures of designed analogs of RUC-1 and RUC-2.

## 7.5. Virtual screening of designed compounds

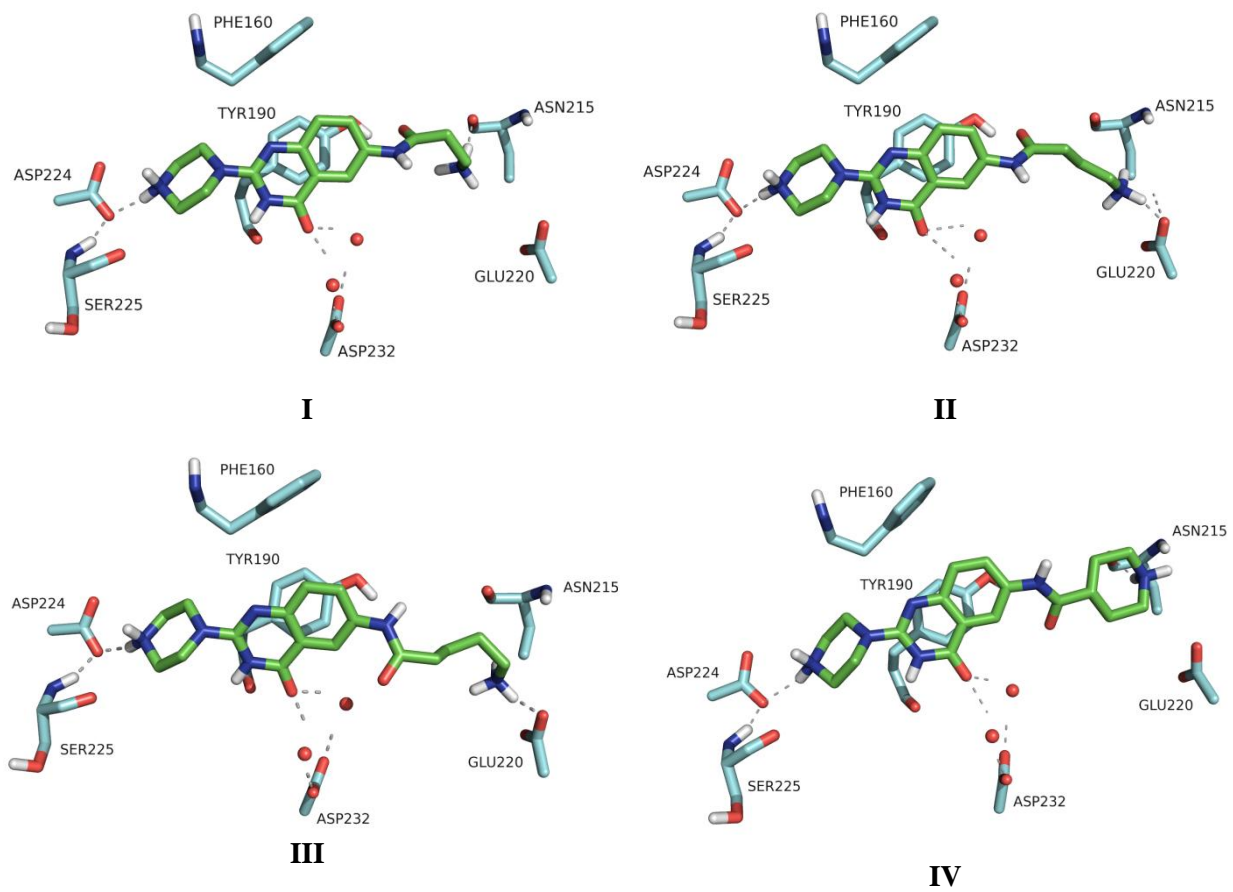
For the purpose of screening all designed compounds were standardized using *Standardizer* of ChemAxon software. Afterwards the compounds were ionized using *Filter*. Since compounds of focused libraries have been designed with consideration of the main ligand-protein interactions reflected in pharmacophore models, these compounds have not been filtered by 3D pharmacophore models, but have been docked. 19 analogs of RUC-1 and 4 analogs of RUC-2 have been kept after the docking.

Additionally, as in previous case kept compounds were assessed for the potential pharmacological activities with help of PASS program (the compounds were filtered out if the difference between  $P_a$  and  $P_i$  of possible adverse effects or toxicity exceeded 0.6) and for some ADME/Tox properties, like aqueous solubility and mutagenicity by earlier published QSAR models [149, 155, 160-161]. All compounds were not mutagenic and possessed satisfactory predicted solubility values ( $\sim 0.001$  mol/l). According to prediction of PASS these compounds didn't have any toxic or side effects.

## 7.6. Experimental validation

One of the analogs of RUC-1 and four analogs of RUC-2 were synthesized tested on their affinity for  $\alpha_{11b}\beta_3$  in PCI. All of the compounds showed high affinity for  $\alpha_{11b}\beta_3$  receptor. According to the molecular docking of RUC-2 analogs two of them (derivatives of  $\gamma$ -aminobutyric (**II**),  $\delta$ -aminovaleric acids (**III**)) had the same binding mode as RUC-2 ligand, while two others (derivatives of  $\beta$ -alanine (**I**) and piperidine-carboxylic acid (**IV**)) bind to  $\beta_3$ Asn215 instead of  $\beta_3$ Glu220 residue. But it cannot be excluded that the latter ones can bind to  $\beta_3$ Glu220 via water molecules which are present in the binding site.

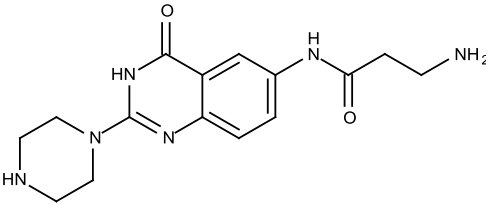
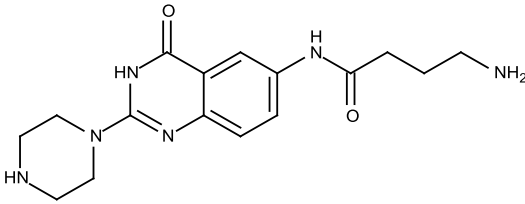
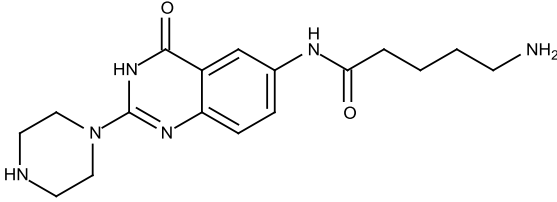
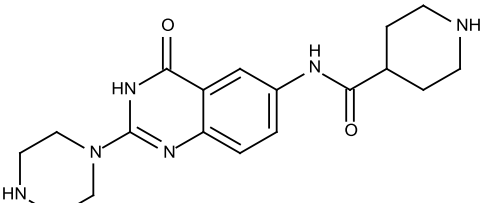
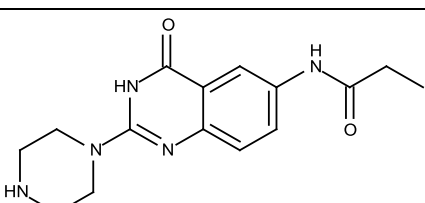
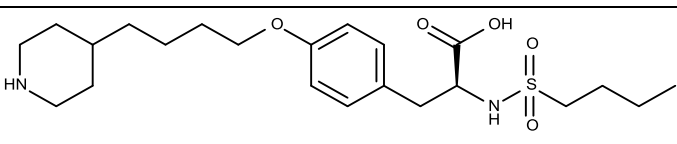
It also can be observed that compound (**I**) can be compared with synthesized RUC-1 analog (derivatives of  $\beta$ -alanine (**V**)). As can be seen this compound has one order of magnitude higher affinity value than analog of RUC-1. This can be an evidence of importance of amino group, the second positively charged group, for the binding of non-classical ligands.



**Figure 67.** Docking poses of remained analogs of RUC-2



Table 14. Affinity for  $\alpha_{IIb}\beta_3$  receptor of designed RUC-2 analogs

	Compound	Affinity for $\alpha_{IIb}\beta_3$ , pIC <sub>50</sub>
I		8.30
II		8.66
III		8.85
IV		8.42
V		7.26
Tirofiban		8.62

## 7.7. Conclusions

This section of the work is dedicated to computer-aided design of new potent antagonists of the closed form of integrin  $\alpha_{IIb}\beta_3$ , analogs of RUC-1 and RUC-2 ligands. The structure of protein-ligand complexes were used to develop the structure-based 3D pharmacophore models

(LigandScout), which subsequently were used to obtain 2D pharmacophore models. Using of these models and ligand to protein docking (FlexX) for the virtual screening of the ~35M commercially available compounds resulted in 18 analogs of RUC-1 ligand and 2 analogs of RUC-2 ligand. Kept compounds were evaluated on the basis of some ADME/Tox properties (mutagenicity, and solubility) and potential pharmacological activities (PASS). One of the RUC-2 analogs was a known drug, *Nafamostat*, with already described anti-aggregation properties, whereas others analogs were no longer commercially available. Therefore, the virtual library which contains 19 analogs of RUC-1 and 28 of RUC-2 has been designed from the list of preselected fragments. Since the compounds have been designed with consideration of the main ligand-protein interactions reflected in pharmacophore models, these compounds have not been filtered by 3D pharmacophore models but have been docked. Kept compounds were also assessed on some ADME/Tox properties and potential pharmacological activities. 19 hits analogs of RUC-1 and 4 hits analogs of RUC-2 were retrieved by virtual screening. One RUC-1 analog and four analogs of RUC-2 were synthesized and experimentally studied in PCI. It was observed that all synthesized analogs of RUC-2 ligand possess affinity comparable to tirofiban.

*Chapter 5.*

**Development of antagonists of thromboxane A2  
receptor**

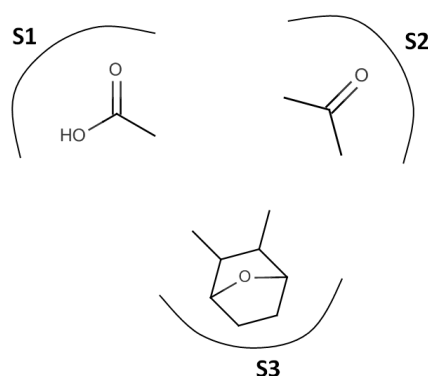
In this part of the work, some of the previously performed modeling studies on antagonists of thromboxane A2 receptor (TP) (Section 8) and the results of computer-aided design obtained in this work (Section 9) will be reported.

## 8. Previously reported models

According to the mutation analysis, there are hypothesis about the existence of TP-specific ligands which interact with different residues in the ligand-binding pocket. Khasawneh *et al.* [261] provided investigation of coordination of two structurally different antagonists of the TP receptor (i.e. SQ29,548 and BM13.505). The results have identified four key amino acids Phe184, Thr186, Ser191, and Asp193 that participate in TP antagonists binding. Mutation of all four amino acids produced dramatic effects on SQ29,548 binding, while for BM13.505 Phe184, Thr186 and Ser191 did not appear to be critical. These findings suggest the possibility of different ways of binding of structurally different TP antagonists.

### 8.1. Pharmacophore models

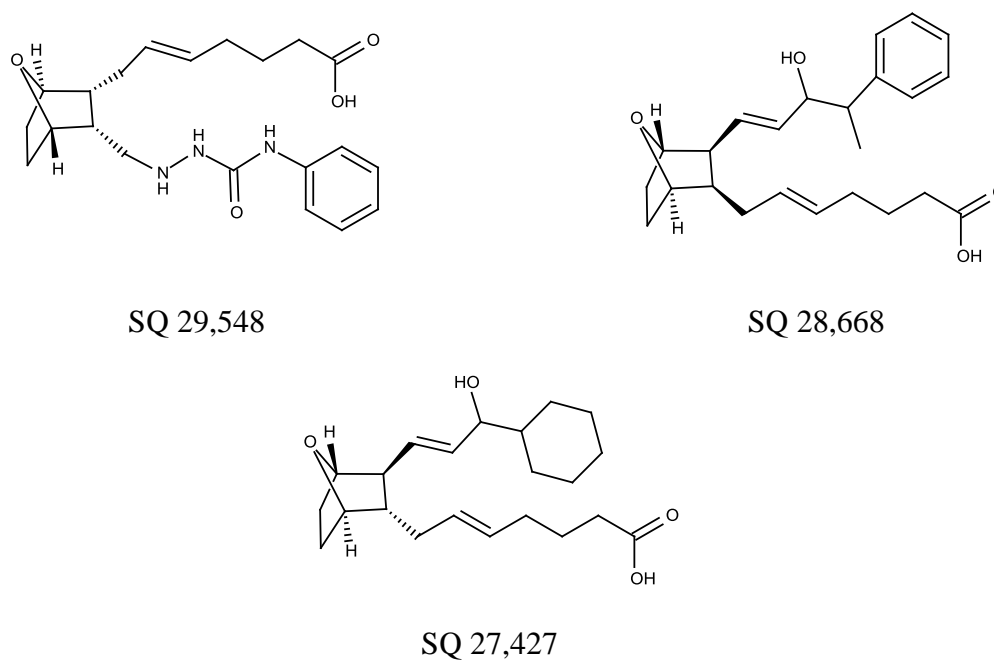
According to Farmer's three site-ligand concept [262], at least three recognition sites of the antagonists are required for high affinity. Based on this assertion, Jin *et al.*[263] distinguish three recognition sites (S1, S2, S3) in TP (see Figure 68): (i) S1 – binding a carboxylic acid group, which is hypothesized to act as a primary recognition site, (ii) S2, binder of a carbonyl or carbonyl-like hydrogen bond acceptor group and (iii) S3 is assumed to harbor the head of the SQ series of compounds (7-oxabicyclo[2.2.1]-heptane ring).



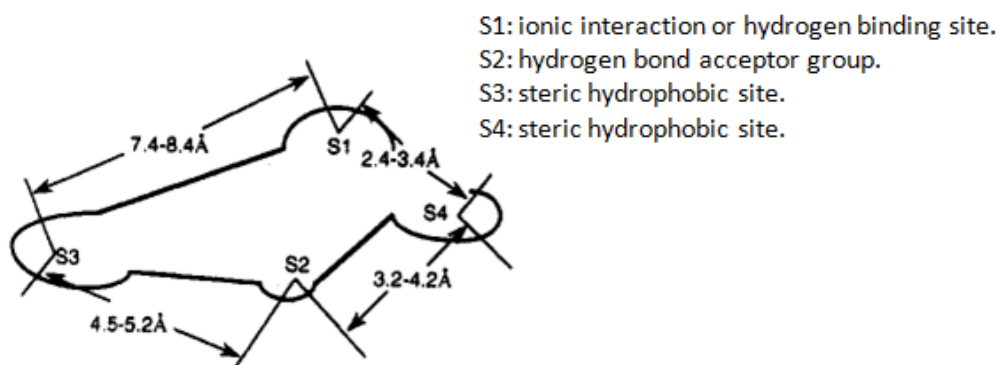
**Figure 68.** Three recognition sites presented by Jin *et al.* [263]

To refine the existing pharmacophore model, Jin *et al* identified the distance between the sites using low-energy conformations of three TP antagonists SQ 29,548, SQ 28,668, SQ 27,427

(see Figure 69 and Figure 70). After analysis, the authors concluded that obtained conformations fit well within a common envelope, with exception of the benzene ring of SQ 29,548. Since SQ 29,548 is more potent than SQ 28,668 or SQ 27,427, the space occupied by the benzene ring may permit intermolecular interactions which are responsible for the potency differences among these three antagonists. It is reasonable to hypothesize that the hydrophobic benzene ring of SQ 29,548 may interact with an additional hydrophobic pocket (S4) in the TP receptor to enhance potency.



**Figure 69.** Studied antagonists of TP receptor from reference [263].

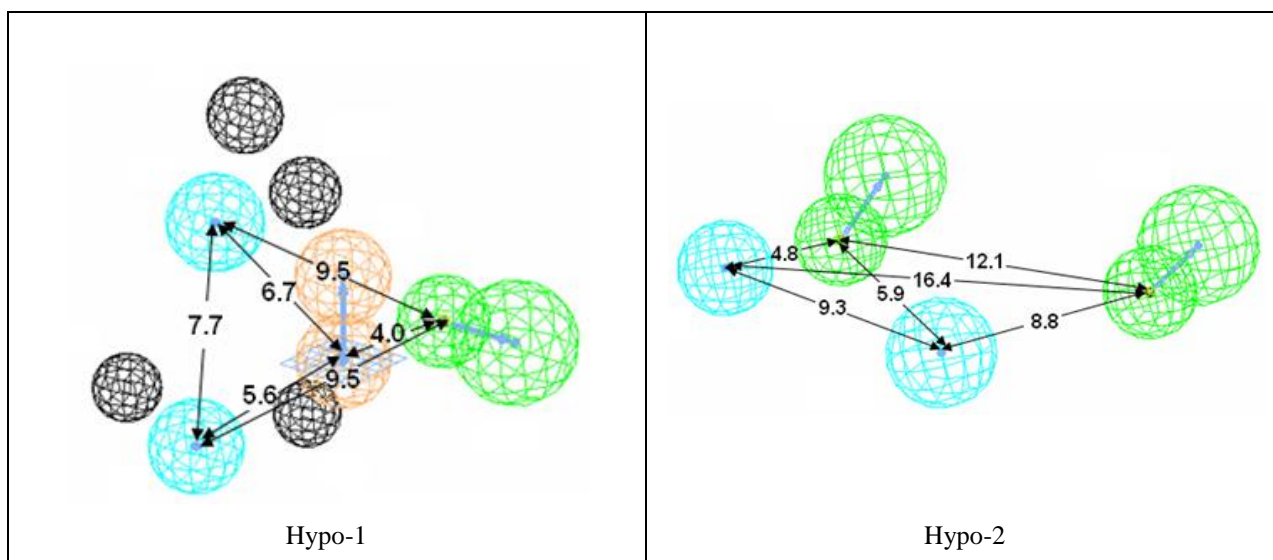


**Figure 70.** The proposed pharmacophore model from reference [263]

Later on, the group of Wei *et al* [264] presented generation of 3D pharmacophore models for TP antagonists using both HypoGenRefine and HipHop modules of the CATALYST software package.

The HypoGenRefine is extensions of the HypoGen algorithm with addition of excluded volumes, which define the space which the ligand shouldn't occupy [265]. HypoGen generates an activity-based pharmacophore model which can be used to estimate activities of new compounds [266]. It tries to correlate the three-dimensional arrangement of chemical features with the biological activities of training set molecules [265]. HipHop provides feature-based alignment of a collection of compounds without considering activity [266].

Two specific ligand-based pharmacophore models were obtained by Wei *et al* [264]. For validation purpose, the authors selected 30 TP antagonists to form the test set for the HypoGenRefine model, and 12 TP antagonists – for the HipHop model. The HypoGenRefine model (Hypo-1, see Figure 71) obtained was developed on the basis of 25 antagonists. It consists of two hydrophobic groups, one aromatic ring, one hydrogen-bond acceptor and excluded volumes. It is characterized by a correlation coefficient between measured and estimated activities of 0.91.



**Figure 71.** The best quantitative model Hypo-1 obtained using HypoGenRefine (left) and the best qualitative model Hypo-2 built using HipHop (right) from reference [264]. Pharmacophore features are color-code with green for hydrogen bond acceptor, cyan for hydrophobic, orange for aromatic ring, and black for excluded volumes.

Seven highly active antagonists with various structures were used to generate the second model (see Figure 71, right) using HipHop module. Hypo-2 model contains two hydrophobic groups and two hydrogen-bond acceptors.

In spite of above-mentioned existence of various different binding modes [261] allowing for putatively different pharmacophore models, the developed models have some common attributes. All include the key carboxylic group: in the first model is represented as anionic center (Figure 70), and as a hydrogen bond acceptor in the other (see Figure 71). Also, developed models confirm the importance of hydrophobic interactions for binding of TP antagonists.

## 8.2. QSAR modeling

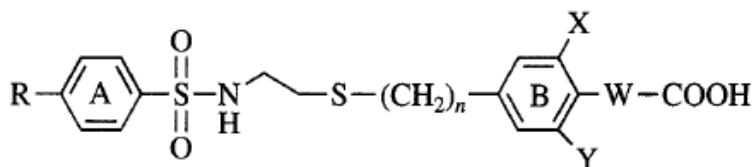
Kawashima *et al.* [267] have reported the QSAR analysis of 4-[2-(4-substituted phenylsulfonylamino)ethylthio]phenoxyacetic acids (see Figure 72), which showed potent TP receptor inhibition activity. The dataset contained 36 antagonists for which inhibition of platelet aggregation was measured ( $IC_{50}$ ). The QSAR studies used the Hansch-Fujita method. In the parameterization of structural features for the Hansch study, the authors investigated physicochemical descriptors generally used in QSAR studies. For validation of the QSAR models leave-one-out cross-validation procedure was applied.

In the obtained eqs. 20-22 (see Table 15),  $\pi_R$  and  $\pi_{X,Y,W-COOH}$  are hydrophobicity parameters, and  $F_R$  is the Swain-Lupton field constant for substituent **R** cited from the compilation by Hansch and Leo.  $\sum Q_{(1)-(6)}$  is the total electronic charge on the **B** benzene ring.  $Q_{\alpha C}$  is the electronic charge of the carbon atom adjacent to the carboxylate anion.  $L_R$ ,  $B_1$ , and  $B_5$  are the Sterimol parameters for the length, minimum width, and maximum width of substituted **R**;  $L_{W-COOH}$  is for the length of the **W-COOH** moiety.

Table 15. Results of Hansch-Fujita analysis for 36 TP receptor antagonists from reference [267]

No	Regression equation	$R^a$	$S^b$	$R_{(pred)}^c$
20	$-\log C = 1.06 B_1 - 0.98 L_{W-COOH} + 2.09 \pi_{X,Y,W-COOH} + 3.04 \sum Q_{(1)-(6)} + 10.12$	0.93	0.27	0.91
21	$-\log C = 0.61 \pi_R + 1.14 F_R - 9.33 Q_{\alpha C} + 4.99 \sum Q_{(1)-(6)} + 3.93$	0.94	0.25	0.92
22	$-\log C = 0.78 L_R - 0.12 (B_5)^2 - 9.47 Q_{\alpha C} + 5.00 \sum Q_{(1)-(6)} + 2.48$	0.94	0.25	0.92

a) Correlation coefficient for training set. b) Standard error of estimate. c) Correlation coefficient obtained in leave-one-out cross-validation procedure.



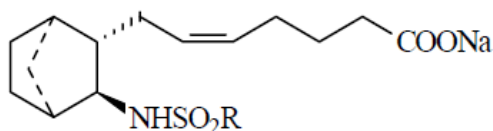
**Figure 72.** 4-[2-(4-substituted phenylsulfonylamino)ethylthio]phenoxyacetic acids (**R**: H, F, Cl, Br,  $NO_2$ ,  $CH_3$ ,  $OCH_3$ ; **W**:  $CH_2$ ,  $CH_2CH_2$ ,  $OCH_2$ ; **X**, **Y**: H, F; **n**: 0, 1)

Analyzing the results of the QSAR study the authors suggested that a hydrophobic and electron-withdrawing substituent **R** (F, Cl,  $NO_2$ , etc.) at the *para*-position of the phenylsulfonyl moiety is required to improve the activity. Furthermore, a long length and moderate width for a

substituent **R** was suggested to be preferable for the higher activity. For the benzene moiety **B**, the positive coefficients for  $\pi_{X,Y,W-COOH}$  and  $\sum Q_{(1)-(6)}$  may indicate that insertion of a hydrophobic electron-withdrawing group on the benzene ring enhances the activity. Also the length of the **W**-COOH moiety may be important [267].

The group of Kontogiorgis et al [268] built a QSAR of TP antagonists by using the C-QSAR suite of programs. The C-QSAR is a database of over 18000 equations that relate to biological or physico-chemical properties of molecules to various molecular descriptors. The data used to derive the quantitative structure activity relationships are taken from various high quality journals. C-QSAR comprises two databases, one for structure-activity information biological systems and the other for physico-chemical properties of organic systems.[269]

For QSAR model development only calculated logP values and CMR calculated molar refractivities have been used. All calculated values were obtained for the neutral forms. The values of substituent constants  $F$  and  $B_{1-4ph}$  have been taken from the literature.



**Figure 73.** The general structure of studied compounds from reference [268]. (R: -C<sub>6</sub>H<sub>5</sub>, -(CH<sub>2</sub>)<sub>n</sub>-C<sub>6</sub>H<sub>4</sub>, -C<sub>6</sub>H<sub>4</sub>NO<sub>2</sub>, -C<sub>6</sub>H<sub>4</sub>OH, etc.)

The dataset of 14 compounds for which IC<sub>50</sub> values, referred to the inhibition of rabbit PRP aggregation induced by arachidonic acid, was used to derive the quantitative structure-activity relationship given in Eq. 23. In all equations  $n$  represents the number of data points,  $R$  is the correlation coefficient between observed and predicted from the equation activity values,  $S$  is the standard deviation of the regression equation,  $Q^2$  defines the cross-validated  $R^2$ ,  $F$  the Swain-Lupton factor for inductive field electronic effect refers to substituents in all positions of phenyl ring.

$$\log 1/IC_{50} = -1.056(\pm 0.269)CMR + 0.955(\pm 0.916)F_{.ph} + 16.841(\pm 3.024) \quad (23)$$

$$n=14, R=0.953, Q^2=0.830, R^2=0.909, S=0.319, F_{2,11}=37.263$$

In this expression CMR is the overall calculated molar refractivity of the molecule. Authors concluded that its negative sign suggested the steric hindrance either directly or through



a conformational change in the receptor. The positive sign of **F** suggests that electron withdrawing property of substituents in benzene ring enhances activity as it has already been shown in the work of Kawashima et al [267].

In order to better delineate the physico-chemical parameter that governs the effect of the substituent in the position 4, the second equation (eq. 24) was derived using dataset of 15 compounds, for which inhibition of the aggregation of rabbit PRP is available:

$$\log 1/IC_{50} = -1.278(\pm 0.244)CMR + 0.851(\pm 0.643)B_{1-4ph} + 18.314(\pm 2.568) \quad (24)$$

$$n=15, R=0.957, R^2=0.916, Q^2=0.848, S=0.303, F_{2,12}=100.866$$

The  $B_{1-4ph}$  term (the Sterimol smallest width of the substituent) appears to confirm a positive steric effect for 4-substituents of the phenyl ring, which is in accordance with the previous study of Kawashima et al. [267]. The larger the atom attached to the ring, the more effective the acid derivative will be.

Another attempt was made by authors to develop a QSAR model, for a dataset of 20 compounds with the inhibitory activity on rat washed platelets aggregation:

$$\log 1/IC_{50} = 4.608(\pm 1.784)CMR - 0.207(\pm 0.084)CMR^2 - 20.257(\pm 9.445) \quad (25)$$

$$n= 17, R=0.859, R^2=0.738, Q^2=-0.197, S=0.290, F_{2,14}=19.69$$

Three compounds were omitted from the studied dataset since they were poorly predicted and didn't fit well the parabolic relationship. One of the compounds was a benzamide, with the lowest activity, whereas other compounds are sulfonyl derivatives. No role for an electronic factor was found from this equation. Also no correlation for a lipophilic effect was found, although in the previous work [267] the lipophilic and steric properties were found as important.

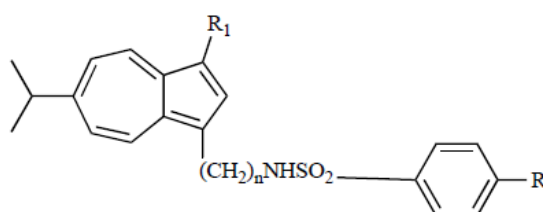
Later on, the same group of researchers [270] performed several QSAR studies using C-QSAR. The first one was devoted to analysis of the dataset contained 11 azulene derivatives (Figure 74), which were tested for contraction's inhibitory activity on rabbit aorta, which is highly populated by TP receptors and often used as a test system for studying of TP antagonists. The following equation was derived:

$$\log 1/IC_{50} = -0.810(\pm 0.640) I_{COOH} + 2.128(\pm 0.926) B_{1R} + 3.126(1.104) \quad (26)$$

$$n=11, R=0.933, R^2=0.870, Q^2=0.798, S=0.385, F_{2,9}=26.789$$

$B_{1R}$  is the Sterimol parameter of Verloop for the smallest width of **R**.  $I_{COOH}$  is indicator, which assigns the value 1 for the presence of carboxylic group at substituent  $R_1$  and 0 for the absence.

Analyzing the results of QSAR model, the authors make an assumption that the smaller width of substituent increases the biological activity. Also the presence of a COOH group as the  $R_1$  substituent seems to decrease the biological response.



**Figure 74.** General structure of studied azulene derivatives (R: H, Cl, OCH<sub>3</sub>; R<sub>1</sub>: COOH, SO<sub>3</sub>Na, CH<sub>2</sub>COOH, CH=CHCOOH; n=2-5)

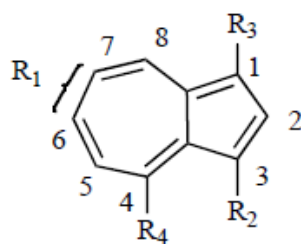
The conclusion of the authors about COOH group is doubtful, since in the dataset the substitution  $R_1$  may be COOH or SO<sub>3</sub>Na groups, and they are mostly identical in possibility to form hydrogen bonds.

The second analysis describes a series of novel azulene-1carboxylic acid derivatives (Figure 75). The thromboxane receptor antagonistic activity is expressed as  $\log 1/IC_{50}$  of inhibition of U-46619 induced contraction of rat aorta. Based on the dataset of 45 compounds, the following equation (see eq. 27) was derived:

$$\log 1/IC_{50} = 0.470(\pm 0.186) ClogP - 0.606(\pm 0.313) MR-R_{1-7} - 0.348 (\pm 0.075) MR-R_3 + 3.923(\pm 1.222) \quad (27)$$

$$n=43, R = 0.912, R^2=0.832, S=0.575, F_{3,41}=64.488$$

Two compounds from the dataset were omitted. One of which was least active and presented the lowest  $MR-R_3$  values (molar refractivity for substituent  $R_3$ ), whereas the second one was the only derivative with a trans  $R_3$ . Analyzing the equation the authors made a suggestion that the larger  $R_3$  substituent is the higher the antagonistic activity. Also, the electronic effects are appeared unimportant.

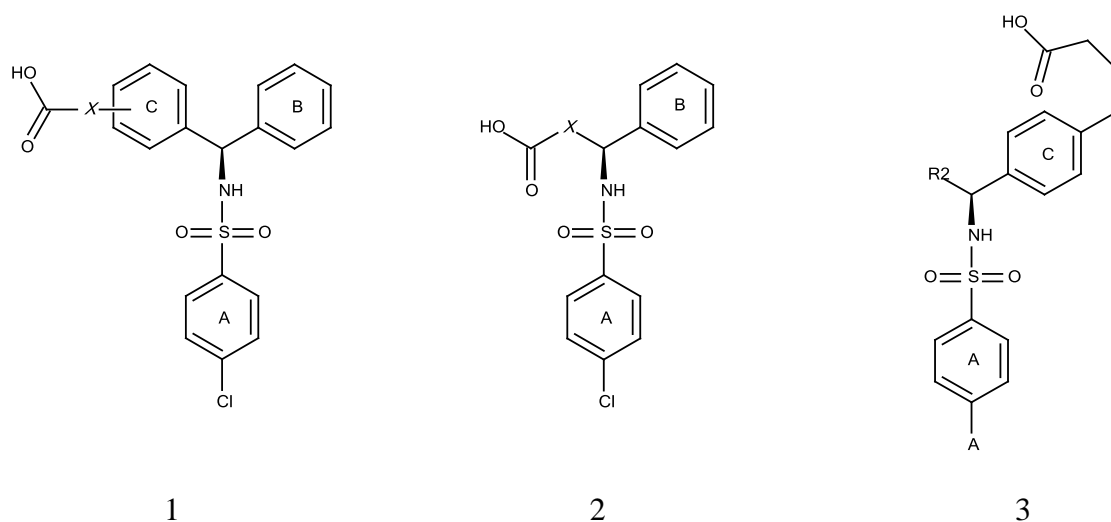


**Figure 75.** General structure of azulene-1-carboxylic acid derivatives ( $R_1$ :H, 6, 7-isopropyl;  $R_2$ :  $\text{CH}_3$ , n, i- $\text{C}_3\text{H}_7$ ,  $(\text{CH}_2)_n\text{NHSO}_2\text{-Ph}$ , cis, trans- $(\text{CH}_2)_2\text{CH}=\text{CH}(\text{CH}_2)_2$ , etc.;  $R_3$ :  $\text{SO}_2\text{NH}_2\text{CH}_2\text{-p-C}_6\text{H}_4\text{-OCH}_2\text{COOH}$ ,  $\text{COOH}$ ,  $\text{CH}_2\text{COOH}$ , trans- $\text{CH}=\text{CHCOOH}$ )

Besides the 2D QSAR, a 3D QSAR study has been made by Sairam et al [271]. They investigated some [[1-Aryl(or Benzyl)-1-(benzenesulphonamino)methyl]phenyl] Alkanoic acid derivatives by using *the receptor surface analysis* (RSA) method. This method is effective for the analysis of data sets where activity information is available but the structure of the receptor site is unknown. Receptor surface models provide compact, quantitative descriptors which capture three-dimensional information about a putative receptor site. These descriptors may be used alone or in combination with more traditional 2D descriptors [272].

The studied dataset contained 31 antagonists from which 25 were considered as training set and the rest – as a test set. For all compounds concentration needed to inhibit U-46619 ( $2\mu\text{g/ml}$ )-induced platelet aggregation in guinea-pig platelet-rich plasma ( $\text{pIC}_{50}$ ) was available. The 3D-molecular structure were generated and optimized with OFF in Cerius2, while AM1 calculations were used for further geometric optimizations following eigenvector methods. Optimized molecular structures and partial atomic charges were used for the molecular alignment with respect to the most active molecule. The receptor surface was generated for the aligned molecules with weights proportional to the biological activity. The steric and electrostatic interaction energies between each molecule and the receptor surface were evaluated and included in the QSAR study. Regression analysis was carried out using the Genetic Partial Least Squares method consisting nearly 20000 cross-over generations.

The QSAR model with 25 molecules yielded  $R_{cv}^2 = 0.758$  for cross-validation and  $\text{PRESS} = 3.613$  (the root mean square error of all target predictions). Analyzing the results the authors make an assumption that the polar substituents on the aromatic rings **A** and **B** are almost ineffective. Increase in electron withdrawing power near to acid group will enhance the electrostatic interaction energy in that region leading to higher activities. Thus, it seems that non-polar substituents on ring **B** may increase the activity while subtle variations in the nature as well as conformation of the acid group may result in a very significant change towards the activity. Furthermore, the substituents on the **A** ring are almost totally ineffective and may not directly contribute towards activity.



**Figure 76.** [[1-Aryl(or Benzyl)-1-(benzenesulphonamino)methyl]phenyl] Alkanoic Acid derivatives.

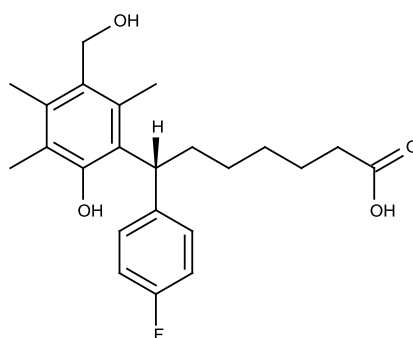
Most described QSAR studies have some drawbacks, such as: 1) the datasets are small and homogeneous, so the developed models have too narrow applicability domains; 2) there is no external validation, so it is impossible to estimate the real predictive ability of the model. Nevertheless in most cases the authors point out that the increase in electron withdrawing power near to acid group will enhance the electrostatic interaction energy in that region, leading to higher activities.

### 8.3. Ligand to Protein docking

The TP receptor is a member of the G protein-coupled receptor family with seven transmembrane segments. Since its X-ray structure could not be determined, there are a lot of studies of docking study with a modeled receptor.

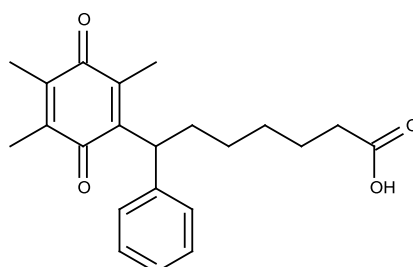
The first attempts of TP modeling have been made when only few 3D structures of GPCRs have been determined – by Yamamoto *et al.*, for example [273], using the only so-far solved GPCR structure, bacteriorhodopsin (bR) as a template. As a first step, the authors constructed a model for the  $\beta_2$ -adrenergic receptor using the structure of bR receptor as a template, since abundant mutational data was available [274]. Then, a model for the TP receptor was constructed from the model of  $\beta_2$  receptor. In order to examine the mode of the receptor-ligand interaction, the authors first docked (R)-(+)-7-(4-fluomphenyl)-7[2-hydroxy-5-(hydroxymethyl)-3,4,6-trimethylphenyllheptanoic acid (see Figure 77), a potent nonprostanoid antagonist with relatively rigid conformation. It was observed that its nonphenolic benzene ring was directed downward and the methylene chain adopted an extended conformation. The hydroxyl group of  $\text{CH}_2\text{OH}$  formed a hydrogen bond with Ser201, the carboxyl group interacted

with Arg295, and the nonphenolic benzene ring was surrounded by hydrophobic residues and the phenolic hydroxyl formed a hydrogen bond with the backbone oxygen of Trp258.



**Figure 77.** (R)-(+)-7-(4-fluomphenyl)-7-[2-hydroxy-5-(hydroxymethyl)-3,4,6- trimethylphenyllh eptanoic acid, the compounds used for docking from reference [273].

The same strategy for modeling of the TP receptor was used by Wouters *et al.* [275]. Seratrobast (see Figure 32), a potent thromboxane A2 receptor antagonist, has been docked in the modeled receptor. Both the R and S isomers were studied. Manual docking of the ligand into the receptor placed the terminal carboxylic group of Seratrobast within binding distance to the lateral chain of Arg295, as suggested in the previous work [273]. The rest part of the antagonist was adjusted in the binding site of the receptor in order to minimize unfavorable steric contacts. The binding of the R isomer to the TP receptor is predicted to be favored with respect to the binding of the S isomer, which is consistent with biological data [276].



**Figure 78.** Structure of Seratrobast [275].

Later on, a different strategy was adopted by Ruan *et al.* [277], who tried to mimic the extracellular loops of the TP receptor by “computation-guided constrained peptide synthesis” for the structural determination using 2D NMR spectroscopy.

Since three-dimensional structures of the second [278] and the third extracellular loop [279] regions of the TP receptor have been already individually experimentally determined, only the first one was determined by authors, using two-dimensional NMR spectroscopy. The obtained information was combined with the known NMR structures of the second and the third extracellular loop domains to construct a solution structure, which includes all three extracellular

loops connected to the conserved transmembrane helices of the TP receptor. The three-dimensional structural model for the seven transmembrane helices was constructed using homology modeling based on the bovine rhodopsin crystallographic structure.

The obtained model of TP receptor has been used to carry out the ligand to protein docking of thromboxane antagonist SQ 29,548 [280]. The initial docking was set up by the three contacts of the antagonist with Val176, Thr186 and Leu187.

The authors hypothesize two ligand interaction sites: first, the ligand is coming into contact with the recognition site on the extracellular domain and then it enters the TM (transmembrane) pocket causing the conformation change of the receptor with the G protein in the intracellular domains. To test this hypothesis, the authors docked the ligand (SQ29,548) into the identified ligand recognition pocket, and then the ligand was moved into the transmembrane binding pocket. It was observed that in the transmembrane domain SQ29,548 interact with Ser201 and Arg295, similarly to other docking studies described above [273, 275]. The distance between the two sites was about 23.0Å based on the NMR structural model.

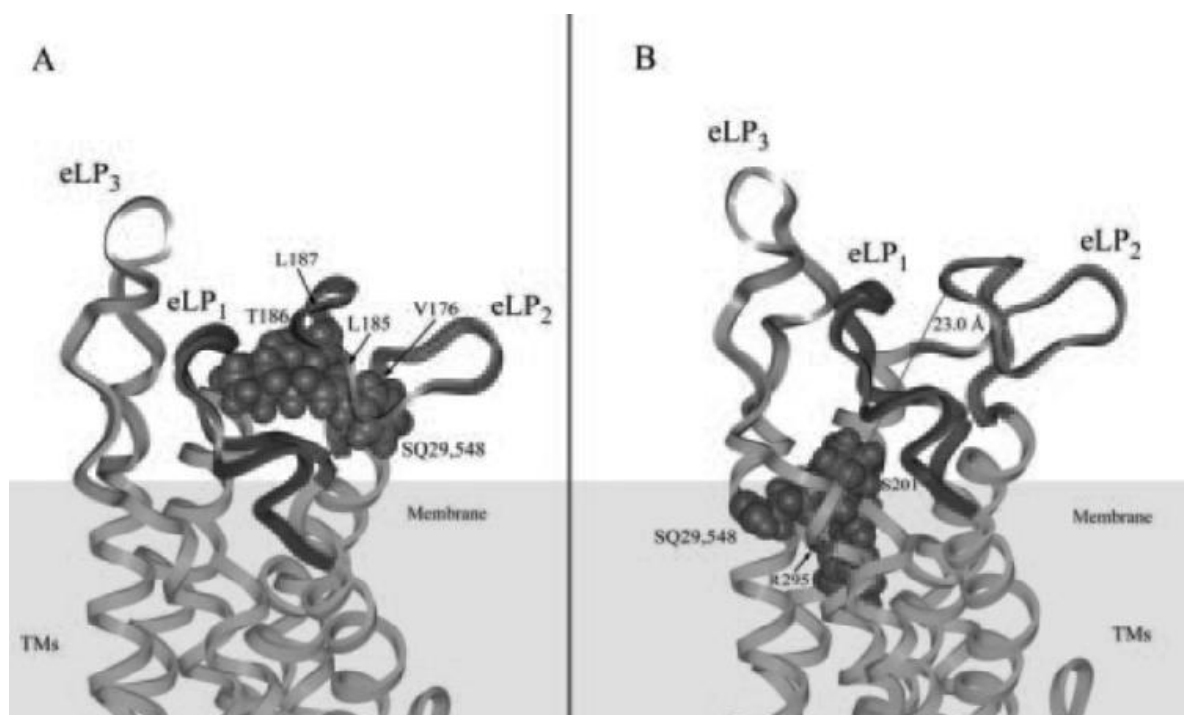


Figure 79. Docking of the TP receptor with SQ29,548 ligand from reference [277]. (A) SQ29 548 docking onto the identified ligand recognition pocket. (B) SQ29 548 at the TM binding pocket. TM- transmembrane domain; eLP1, eLP2, eLP3 - first, second and third extracellular loops, respectively.

#### 8.4. Conclusions

According to the mutagenic studies there are two possible binding modes of TP antagonists. Therefore comparison between the results of the models obtained on different datasets should be carried out cautiously. Despite of this fact the pharmacophore models have

some common attributes: ionic and hydrophobic interactions which can be important for activity of TP antagonists. QSAR studies confirm the importance of hydrophobic interactions and showed that increasing in electron withdrawing power of substituents near carboxylic and sulfonic acid groups of ligands is favorable for higher activity. Docking studies were performed on modeled TP receptors since no X-ray structure is available. All docked ligands were different in their shape, size and flexibility. Besides the ligands adopt different binding poses all of them interact with the same residues Arg295 and Ser201.

Since the datasets used in QSAR modeling contained small congeneric sets of TP antagonists they have very little applicability for virtual screening. Obtained models were not validated on external datasets, so their real predictive ability is undefined. Therefore most of the described models can't be used for virtual screening.

## 9. Computer-aided design of thromboxane A2 receptor antagonists

Here, we will describe virtual screening for discovering of new antagonists of thromboxane A2 receptor.

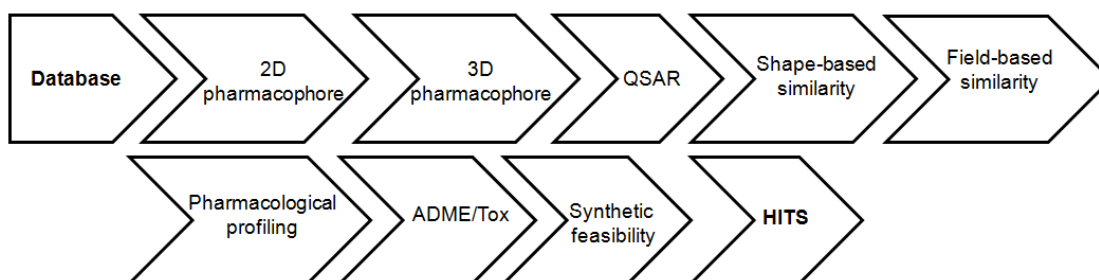


Figure 80. Virtual screening workflow used to discover novel antagonists of thromboxane A2 receptor.

### 9.1. Dataset preparation

The ChEMBL database version 9 has been chosen as source of affinity data for thromboxane A2 receptor and anti-aggregation activity. Compounds with assessed affinity for TP, in assays on human platelets, have been included in the *Affinity* dataset. Values of inhibition constants ( $K_i$ ) expressed in mol/l units have been converted to  $pK_i = -\lg K_i$ . Compounds which had been tested by Born's turbidimetry assay [234] on human platelet rich plasma using U-46619 as agonist have been collected in the *Anti-aggregation* dataset ( $IC_{50}$ ). All values of  $IC_{50}$  expressed in mol/l units have been converted to  $pIC_{50} = -\lg IC_{50}$ . The observed significant imbalance in distribution of  $pIC_{50}$  values for the *Anti-aggregation* dataset (Figure 81) is caused by the presence of qualitative results like  $pIC_{50} < 6$ , which were widely reported by authors. If such results are excluded, the representativeness of the dataset would be significantly decreased.

Both datasets have been standardized with ChemAxon *Standardizer*, afterwards the duplicates have been removed using ChemAxon *InstantJChem*. During datasets curation single stereoisomers have been removed if corresponding compounds with unspecified stereo centers (racemate) have been present in the dataset.

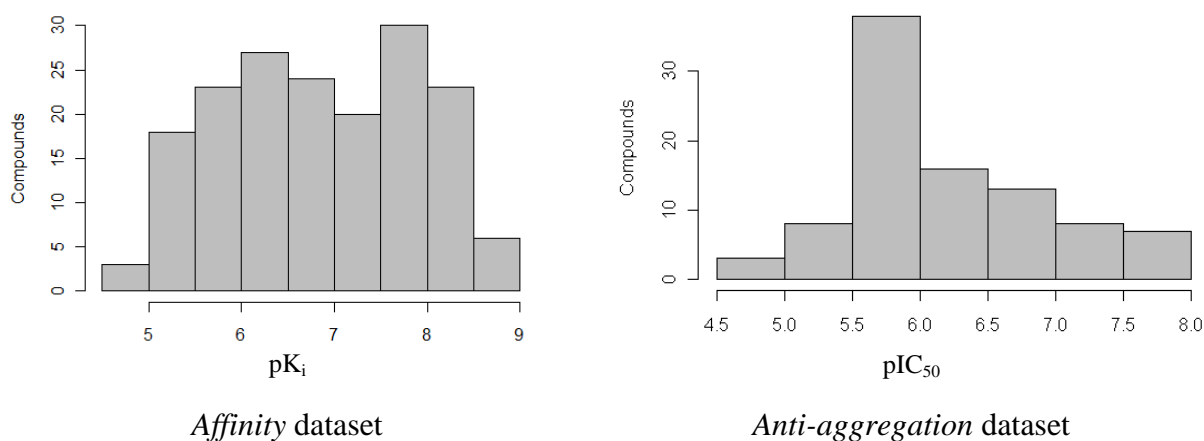
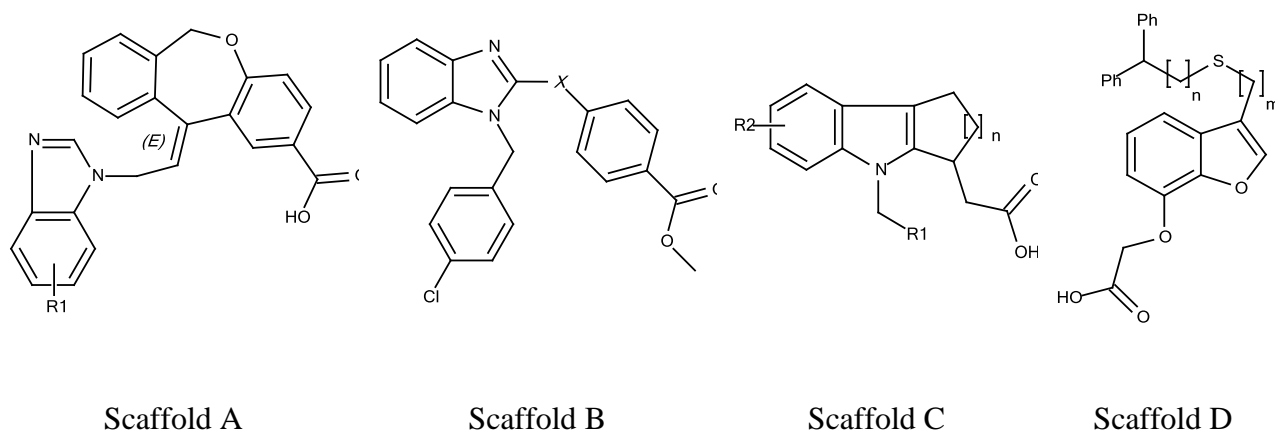


Figure 81. Distribution of experimental  $pK_i$  ( $-\lg K_i$ ,  $K_i$  in mol/l) and  $pIC_{50}$  ( $-\lg IC_{50}$ ,  $IC_{50}$  in mol/l) values for the *Affinity* (80 actives and 94 inactives) and *Anti-aggregation* (44 actives and 49 inactives) datasets.

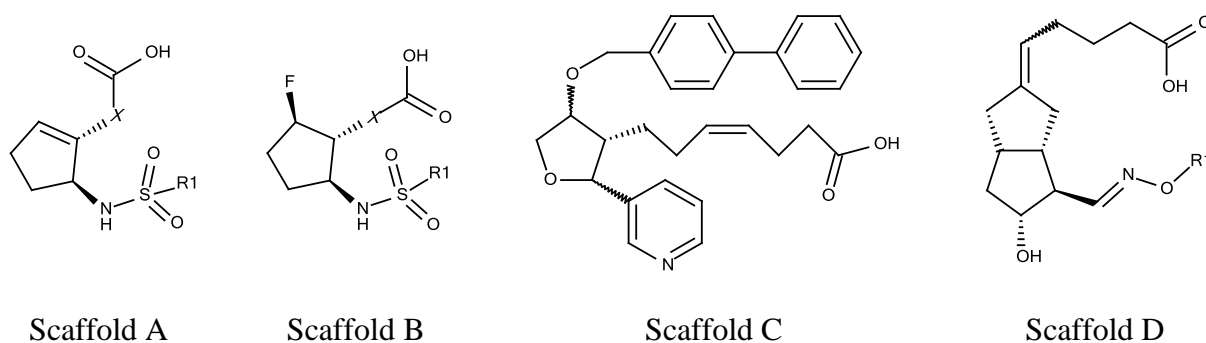


The TP antagonists from both collected datasets belong to the classes of carboxylic acids and esters. The compounds from the *Affinity* dataset are represented mainly by four scaffolds (Figure 82): (i) 11-substituted-6,11-dihydrodibenz[b,e]oxepin derivatives (scaffold A), (ii) 2-substituted-N-benzyl benzimidazole derivatives (scaffold B), (iii) tetrahydropyridoindeole derivatives, (iv) benzofuran-7-oxyacetic acid derivatives (scaffold C).



**Figure 82.** Most commonly found scaffolds in the *Affinity* dataset. Scaffold A:  $R_1 = H, NO_2, (MeO)_2, Me, Me_2, Cl,$  etc. Scaffold B:  $X = -SCH_2-, -SOCH_2-, -CH_2-, -NH-, -NHCH_2-, -(CH)_2-, -(CH_2)_2-$ . Scaffold C:  $n=0-1, R_1 = aryl, R_2 = halogen, SO_2Me, SO_2NMe_2, H,$  etc. Scaffold D:  $n=0-2, m=1-3; Ph = phenyl$

The *Anti-aggregation* dataset contains four main series of compounds: (i) prostenoic acid derivatives (scaffold A and B), (ii) oxa-prostanoid derivatives (scaffold C) and (iii) bicyclo[3.3.0]octanecarbacyclic derivatives (scaffold D).



**Figure 83.** Scaffold A and B:  $R_1 = phenyl, fluorophenyl, difluorophenyl, nitrophenyl, chlorophenyl, naphthyl;$   $X = -CH=CH-(CH_2)_3-, -CH=CH-(CH_2)_2-, -(CH_2)_3OCH_2-;$  Scaffold D:  $R_1 = mono/di- aryl-substituted methyl.$

## 9.2. QSAR modeling

The *Affinity* and *Anti-aggregation* datasets have been used for QSAR modeling. In both cases three different types of 2D molecular descriptors in combination with Random Forest

method (CF software) have been used in this study: SiRMS and two types of ISIDA descriptors – SMF and FPT. The prediction accuracy of the developed regression models have been estimated by determination coefficient ( $R^2$ ) and root mean square error (RMSE). Classification model quality was reported in terms of specificity, sensitivity and balanced accuracy (see section 2.2.4).

Three individual regression QSAR models have been developed for the *Affinity* dataset. Predictive ability of the obtained models was estimated by a 5-fold external cross-validation procedure (Table 16). For more reliable prediction, the consensus model was defined by averaging of predictions of the individual models. Predictive ability of the consensus model (Figure 84, Table 16) was a bit better than individual ones, and it was chosen for the virtual screening purpose.

To estimate the applicability domain for the consensus models the standard deviation of the prediction, obtained from the ensemble of models, was used. As a threshold 0.5 of the standard deviation was taken. 8 compounds were found out of the applicability domain.

Table 16. 5-Fold external cross-validation statistics for 2D QSAR models of affinity for TP receptor

Descriptors <sup>1</sup>	$R^2$	RMSE	$R^2_{AD}$ <sup>a</sup>	RMSE <sub>AD</sub> <sup>a</sup>	AD coverage
SiRMS	0.62	0.65	0.63	0.61	0.95
SMF	0.59	0.67	0.62	0.62	0.95
FPT	0.66	0.61	0.66	0.61	0.99
<b>consensus</b>	<b>0.66</b>	<b>0.61</b>	<b>0.66</b>	<b>0.59</b>	<b>0.95</b>

<sup>1</sup> SiRMS – simplex representation of molecular structure, FPT – pH-dependent fuzzy pharmacophoric triplets, SMF – ISIDA fragmental descriptors of 2-15 length and terminal groups. <sup>a</sup>Results within applicability domain of the model.

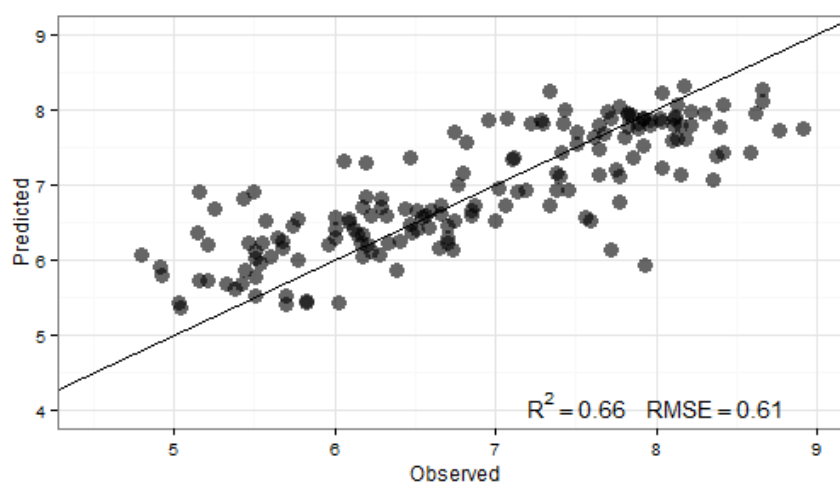


Figure 84. Observed vs. predicted values of affinity for TP receptors for the consensus model.

Since the *Anti-aggregation* dataset contains many qualitative experimental results, it has been used to develop classification models. The compounds of the dataset have been split on active and inactive according to specified threshold –  $pIC_{50} \geq 6$  for active compounds and  $pIC_{50} < 6$  for inactive ones. Since the *Anti-aggregation* dataset contained several compounds represented by different diastereoisomers, the corresponding racemic compounds have been included in the training set if all their diastereoisomers got the same activity class, otherwise such compounds have been discarded.

Three individual classification models have been developed. Since they had low accuracy, the consensus predictions have been made by majority vote of individual classification models (Table 17). The obtained consensus Random Forest model had higher accuracy and sensitivity than individual ones. Therefore, the consensus model has been chosen to use in the virtual screening process.

Table 17. Statistical characteristics of classification 2D QSAR models for *Anti-aggregation* dataset.

<b>Descriptors<sup>1</sup></b>	<b>TN</b>	<b>FP</b>	<b>FN</b>	<b>TP</b>	<b>Specificity</b>	<b>Sensitivity</b>	<b>Balanced accuracy</b>
SiRMS	34	6	12	17	0.85	0.59	0.72
FPT	30	10	8	21	0.75	0.72	0.74
SMF	33	7	11	18	0.83	0.62	0.72
Consensus	33	7	6	23	0.83	0.79	0.81

<sup>1</sup> SiRMS – simplex representation of molecular structure, FPT – pH-dependent fuzzy pharmacophoric triplets, SMF – ISIDA fragmental descriptors of 2-15 length and terminal groups.

### 9.3. Pharmacophore models

In absence of an X-ray structure, only ligand-based pharmacophore models have been produced based on the *Affinity* dataset by LigandScout program.

Validation sets have been prepared from the *Affinity* dataset. The molecules used for the model building were excluded from the validation set. For remaining compounds, in the validation dataset all possible stereoisomers and then at most 50 conformers within 10 kcal/mol generated by *Omega* have been considered. Compounds have been split on active ( $pK_i \geq 6$ ) and inactive ( $pK_i < 6$ ) ones. A compound considered as active if at least one of its generated stereoisomers fit this model. Prediction performance of the validation set was estimated by precision and recall (see eq. 3 and 4).

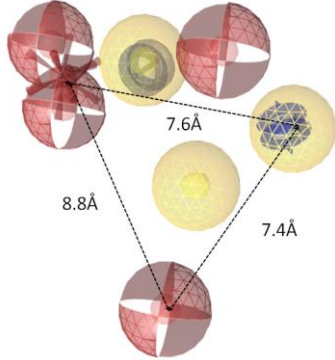
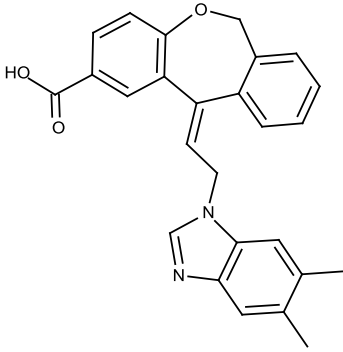
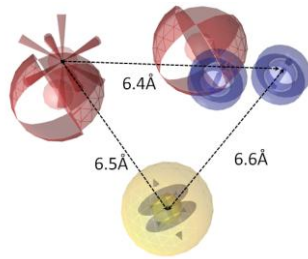
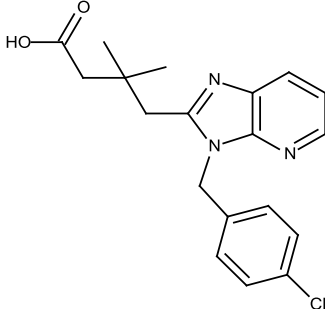
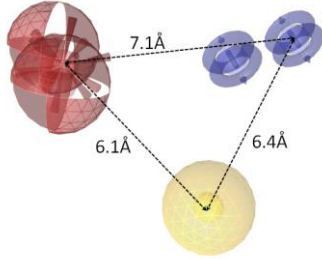
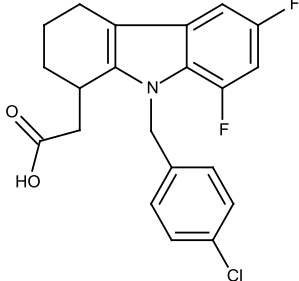
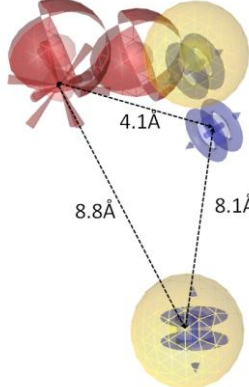
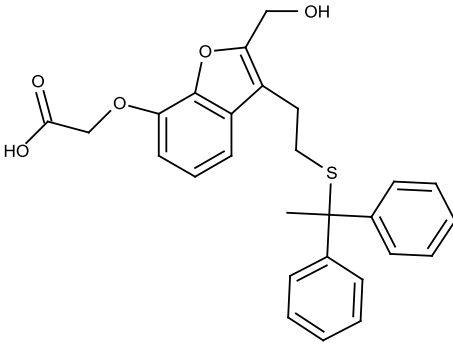
For the purpose of pharmacophore model generation, 77 compounds from the *Affinity* dataset, which had at most one (specified or unspecified) stereocenter and  $pK_i \geq 7$ , have been

selected. If unspecified, the stereocenter was systematically set to *S* configuration. The selected compounds have been ionized with *Filter* tool of OpenEye software and for each compound at most 50 conformers within 10 kcal/mol have been generated by *Omega*.

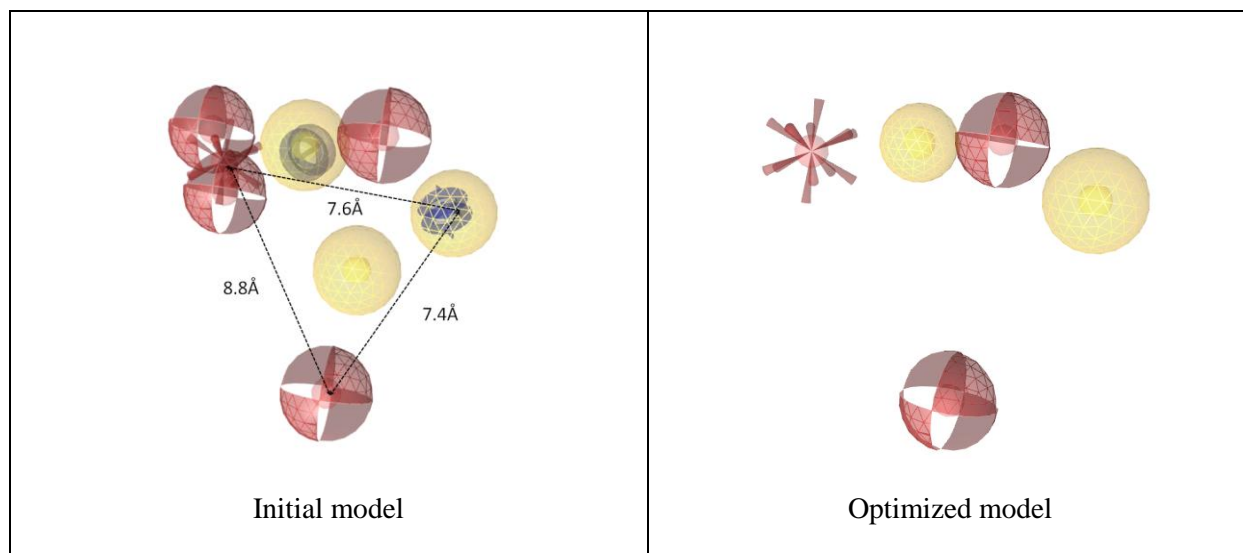
The prepared set has been clustered using LigandScout (default parameters), what resulted in 4 clusters and one singleton. Thus, four shared pharmacophore models have been produced (see Figure 85).

Originally obtained pharmacophore models were very specific or have low selectivity. Therefore, we made an attempt to optimize the model obtained on first cluster. We removed both aromatic features and decreased the tolerance of the hydrophobic feature between negative charge and H-bond acceptor, this modification helped to include the compounds from other clusters (*precision* 0.81 and *recall* 0.17, see Figure 86). As one can see, the modification didn't help to increase the performance of the model. Therefore, we joined the results of all three initially built pharmacophore models and the optimized one on first cluster (see Figure 85), and obtained reasonable *precision* = 0.60 and *recall* = 0.73. The ensemble of these models was chosen for virtual screening.

As it was explained in section 8, the pharmacophore model for TP antagonists may diverge if they were obtained on compounds with possibly a different way of binding. Thus, the developed pharmacophore models (see Figure 85) can correlate only with model shown on Figure 71 (left). They have some common features: the anionic group represented by carboxylic group, which corresponds to hydrogen bond acceptor in other model (Figure 71) and two hydrophobic features. Moreover, despite the fact that other described models were obtained on compounds different from models on Figure 85, they also have an ionic center represented by carboxylic group, which confirms its importance for antagonist binding.

№	Pharmacophore models	The most representative compounds
1	 <p data-bbox="432 591 762 622">Precision = 1, Recall = 0.01</p>	
2	 <p data-bbox="411 994 783 1025">Precision = 0.56, Recall = 0.05</p>	
3	 <p data-bbox="411 1375 783 1406">Precision = 0.41, Recall = 0.50</p>	
4	 <p data-bbox="432 1854 762 1886">Precision = 1, Recall = 0.03</p>	

**Figure 85.** Shared pharmacophore models (left) and the most representative compounds from each cluster (right). See Figure 18 for the details of feature color-coding. Exclusion volumes are not shown.



**Figure 86.** Optimization of ligand-based model built on first cluster (*precision* 0.81 and *recall* 0.17).

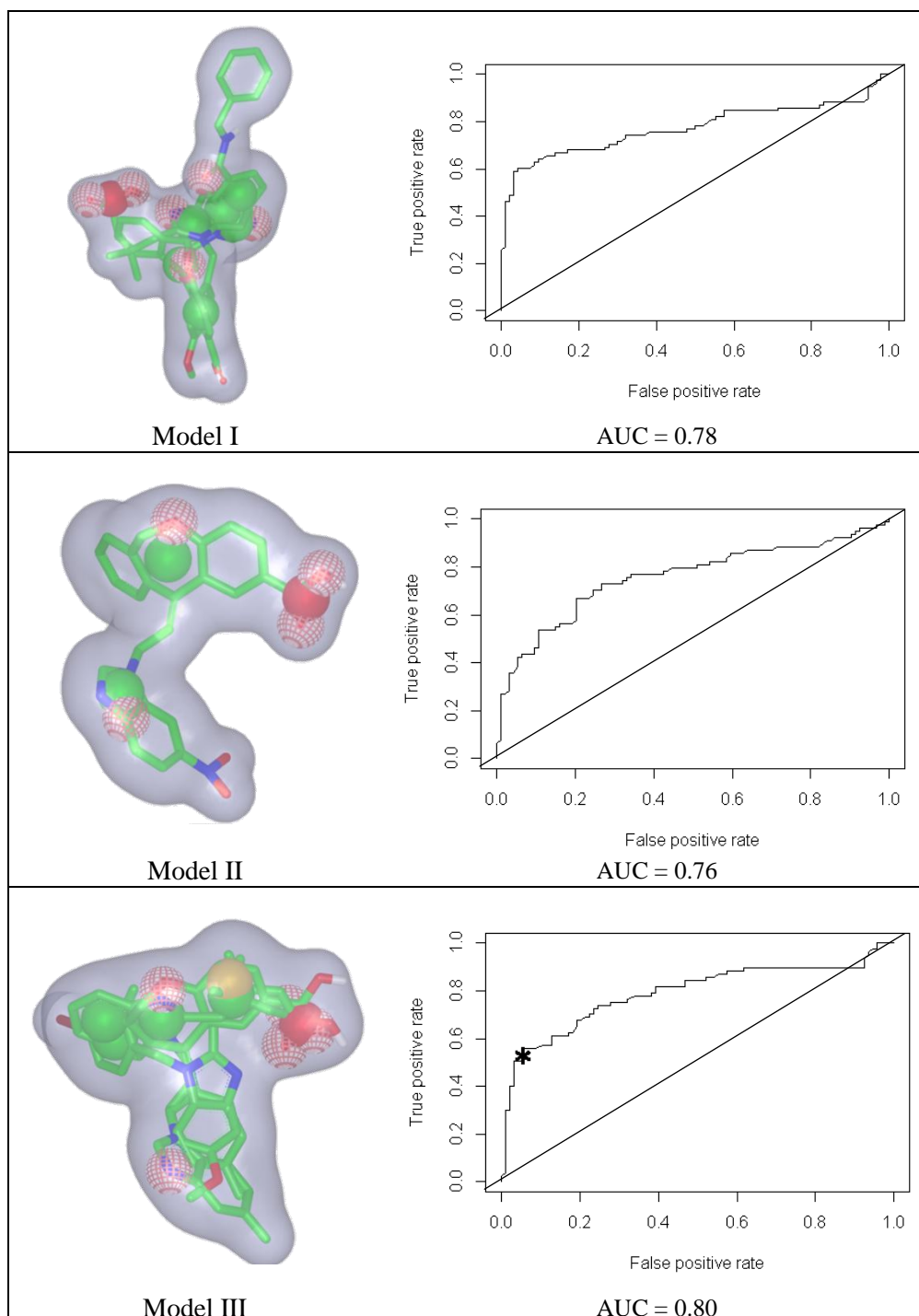
#### 9.4. Molecular shape-based comparison

The shape-based superposition method implemented in ROCS package of OpenEye software was used for generation of molecular shape models for the purpose of virtual screening. Since ROCS is a 3D approach it's necessary to choose proper “bioactive” conformers for model generation. Two ways of solving this problem have been applied and two training sets were formed. The first training set was prepared from 51 active compounds ( $pK_i \geq 7$ ) which were achiral or had specified configuration of all stereocenters and double bonds. The conformational search based on stochastic method implemented in MOE was used and the conformers with the lowest energy were considered as “bioactive”. The second one contained conformers of 28 active compounds ( $pK_i \geq 7$ ) from the entire *Affinity* dataset, which fit the pharmacophore model obtained above (cluster 1, Figure 86). Validation sets for each generated model have been prepared in the same way as for the pharmacophore models described above. Goodness-of-fit measure of screened compounds has been estimated using the TanimotoCombo score, averaged over different stereoisomers of the same compound from the validation set. For consensus prediction TanimotoCombo score of selected individual models have been averaged. Models quality has been estimated by AUC value calculated from ROC curves.

For the first training set the generated conformers were pre-aligned with MOE and all combinations of two and three compounds have been used for model generation. The best model (see Figure 87, Model I) contained three compounds had a good predictive ability  $AUC = 0.78$ .

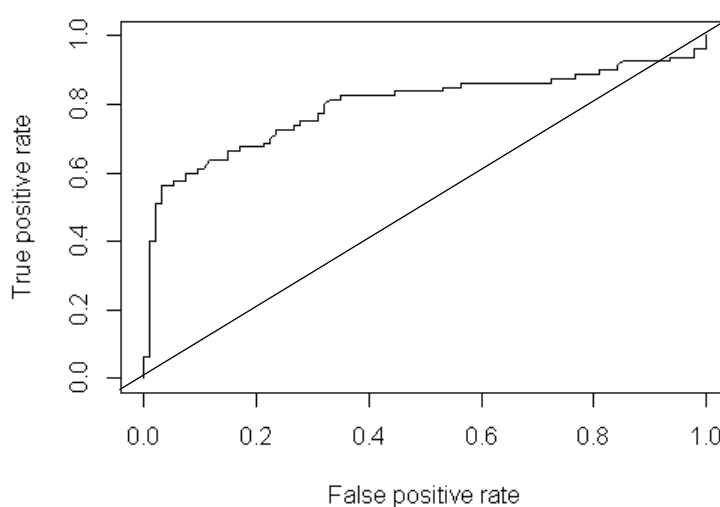
For the second training set initially no pre-alignment has been made. All combinations of two and three compounds have been used for model development. The best model (see Figure 87, Model II) contained two molecules and had satisfactory AUC value 0.76. Before a second

run for comparison purposes, all molecules have been pre-aligned by the rigid alignment method implemented in MOE. Again, all combinations of two and three compounds have been used for model generation. The best model contained three compounds and had insignificantly high AUC value 0.80 (see Figure 87, Model III). It can be concluded that using the pharmacophore model as a pre-alignment tool for shape-based models development can be a reasonable strategy, but in this case the models obtained with pre-alignment showed better results.



**Figure 87.** The shape-based ROCS models and corresponding ROC curves. Model I is obtained on ligands from first training; Model II is obtained on ligands from the second training set which fit the pharmacophore model; Model III is obtained on ligands from the second training set pre-aligned in MOE (the threshold (TanimotoCombo score = 0.67) for virtual screening corresponds to the point (\*) where the gradient of ROC decreases.).

All possible combinations of the obtained models were tested for usage in consensus prediction. Below, the prediction performance derived from consensus prediction (Figure 88) of all 3 models was demonstrated, at AUC = 0.80. Using the consensus predictions didn't significantly improve the predictive power, that's why for the purpose of virtual screening it was decided to choose the model obtained on the second training set with AUC = 0.80 (see Figure 87, Model III). TanimotoCombo score = 0.67 was chosen as threshold, it leads to *precision* = 0.91 and *recall* = 0.49. The chosen model is comparable to developed pharmacophore model since it contains the same set of features in same position: anionic center, five hydrogen bond acceptors, four aromatic rings and hydrophobic feature.



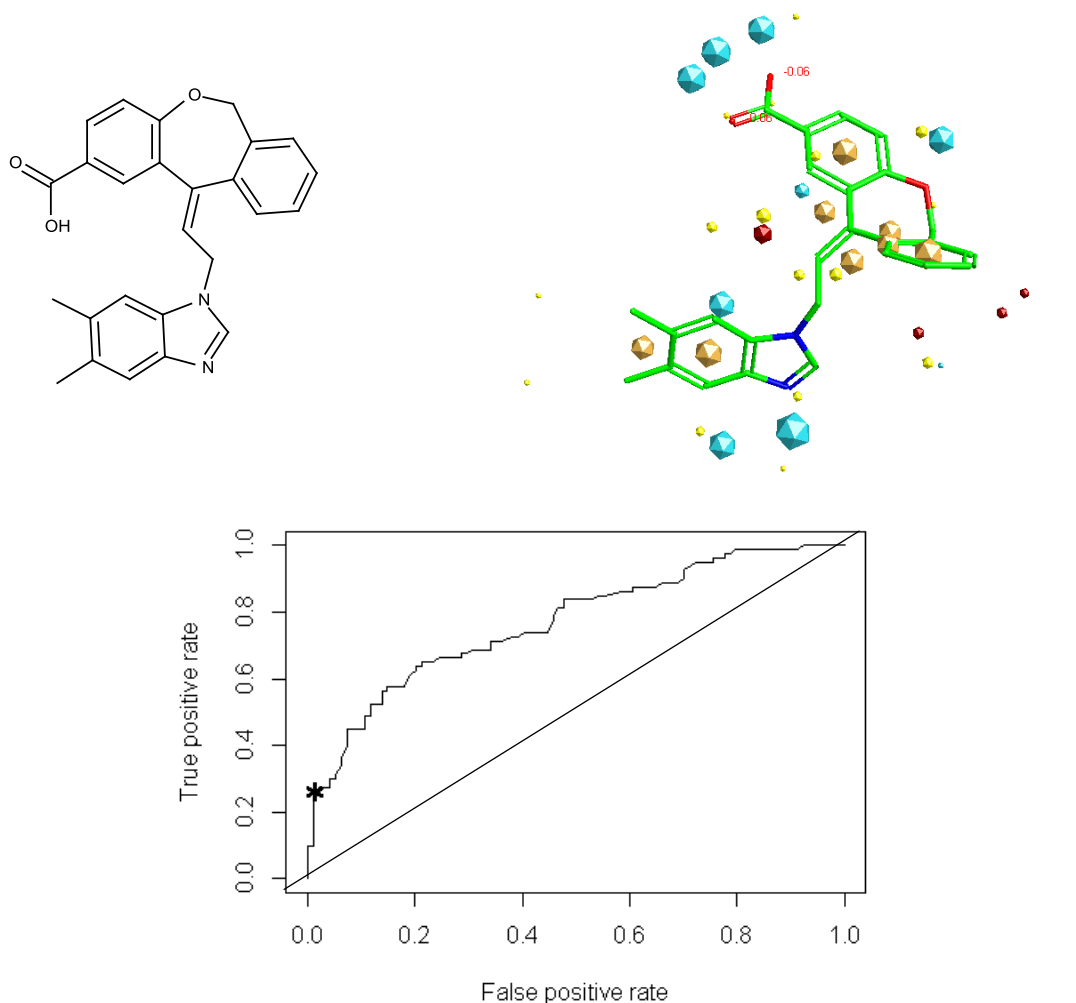
**Figure 88.** The consensus prediction based on five generated models (AUC = 0.80).

## 9.5. Molecular fields similarity

The FieldAlign software has been used for molecular field similarity of TP antagonists. As well as for ROCS it's necessary to choose proper "bioactive" conformers for molecular field generation in FieldAlign. For model generation, the most active compound which is rather conformationally rigid with  $pK_i = 8.92$  was chosen as a reference. The conformation of the chosen compound corresponded to the one obtained by pharmacophore model (cluster 1, Figure 86) was taken for model generation. The molecular fields were automatically generated by the FieldAlign software. The developed model was validated with *Affinity* dataset. The screening was performed using Normal settings from FieldAlign, including generation of 100 conformations with 0.5kcal/mol for each compound. The obtained field similarity was used to obtain ROC curve and AUC value to assess the generated model. As well as for ROCS models the *Affinity* dataset was divided on classes with threshold  $pK_i = 7$ . The generated molecular field



model showed high AUC value (see Figure 89) and was chosen for virtual screening. For the purpose of virtual screening as a threshold the Similarity value = 0.67 was chosen, it correspond to *precision* = 0.78 and *recall* = 0.49.

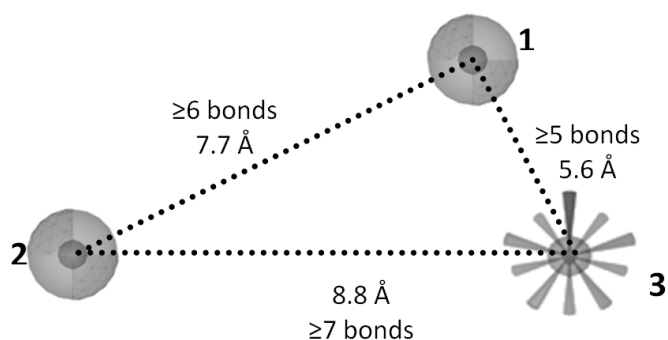


**Figure 89.** The compound chosen for virtual screening (AUC=0.79). The molecular fields are represented as points: negative field (blue points), positive field (red), van der Waals surface field (yellow), hydrophobic field (beige). The threshold (Similarity value = 0.67) for virtual screening corresponds to the point (\*) where the gradient of ROC decreases.

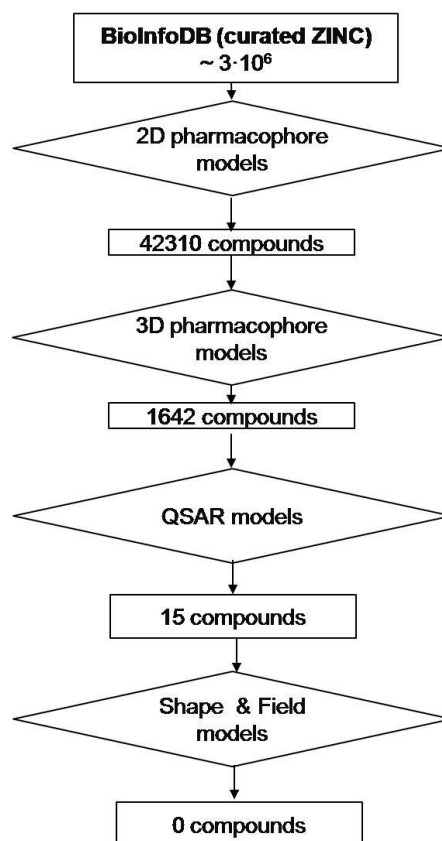
## 9.6. Virtual screening of BioInfoDB

The BioInfoDB database has been screened for novel TP receptor antagonists, using the multistage workflow comprised all developed models. First, all compounds have been filtered by a simplified 2D pharmacophore (Figure 90), which had been derived from 3D ligand-based pharmacophore (cluster 1, Figure 85). The three most restrictive features of the 3D pharmacophore model have been transferred to the 2D pharmacophore: the negative charge and two H-bond acceptors. Distances between selected features in 3D pharmacophore model have

been translated in minimum number of single carbon-carbon bonds, which could cover specified distances. Filtering has been performed by an in-house Python script, which used OpenEye OEChem toolkit. About 42000 compounds have been fit the obtained simplified 2D pharmacophore model (Figure 90). For the passing compounds, in case of unspecified stereocenters all possible stereoisomers have been generated with *Omega*, and then for all compounds at most 50 conformers, within 10 kcal/mol, were built. These have been subsequently screened with the 3D pharmacophore model. Compounds fitting the pharmacophore model have been, at a next stage, evaluated by the selected ROCS and molecular field models and 2D QSAR models.



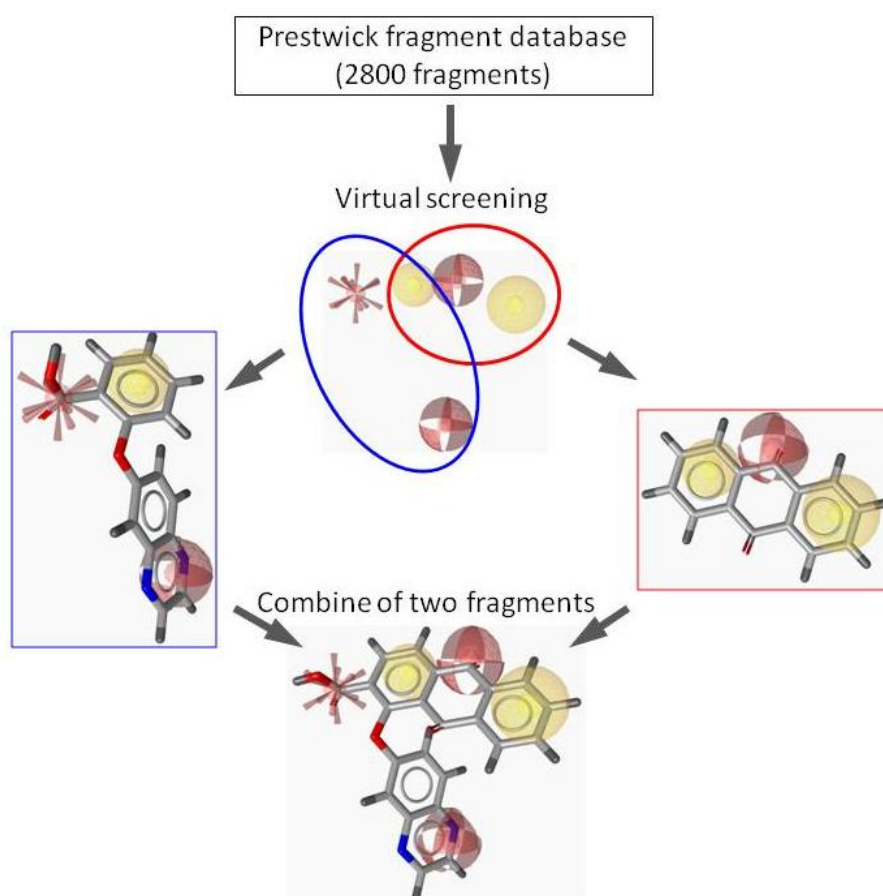
**Figure 90.** The simplified 2D pharmacophore model derived from the 3D ligand-based pharmacophore model (the spheres – H-bond acceptors, a star – center of negative charge).



**Figure 91.** Multistage workflow used for virtual screening of BioInfoDB.

### 9.7. Library design by pharmacophore fragment-based approach

Since screening of the database of commercially available compounds didn't return any reliable hits, a virtual library of potential antagonists of TP receptor has been generated using the novel pharmacophore fragment-based approach. 3D ligand-based pharmacophore model obtained earlier (cluster 1, see Figure 86) has been split on two parts each consisted of three features and one feature of them has been common (Figure 92). The Prestwick fragment database, which contains 2800 fragments, derived from known drugs, has been screened on both new pharmacophores, and fitted fragments have been combinatorial linked together in molecules of new library using MOE. Thus 171 compounds have been generated.

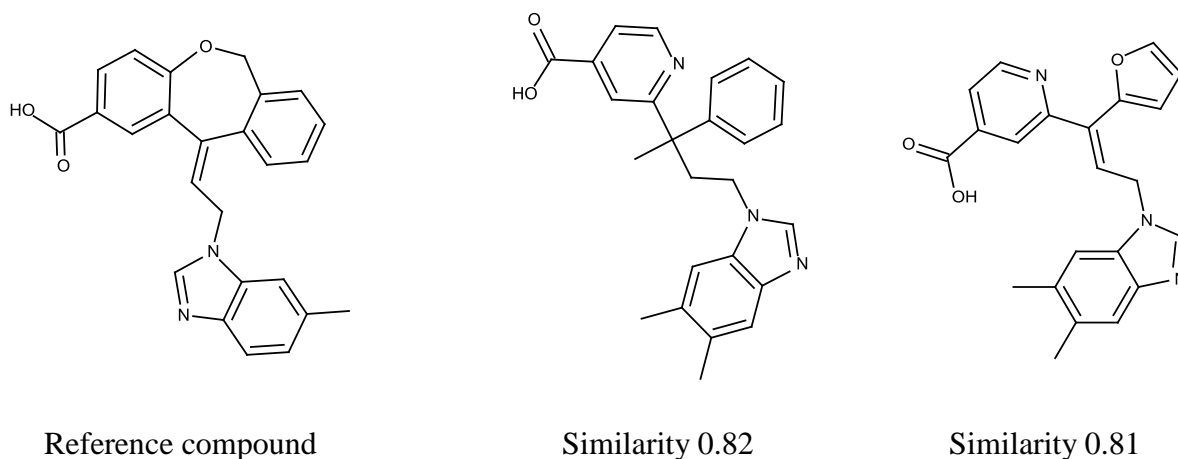


**Figure 92.** Library design using pharmacophore fragment-based approach.

### 9.8. Library design based on molecular fields similarity

As an alternative for the previous approach, bioisosteric replacement based on molecular fields similarity has been applied for generation of another virtual library of potential antagonists of TP receptor. The highly active compound ( $pK_i = 8.92$ ) from the *Affinity* dataset (Figure 89) has been chosen as a lead for further structural optimization. Conformation of this compound has been fixed and corresponded to the 3D pharmacophore model.

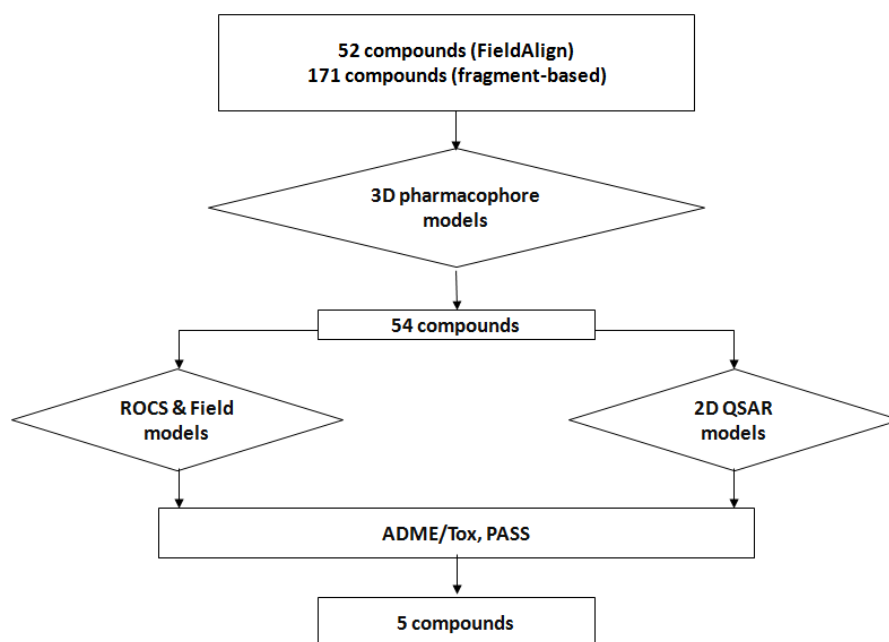
The complementarity of the molecular fields in terms of four properties, positive and negative electrostatics, van der Waals attraction and hydrophobicity, are considered while comparison of molecules with the lead one. Manual replacement and changing of some structural fragments gave 52 novel compounds with high molecular fields similarity, which has been estimated by FieldAlign program.



**Figure 93.** Example of manual modification.

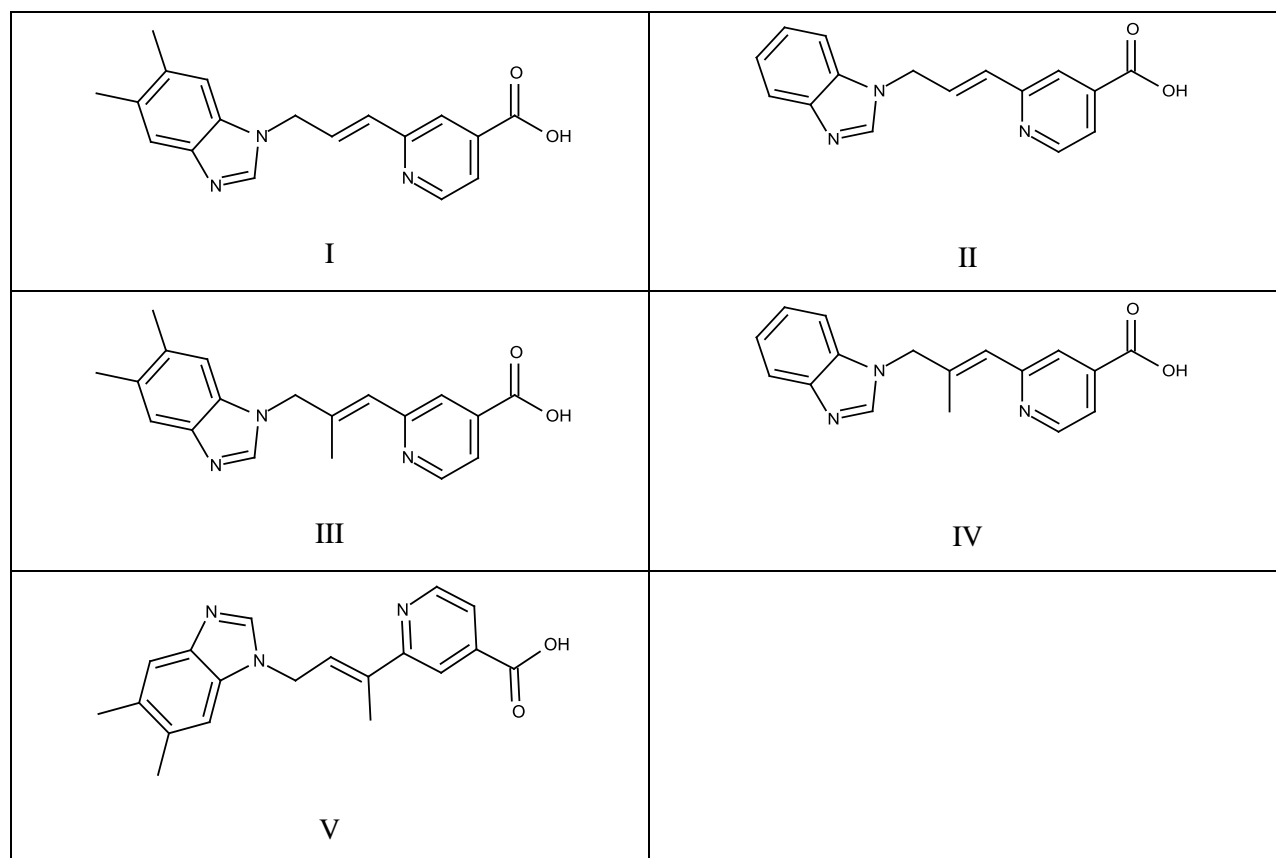
### 9.9. Virtual screening of designed libraries

The same workflow as for the BioInfoDB has been applied. All designed datasets have been combined and screened on 3D ligand-based pharmacophore model, followed by comparison of molecular fields, molecular shapes and prediction of affinity for TP receptor and anti-aggregation activity by consensus QSAR models. Additionally selected compounds have been estimated by PASS. The compounds were filtered out if the difference between  $P_a$  and  $P_i$  of possible adverse effects or toxicity exceeded 0.6. The remained compounds possessed satisfactory predicted solubility values (0.001-0.0004 mol/l) and were not mutagenic. According to prediction of PASS these compounds didn't have any toxic or side effects.



**Figure 94.** Workflow used for virtual screening of designed library.

As a result of virtual screening five perspective compounds (Figure 95) were selected for synthesis and further biological evaluation.



**Figure 95.** Design of new antagonists of thromboxane A<sub>2</sub> receptor: the hits recommended for the synthesis and experimental tests.

## 9.10. Conclusions

The computer-aided design of new potent antagonists of thromboxane A2 receptor is described in the current section. At first, experimentally available information on affinity for TP and anti-aggregation activity have been used to develop different models, such as QSAR models (CF), ligand-based 3D pharmacophore model (LigandScout), 2D pharmacophore model, shape-based model (ROCS) and molecular field-based model (FieldAlign, Cresset). Application of these models to virtual screening of BioInfoDB database resulted in no hits. For this reason, the design of virtual library was performed using two approaches: pharmacophore fragments-based approach and bioisosteric replacement based on molecular fields similarity. The virtual screening of the designed virtual library resulted in five hits. Additionally the found hits were assessed on some ADME/Tox properties (mutagenicity, solubility) and on possible biological responses (PASS). These compounds have been already synthesized in PCI but biological tests have not been completed yet.

## **General conclusions**

This work deals with computer-aided design of new anti-thrombotic agents able to inhibit two types of receptors located on the surface of the platelets. The first one - integrin  $\alpha_{IIb}\beta_3$  - is responsible for the interaction of activated platelets with fibrinogen to form clots, whereas the second one - thromboxane A2 - is responsible for platelet activation by one of agonists excreted by adjacent platelets.

Three commercialized antagonists of the  $\alpha_{IIb}\beta_3$  receptor - abciximab, eptifibatide and tirofiban - have severe side effects, such as thrombocytopenia and bleeding. The former is associated with the conformational changes of integrin  $\alpha_{IIb}\beta_3$ , which are induced by binding of "classical" antagonists – RGD peptidomimetics. In contrast, recently discovered non-classical  $\alpha_{IIb}\beta_3$  receptor antagonists RUC-1 and RUC-2 bind its closed form. This doesn't induce the conformational changes of the protein and, hence, reduces the risk of undesirable side effects.

Antagonists of thromboxane A2 receptor could be particularly useful in treatment of acute myocardial ischemia, heart failure and exhibit cardioprotective effects. However, up to now no marketed drugs of this type are available because of insufficient potency of studied compounds.

Our goal was to design new anti-thrombotic agents, which are, at least, as potent as commercial drugs and, desirably, don't possess side effects. For  $\alpha_{IIb}\beta_3$  receptor, we considered inhibitors with classical and non-classical binding modes. Thus, the entire work was split onto three projects: development of (i) classical RGD-peptidomimetics - antagonists of  $\alpha_{IIb}\beta_3$  receptor; (ii) RUC-1 and RUC-2 analogs and (iii) antagonists of thromboxane A2 receptor. Each project included several stages: (i) assembling and curation of available experimental data, (ii) models obtaining and validation, (iii) screening a database of commercial compounds, and (iv) generation and screening a virtual library.

Various structure- and ligand-based methods have been used: QSAR, 2D and 3D pharmacophores, ligand to protein docking, shape and field similarity techniques. Although these approaches are conceptually different, they are closely related. Thus, in pharmacophore modeling of peptidomimetics their conformational space was restricted using the information extracted from X-ray structure of  $\alpha_{IIb}\beta_3$  – ligand complexes. In turn, ligand- and structure-based pharmacophores obtained with LigandScout were used to select several reference molecules for shape and field similarity studies performed, respectively, with the ROCS (OpenEye) and FieldAlign (Cresset) programs. Rigorous re-docking and cross-docking procedures have been used to prepare the protein binding pocket (with or without water molecules) and to select the most appropriate tool among three available programs – MOE, FlexX and PLANTS. Additionally, a correlation between scoring function and experimental activities of the compounds has been checked. The screening workflow included also 2D pharmacophores



(Python script developed in this work) and QSAR models based on the fragment descriptors (Simplex and SMF/ISIDA) invented in the partner laboratories in Odessa and Strasbourg. Previously developed in our laboratories models for aqueous solubility, as well as the PASS program have been used to obtain ADME/Tox profiles of selected hits.

Surprisingly, virtual screening of large commercial databases using all these complementary approaches resulted in no valuable hits. This motivated us to generate focused combinatorial library screening of which led to very few compounds suggested for the synthesis and biological studies.

To sum up, two out of three projects have been accomplished by preparation of real compounds with minimal synthetic efforts. Two compounds, each representing by two enantiomers, – “classical” antagonists of the  $\alpha_{IIb}\beta_3$  receptor have been synthesized. They display high binding and anti-aggregation affinities; moreover, one of them was found more active than commercial drug tirofiban. Developed in this work QSAR model correctly estimated their activity values, whereas the docking score was a good indication of relative potencies of R and S enantiomers. Four “non-classical”  $\alpha_{IIb}\beta_3$  antagonists (RUC-2) analogues have been synthesized; one of this compounds was found better protein binder than tirofiban. The only part of this work which still needs experimental validation concerns the development of new antagonists of thromboxane A2 receptors. Five molecules selected in virtual screening will be synthesized and tested in the nearest future in the Bogatsky Physico-Chemical Institute of Odessa.

Although, some success has been achieved in this work, the design of new antithrombotics is still in its infancy. At least, two new projects could be suggested as perspective studies in this field. Up to now, only one binding site of integrin  $\alpha_{IIb}\beta_3$  is considered in drug design projects. We believe that an effort should be done to discover complementary allosteric binding sites which could give rise to discovery of novel types of ligands. Another suggestion concerns design of antagonists of the thromboxane A2 receptor. Experimental X-ray structure of this protein is not still available. Therefore, its modeled structure could become a valuable support for further structure-based studies.

It should be noticed that this work may have important social impact. The point is that the commercial antithrombotic drugs abciximab, eptifibatide and tirofiban are rather expensive, and, therefore, their use may not be afforded by the part of Ukrainian population. Although, the distance between designed here antagonists of the  $\alpha_{IIb}\beta_3$  receptor and real drugs is still too long, this work could become a first step in the development of new generation of antithrombotics.

Last, but not least: this PhD thesis has become a bridge between the research teams from the University of Strasbourg and the Bogatsky Physico-Chemical Institute of Odessa. A success

achieved in this work resulted from joint efforts of theoretical and experimental chemists from both institutions.

## References

1. Blue R, Murcia M, Karan C, Jirouskova M, Collier BS. 2008. Application of high-throughput screening to identify a novel alpha IIb-specific small-molecule inhibitor of alpha IIb beta 3-mediated platelet interaction with fibrinogen. *Blood* 111: 1248-56
2. Zhu JQ, Choi WS, McCoy JG, Negri A, Zhu JH, Naini S, Li JH, Shen M, Huang WW, Bougie D, Rasmussen M, Aster R, Thomas CJ, Filizola M, Springer TA, Collier BS. 2012. Structure-Guided Design of a High-Affinity Platelet Integrin alpha(IIb)beta(3) Receptor Antagonist That Disrupts Mg<sup>2+</sup> Binding to the MIDAS. *Science Translational Medicine* 4
3. Colman RW, Marder VJ, Clowes AW, George JN, Goldhaber SZ. 2005. *Hemostasis and thrombosis: basic principles and clinical practice*: Lippincott Williams & Wilkins
4. Becker RC. 2000. *Fibrinolytic and antithrombotic therapy: theory, practice, and management*: Oxford University Press, USA
5. 2012. WHO. World Health Statistics 2012.
6. Reiningner AJ, Bernlochner I, Penz SM, Ravanat C, Smethurst P, Farndale RW, Gachet C, Brandl R, Siess W. 2010. A 2-Step Mechanism of Arterial Thrombus Formation Induced by Human Atherosclerotic Plaques. *Journal of the American College of Cardiology* 55: 1147-58
7. Lundberg AM, Hansson GK. 2010. Innate immune signals in atherosclerosis. *Clinical Immunology* 134: 5-24
8. Libby P. 2009. Molecular and cellular mechanisms of the thrombotic complications of atherosclerosis. *Journal of Lipid Research* 50: S352-S7
9. Furie B, Furie BC. 2008. Mechanisms of disease: Mechanisms of thrombus formation. *New England Journal of Medicine* 359: 938-49
10. Katsuda S, Kaji T. 2003. Atherosclerosis and extracellular matrix. *Journal of atherosclerosis and thrombosis* 10: 267-74
11. De Meyer GRY, Hoylaerts MF, Kockx MM, Yamamoto H, Herman AG, Bult H. 1999. Intimal Deposition of Functional von Willebrand Factor in Atherogenesis. *Arteriosclerosis, Thrombosis, and Vascular Biology* 19: 2524-34
12. Barnes MJ, Farndale RW. 1999. Collagens and atherosclerosis. *Experimental Gerontology* 34: 513-25
13. Penz S, Reiningner AJ, Brandl R, Goyal P, Rabie T, Bernlochner I, Rother E, Goetz C, Engelmann B, Smethurst PA, Ouwehand WH, Farndale R, Nieswandt B, Siess W. 2005. Human atheromatous plaques stimulate thrombus formation by activating platelet glycoprotein VI. *The FASEB Journal* 19: 898-909
14. Varga-Szabo D, Pleines I, Nieswandt B. 2008. Cell adhesion mechanisms in platelets. *Arteriosclerosis Thrombosis and Vascular Biology* 28: 403-12
15. Savage B, Almus-Jacobs F, Ruggeri ZM. 1998. Specific synergy of multiple substrate-receptor interactions in platelet thrombus formation under flow. *Cell* 94: 657-66
16. Savage B, Saldivar E, Ruggeri ZM. 1996. Initiation of platelet adhesion by arrest onto fibrinogen or translocation on von Willebrand factor. *Cell* 84: 289-97
17. Gralnick HR, Kramer WS, McKeown LP, Garfinkel L, Pinot A, Williams SB, Krutzsch H. 1996. Platelet adhesion at high shear rates: The roles of von Willebrand factor/GPIb and the beta(1) integrin alpha(2)beta(1). *Thrombosis Research* 81: 113-9
18. Siljander PRM, Munnix ICA, Smethurst PA, Deckmyn H, Lindhout T, Ouwehand WH, Farndale RW, Heemskerk JWM. 2004. Platelet receptor interplay regulates collagen-induced thrombus formation in flowing human blood. *Blood* 103: 1333-41
19. Clemetson KJ. 2012. Platelets and Primary Haemostasis. *Thrombosis Research* 129: 220-4
20. Harrison P, Wilbourn B, Debili N, Vainchenker W, Bretongorius J, Lawrie AS, Masse JM, Savidge GF, Cramer EM. 1989. Uptake of Plasma-Fibrinogen into the Alpha-Granules of Human Megakaryocytes and Platelets. *Journal of Clinical Investigation* 84: 1320-4
21. George JN. 2000. Platelets. *Lancet* 355: 1531-9
22. Offermanns S. 2006. Activation of platelet function through G protein-coupled receptors. *Circulation Research* 99: 1293-304

23. Ding ZG, Tuluc F, Bandivadekar KR, Zhang LL, Jin JG, Kunapuli SP. 2005. Arg333 and Arg334 in the COOH terminus of the human P2Y(1) receptor are crucial for G(q) coupling. *American Journal of Physiology-Cell Physiology* 288: C559-C67
24. Jantzen HM, Milstone DS, Gousset L, Conley PB, Mortensen RM. 2001. Impaired activation of murine platelets lacking G alpha(i2). *Journal of Clinical Investigation* 108: 477-83
25. Gachet C. 2001. ADP receptors of platelets and their inhibition. *Thrombosis and Haemostasis* 86: 222-32
26. Jin J, Kunapuli SP. 1998. Coactivation of two different G protein-coupled receptors is essential for ADP-induced platelet aggregation. *Proc Natl Acad Sci U S A* 95: 8070-4
27. Ally AI, Horrobin DF. 1980. Thromboxane-A2 in Blood-Vessel Walls and Its Physiological Significance - Relevance to Thrombosis and Hypertension. *Prostaglandins and Medicine* 4: 431-8
28. Arita H, Nakano T, Hanasaki K. 1989. Thromboxane-A2 - Its Generation and Role in Platelet Activation. *Progress in Lipid Research* 28: 273-301
29. Namba T, Sugimoto Y, Hirata M, Hayashi Y, Honda A, Watabe A, Negishi M, Ichikawa A, Narumiya S. 1992. Mouse Thromboxane-A2 Receptor - Cdna Cloning, Expression and Northern Blot Analysis. *Biochemical and Biophysical Research Communications* 184: 1197-203
30. Offermanns S, Laugwitz KL, Spicher K, Schultz G. 1994. G-Proteins of the G(12) Family Are Activated Via Thromboxane a(2) and Thrombin Receptors in Human Platelets. *Proceedings of the National Academy of Sciences of the United States of America* 91: 504-8
31. Klages B, Brandt U, Simon MI, Schultz G, Offermanns S. 1999. Activation of G(12)/G(13) results in shape change and Rho/Rho-kinase-mediated myosin light chain phosphorylation in mouse platelets. *Journal of Cell Biology* 144: 745-54
32. Siffert W, Siffert G, Scheid P, Akkerman JWN. 1990. Na<sup>+</sup>/H<sup>+</sup> Exchange Modulates Ca<sup>2+</sup> Mobilization in Human-Platelets Stimulated by Adp and the Thromboxane Mimetic U-46619. *Journal of Biological Chemistry* 265: 719-25
33. Giesen PLA, Rauch U, Bohrmann B, Kling D, Roqué M, Fallon JT, Badimon JJ, Himer J, Riederer MA, Nemerson Y. 1999. Blood-borne tissue factor: Another view of thrombosis. *Proceedings of the National Academy of Sciences* 96: 2311-5
34. Wilcox JN, Smith KM, Schwartz SM, Gordon D. 1989. Localization of tissue factor in the normal vessel wall and in the atherosclerotic plaque. *Proceedings of the National Academy of Sciences* 86: 2839-43
35. Toschi V, Gallo R, Lettino M, Fallon JT, Gertz SD, Ferná'ndez-Ortiz A, Chesebro JH, Badimon L, Nemerson Y, Fuster V, Badimon JJ. 1997. Tissue Factor Modulates the Thrombogenicity of Human Atherosclerotic Plaques. *Circulation* 95: 594-9
36. Nesheim ME, Taswell JB, Mann KG. 1979. Contribution of Bovine Factor-V and Factor-Va to the Activity of Prothrombinase. *Journal of Biological Chemistry* 254: 952-62
37. Osterud B, Rapaport SI. 1977. Activation of factor IX by the reaction product of tissue factor and factor VII: additional pathway for initiating blood coagulation. *Proceedings of the National Academy of Sciences* 74: 5260-4
38. von dem Borne PA, Meijers JC, Bouma BN. 1995. Feedback activation of factor XI by thrombin in plasma results in additional formation of thrombin that protects fibrin clots from fibrinolysis. *Blood* 86: 3035-42
39. Hockin MF, Jones KC, Everse SJ, Mann KG. 2002. A Model for the Stoichiometric Regulation of Blood Coagulation. *Journal of Biological Chemistry* 277: 18322-33
40. Gailani D, Renne T. 2007. Intrinsic pathway of coagulation and arterial thrombosis. *Arterioscler Thromb Vasc Biol* 27: 2507-13
41. Coughlin SR. 2000. Thrombin signalling and protease-activated receptors. *Nature* 407: 258-64
42. Vu TKH, Hung DT, Wheaton VI, Coughlin SR. 1991. Molecular-Cloning of a Functional Thrombin Receptor Reveals a Novel Proteolytic Mechanism of Receptor Activation. *Cell* 64: 1057-68
43. Fukuhara S, Murga C, Zohar M, Igishi T, Gutkind JS. 1999. A novel PDZ domain containing guanine nucleotide exchange factor links heterotrimeric G proteins to Rho. *Journal of Biological Chemistry* 274: 5868-79

44. Falker K, Haglund L, Gunnarsson P, Nylander M, Lindahl TL, Grenegard M. 2011. Protease-activated receptor 1 (PAR1) signalling desensitization is counteracted via PAR4 signalling in human platelets. *Biochemical Journal* 436: 469-80
45. Henschen A, Lottspeich F, Kehl M, Southan C. 1983. Covalent Structure of Fibrinogen. *Annals of the New York Academy of Sciences* 408: 28-43
46. Henshen-Edman AH. 1996. Human Fibrinogen Occurs as over 1 Million Non-Identical Molecules. *International Conference: Methods in protein structure analysis*: 1 - 527
47. Laudano AP, Doolittle RF. 1980. Studies on Synthetic Peptides That Bind to Fibrinogen and Prevent Fibrin Polymerization - Structural Requirements, Number of Binding-Sites, and Species-Differences. *Biochemistry* 19: 1013-9
48. La Corte ALC, Philippou H, Ariens RAS. 2011. Role of Fibrin Structure in Thrombosis and Vascular Disease. *Advances in Protein Chemistry and Structural Biology: Protein Structure and Diseases, Vol 83* 83: 75-127
49. Doolittle RF. 1984. Fibrinogen and Fibrin. *Annual Review of Biochemistry* 53: 195-229
50. Mosesson MW, Siebenlist KR, Meh DA. 2001. The structure and biological features of fibrinogen and fibrin. *Fibrinogen* 936: 11-30
51. Wagner CL, Mascelli MA, Neblock DS, Weisman HF, Collier BS, Jordan RE. 1996. Analysis of GPIIb/IIIa receptor number by quantification of 7E3 binding to human platelets. *Blood* 88: 907-14
52. Cramer EM, Savidge GF, Vainchenker W, Berndt MC, Pidard D, Caen JP, Masse JM, Bretonogorius J. 1990. Alpha-Granule Pool of Glycoprotein-IIb-IIIa in Normal and Pathological Platelets and Megakaryocytes. *Blood* 75: 1220-7
53. Gogstad GO, Hagen I, Korsmo R, Solum NO. 1981. Characterization of the Proteins of Isolated Human-Platelet Alpha-Granules - Evidence for a Separate Alpha-Granule-Pool of the Glycoprotein-IIb and Glycoprotein-IIia. *Biochimica et Biophysica Acta* 670: 150-62
54. Hynes RO. 1992. Integrins - Versatility, Modulation, and Signaling in Cell-Adhesion. *Cell* 69: 11-25
55. Takagi J, Petre BM, Walz T, Springer TA. 2002. Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling. *Cell* 110: 599-611
56. Luo BH, Carman CV, Springer TA. 2007. Structural basis of integrin regulation and signaling. *Annual Review of Immunology* 25: 619-47
57. Shimaoka M, Takagi J, Springer TA. 2002. Conformational regulation of integrin structure and function. *Annual Review of Biophysics and Biomolecular Structure* 31: 485-516
58. Zhu JQ, Zhu JH, Negri A, Provasi D, Filizola M, Collier BS, Springer TA. 2010. Closed headpiece of integrin alpha(IIb)beta(3) and its complex with an alpha(IIb)beta(3)-specific antagonist that does not induce opening. *Blood* 116: 5050-9
59. Otoole TE, Ylanne J, Culley BM. 1995. Regulation of Integrin Affinity States through an Npxy Motif in the Beta-Subunit Cytoplasmic Domain. *Journal of Biological Chemistry* 270: 8553-8
60. Hughes PE, DiazGonzalez F, Leong L, Wu CY, McDonald JA, Shattil SJ, Ginsberg MH. 1996. Breaking the integrin hinge - A defined structural constraint regulates integrin signaling. *Journal of Biological Chemistry* 271: 6571-4
61. Moser M, Nieswandt B, Ussar S, Pozgajova M, Fassler R. 2008. Kindlin-3 is essential for integrin activation and platelet aggregation. *Nature Medicine* 14: 325-30
62. Tadokoro S, Shattil SJ, Eto K, Tai V, Liddington RC, de Pereda JM, Ginsberg MH, Calderwood DA. 2003. Talin binding to integrin beta tails: A final common step in integrin activation. *Science* 302: 103-6
63. Anthis NJ, Campbell ID. 2011. The tail of integrin activation. *Trends Biochem Sci* 36: 191-8
64. Gartner TK, Bennett JS. 1985. The Tetrapeptide Analog of the Cell Attachment Site of Fibronectin Inhibits Platelet-Aggregation and Fibrinogen Binding to Activated Platelets. *Journal of Biological Chemistry* 260: 1891-4
65. Andrieux A, Hudryclergeon G, Ryckewaert JJ, Chapel A, Ginsberg MH, Plow EF, Marguerie G. 1989. Amino-Acid Sequences in Fibrinogen Mediating Its Interaction with Its Platelet Receptor, Gpiibiii. *Journal of Biological Chemistry* 264: 9258-65

66. Kloczewiak M, Timmons S, Lukas TJ, Hawiger J. 1984. Platelet Receptor Recognition Site on Human-Fibrinogen - Synthesis and Structure-Function Relationship of Peptides Corresponding to the Carboxy-Terminal Segment of the Gamma-Chain. *Biochemistry* 23: 1767-74
67. James SH. 2009. Hematology pharmacology: anticoagulant, antiplatelet, and procoagulant agents in practice. *AACN Adv Crit Care* 20: 177-92
68. Di Nisio M, Middeldorp S, Buller HR. 2005. Drug therapy - Direct thrombin inhibitors. *New England Journal of Medicine* 353: 1028-40
69. Kearon C, Kahn SR, Agnelli G, Goldhaber S, Raskob GE, Comerota AJ. 2008. Antithrombotic therapy for venous thromboembolic disease. *Chest* 133: 454s-545s
70. Lansberg MG, O'Donnell MJ, Khatri P, Lang ES, Nguyen-Huynh MN, Schwartz NE, Sonnenberg FA, Schulman S, Vandvik PO, Spencer FA, Alonso-Coello P, Guyatt GH, Akl EA. Antithrombotic and thrombolytic therapy for ischemic stroke: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 141: e601S-36S
71. Kearon C, Akl EA, Comerota AJ, Prandoni P, Bounameaux H, Goldhaber SZ, Nelson ME, Wells PS, Gould MK, Dentali F, Crowther M, Kahn SR. 2012. Antithrombotic Therapy for VTE Disease Antithrombotic Therapy for VTE Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *CHEST Journal* 141: e419S-e94S
72. Becker RC, Green D. 2006. *Antithrombotic Therapy. 4th ed. West Islip, NY: Professional Communications*
73. Harrington RA, Becker RC, Cannon CP, Gutterman D, Lincoff AM, Popma JJ, Steg G, Guyatt GH, Goodman SG. 2008. Antithrombotic therapy for non-ST-segment elevation acute coronary syndromes: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 133: 670S-707S
74. Goodman SG, Menon V, Cannon CP, Steg G, Ohman EM, Harrington RA. 2008. Acute ST-segment elevation myocardial infarction: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 133: 708S-75S
75. Geerts WH, Bergqvist D, Pineo GF, Heit JA, Samama CM, Lassen MR, Colwell CW. 2008. Prevention of venous thromboembolism: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 133: 381S-453S
76. Kearon C, Kahn SR, Agnelli G, Goldhaber S, Raskob GE, Comerota AJ. 2008. Antithrombotic therapy for venous thromboembolic disease: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 133: 454S-545S
77. Schulman S, Beyth RJ, Kearon C, Levine MN. 2008. Hemorrhagic complications of anticoagulant and thrombolytic treatment: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 133: 257S-98S
78. Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G. 2008. Pharmacology and management of the vitamin K antagonists: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 133: 160S-98S
79. Francis CW, KK. 2006. Principles of antithrombotic therapy. *Lichtman MA, Beutler E, Kipps TJ, Seligsohn U, Kaushansky K, Prchal JT, eds. Williams Hematology. 7th ed. New York, NY: McGraw-Hill: 283-300*
80. JS.Fedan. 2004. Anticoagulant, antiplatelet, and fibrinolytic (thrombolytic) drugs. *In: Craig CR, Stitzel RE, eds. Modern Pharmacology With Clinical Applications. 6th ed. Philadelphia, PA: Lippincott Williams & Wilkins: 256-67*
81. Collen D, Stump DC, Gold HK. 1988. Thrombolytic Therapy. *Annual Review of Medicine* 39: 405-23
82. Steg PG, James S, Harrington RA, Ardissino D, Becker RC, Cannon CP, Emanuelsson H, Finkelstein A, Husted S, Katus H, Kilhamn J, Olofsson S, Storey RF, Weaver WD, Wallentin L, for the PSG. 2010. Ticagrelor Versus Clopidogrel in Patients With ST-Elevation Acute Coronary Syndromes Intended for Reperfusion With Primary Percutaneous Coronary Intervention / Clinical Perspective. *Circulation* 122: 2131-41

83. Eikelboom JW, Hirsh J, Spencer FA, Baglin TP, Weitz JI. Antiplatelet drugs: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 141: e89S-119S
84. Michelson AD. 2010. Antiplatelet therapies for the treatment of cardiovascular disease. *Nat Rev Drug Discov* 9: 154-69
85. 2002. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ* 324: 71-86
86. Mehta SR, Yusuf S. 2000. The Clopidogrel in Unstable angina to prevent Recurrent Events (CURE) trial programme; rationale, design and baseline characteristics including a meta-analysis of the effects of thienopyridines in vascular disease. *Eur Heart J* 21: 2033-41
87. Wiviott SD, Braunwald E, McCabe CH, Montalescot G, Ruzyllo W, Gottlieb S, Neumann F, Ardissino D, De Servi S, Murphy SA, Riesmeyer J, Weerakkody G, Gibson CM, Antman EM. 2007. Prasugrel versus clopidogrel in patients with acute coronary syndromes. *New England Journal of Medicine* 357: 2001-15
88. Boersma E, Harrington RA, Moliterno DJ, White H, Theroux P, Van de Werf F, de Torbal A, Armstrong PW, Wallentin LC, Wilcox RG, Simes J, Califf RM, Topol EJ, Simoons ML. 2002. Platelet glycoprotein IIb/IIIa inhibitors in acute coronary syndromes: a meta-analysis of all major randomised clinical trials. *Lancet* 359: 189-98
89. De Luca G, Navarese E, Marino P. 2009. Risk profile and benefits from Gp IIb-IIIa inhibitors among patients with ST-segment elevation myocardial infarction treated with primary angioplasty: a meta-regression analysis of randomized trials. *European Heart Journal* 30: 2705-13
90. Quinn MJ, Fitzgerald DJ. 1999. Ticlopidine and clopidogrel. *Circulation* 100: 1667-72
91. Husted S. 2007. New developments in oral antiplatelet therapy. *European Heart Journal Supplements* 9: D20-D7
92. Savi P, Herbert JM, Pflieger AM, Dol F, Delebassee D, Combalbert J, Defreyne G, Maffrand JP. 1992. Importance of Hepatic-Metabolism in the Antiaggregating Activity of the Thienopyridine Clopidogrel. *Biochemical Pharmacology* 44: 527-32
93. Gurbel PA, Bliden KP, Hiatt BL, O'Connor CM. 2003. Clopidogrel for coronary stenting - Response variability, drug resistance, and the effect of pretreatment platelet reactivity. *Circulation* 107: 2908-13
94. Algaier I, Jakubowski JA, Asai F, Von Kugelgen I. 2008. Interaction of the active metabolite of prasugrel, R-138727, with cysteine 97 and cysteine 175 of the human P2Y(12) receptor. *Journal of Thrombosis and Haemostasis* 6: 1908-14
95. Cattaneo M, Podda GM. 2010. State of the art of new P2Y(12) antagonists. *Internal and Emergency Medicine* 5: 385-91
96. van Giezen JJJ, Humphries RG. 2005. Preclinical and clinical studies with selective reversible direct P2Y(12) antagonists. *Seminars in Thrombosis and Hemostasis* 31: 195-204
97. Harrington RA, Stone GW, McNulty S, White HD, Lincoff AM, Gibson CM, Pollack CV, Montalescot G, Mahaffey KW, Kleiman NS, Goodman SG, Amine M, Angiolillo DJ, Becker RC, Chew DP, French WJ, Leisch F, Parikh KH, Skerjanec S, Bhatt DL. 2009. Platelet Inhibition with Cangrelor in Patients Undergoing PCI. *New England Journal of Medicine* 361: 2318-29
98. Husted S, Emanuelsson H, Heptinstall S, Sandset PM, Wickens M, Peters G. 2006. Pharmacodynamics, pharmacokinetics, and safety of the oral reversible P2Y(12) antagonist AZD6140 with aspirin in patients with atherosclerosis: a double-blind comparison to clopidogrel with aspirin. *European Heart Journal* 27: 1038-47
99. Bliden KP, Tantry US, Storey RF, Jeong YH, Gesheff M, Wei C, Gurbel PA. 2011. The effect of ticagrelor versus clopidogrel on high on-treatment platelet reactivity: Combined analysis of the ONSET/OFFSET and RESPOND studies. *American Heart Journal* 162: 160-5
100. Wallentin L, Becker RC, Budaj A, Cannon CP, Emanuelsson H, Held C, Horrow J, Husted S, James S, Katus H, Mahaffey KW, Scirica BM, Skene A, Steg PG, Storey RF, Harrington RA, Investigators P. 2009. Ticagrelor versus Clopidogrel in Patients with Acute Coronary Syndromes. *New England Journal of Medicine* 361: 1045-57



101. Varenhorst C, Alstrom U, Scirica BM, Hogue CW, Asenblad N, Storey RF, Steg PG, Horrow J, Mahaffey KW, Becker RC, James S, Cannon CP, Brandrup-Wognsen G, Wallentin L, Held C. Factors contributing to the lower mortality with ticagrelor compared with clopidogrel in patients undergoing coronary artery bypass surgery. *J Am Coll Cardiol* 60: 1623-30
102. Kosoglou T, Reyderman L, Tiessen RG, van Vliet AA, Fales RR, Keller R, Yang B, Cutler DL. 2012. Pharmacodynamics and pharmacokinetics of the novel PAR-1 antagonist vorapaxar (formerly SCH 530348) in healthy subjects. *European Journal of Clinical Pharmacology* 68: 249-58
103. Weitz JI, Eikelboom JW, Samama MM. 2012. New antithrombotic drugs: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 141: e120S-51S
104. Tello-Montoliu A, Jover E, Rivera J, Valdes M, Angiolillo DJ, Marin F. 2012. New Perspectives in Antiplatelet Therapy. *Current Medicinal Chemistry* 19: 406-27
105. Angiolillo DJ, Guzman LA. 2008. Clinical overview of promising nonthienopyridine antiplatelet agents. *American Heart Journal* 156: S23-S8
106. Chackalamannil S, Wang YG, Greenlee WJ, Hu ZY, Xia Y, Ahn HS, Boykow G, Hsieh YS, Palamanda J, Agans-Fantuzzi J, Kurowski S, Graziano M, Chintala M. 2008. Discovery of a novel, orally active himbacine-based thrombin receptor antagonist (SCH 530348) with potent antiplatelet activity. *Journal of Medicinal Chemistry* 51: 3061-4
107. O'Donoghue ML, Bhatt DL, Wiviott SD, Goodman SG, Fitzgerald DJ, Angiolillo DJ, Goto S, Montalescot G, Zeymer U, Aylward PE, Guetta V, Dudek D, Ziecina R, Contant CF, Flather MD, Investigators L-A. 2011. Safety and Tolerability of Atopaxar in the Treatment of Patients With Acute Coronary Syndromes The Lessons From Antagonizing the Cellular Effects of Thrombin-Acute Coronary Syndromes Trial. *Circulation* 123: 1843+
108. Tello-Montoliu A, Tomasello SD, Ueno M, Angiolillo DJ. 2011. Antiplatelet therapy: thrombin receptor antagonists. *British Journal of Clinical Pharmacology* 72: 658-71
109. Schlansky B, Hwang JH. 2009. Prevention of nonsteroidal anti-inflammatory drug-induced gastropathy. *Journal of Gastroenterology* 44: 44-52
110. Serebruany VL, Steinhubl SR, Berger PB, Malinin AI, Baggish JS, Bhatt DL, Topol EJ. 2005. Analysis of risk of bleeding complications after different doses of Aspirin in 192,036 patients enrolled in 31 randomized controlled trials. *American Journal of Cardiology* 95: 1218-22
111. Pusch G, Feher G, Kotai K, Tibold A, Gasztonyi B, Feher A, Papp E, Lupkovics G, Szapary L. 2008. Aspirin Resistance: Focus on Clinical Endpoints. *Journal of Cardiovascular Pharmacology* 52: 475-84
112. Davi G, Santilli F, Vazzana N. Thromboxane receptors antagonists and/or synthase inhibitors. *Handb Exp Pharmacol*: 261-86
113. Hennerici MG, Bots ML, Ford I, Laurent S, Touboul PJ. 2010. Rationale, design and population baseline characteristics of the PERFORM Vascular Project: an ancillary study of the Prevention of cerebrovascular and cardiovascular Events of ischemic origin with teRutroban in patients with a history of ischemic stroke or transient ischaemic attack (PERFORM) trial. *Cardiovascular Drugs and Therapy* 24: 175-80
114. Vilahur G, Casani L, Badimon L. 2007. A thromboxane A<sub>2</sub>/prostaglandin H<sub>2</sub> receptor antagonist (S18886) shows high antithrombotic efficacy in an experimental model of stent-induced thrombosis. *Thromb Haemost* 98: 662-9
115. Vilahur G, Casani L, Badimon L. 2007. A thromboxane A<sub>2</sub>/prostaglandin H<sub>2</sub> receptor antagonist (S18886) shows high antithrombotic efficacy in an experimental model of stent-induced thrombosis. *Thrombosis and Haemostasis* 98: 662-9
116. Sollier CBD, Crassard I, Simoneau G, Bergmann JF, Bousser MG, Drouet L. 2009. Effect of the Thromboxane Prostaglandin Receptor Antagonist Terutroban on Arterial Thrombogenesis after Repeated Administration in Patients Treated for the Prevention of Ischemic Stroke. *Cerebrovascular Diseases* 28: 505-13
117. Bousser MG, Amarenco P, Chamorro A, Investigators PS. 2011. Terutroban versus aspirin in patients with cerebral ischaemic events (PERFORM): a randomised, double-blind, parallel-group trial (vol 377, pg 2013, 2011). *Lancet* 378: 402-

118. Gresele P, Deckmyn H, Arnout J, Nenci GG, Vermeylen J. 1989. Characterization of N,N'-Bis(3-Picolyl)-4-Methoxy-Isophthalamide (Picotamide) as a Dual Thromboxane Synthase Inhibitor Thromboxane-A2 Receptor Antagonist in Human-Platelets. *Thrombosis and Haemostasis* 61: 479-84
119. Collier BS. 1985. A new murine monoclonal antibody reports an activation-dependent change in the conformation and/or microenvironment of the platelet glycoprotein IIb/IIIa complex. *J Clin Invest* 76: 101-8
120. Tam SH, Sassoli PM, Jordan RE, Nakada MT. 1998. Abciximab (ReoPro, chimeric 7E3 Fab) demonstrates equivalent affinity and functional blockade of glycoprotein IIb/IIIa and alpha(v)beta3 integrins. *Circulation* 98: 1085-91
121. Altieri DC, Edgington TS. 1988. A monoclonal antibody reacting with distinct adhesion molecules defines a transition in the functional state of the receptor CD11b/CD18 (Mac-1). *J Immunol* 141: 2656-60
122. Scarborough RM, Naughton MA, Teng W, Rose JW, Phillips DR, Nannizzi L, Arfsten A, Campbell AM, Charo IF. 1993. Design of Potent and Specific Integrin Antagonists - Peptide Antagonists with High Specificity for Glycoprotein-IIb-IIIa. *Journal of Biological Chemistry* 268: 1066-73
123. Hartman GD, Egbertson MS, Halczenko W, Laswell WL, Duggan ME, Smith RL, Naylor AM, Manno PD, Lynch RJ, Zhang G, et al. 1992. Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. *J Med Chem* 35: 4640-2
124. Egbertson MS, Chang CT, Duggan ME, Gould RJ, Halczenko W, Hartman GD, Laswell WL, Lynch JJ, Jr., Lynch RJ, Manno PD, et al. 1994. Non-peptide fibrinogen receptor antagonists. 2. Optimization of a tyrosine template as a mimic for Arg-Gly-Asp. *J Med Chem* 37: 2537-51
125. Berger PB. 2010. The Glycoprotein IIb/IIIa Inhibitor Wars An Update. *Journal of the American College of Cardiology* 56: 476-8
126. Berkowitz SD, Sane DC, Sigmon KN, Shavender JH, Harrington RA, Tcheng JE, Topol EJ, Califf RM, Grp EcEPICES. 1998. Occurrence and clinical significance of thrombocytopenia in a population undergoing high-risk percutaneous coronary revascularization. *Journal of the American College of Cardiology* 32: 311-9
127. Topol E, Califf R, Simoons M, Diaz R, Paolasso E, Klein W, Boland J, DeBacker G, Armstrong P, Corbalan R, Isaza D, Widimsky P, Urrutia C, Luomanmaki K, Vahanian A, Karsch K, Cokkinos D, Karatasakis G, Toutouzas P, Rodas M, Keltai M, Chierchia S, Silva E, Erikssen J, Ruzyllo W, Stepinska J, Ribeiro VD, Fernandez-Ortiz A, Macaya C, Goy J, Deckers J, Skene A, Wilcox R, Guerci A, Harrington R, Hochman J, Holmes D, Kleiman N, Kopecky S, Lee K, Lincoff A, Ohman E, Pepine C, Isea J, Investigators PT. 1998. Inhibition of platelet glycoprotein IIb/IIIa with eptifibatid in patients with acute coronary syndromes. *New England Journal of Medicine* 339: 436-43
128. Bazzino O, Barrero C, Garre L, Sosa A, Aylward P, et al. 1998. Inhibition of the platelet glycoprotein IIb/IIIa receptor with tirofiban in unstable angina and non-Q-wave myocardial infarction. *New England Journal of Medicine* 338: 1488-97
129. Leach AR, Gillet, V.J. 2003. An Introduction to Chemoinformatics. 259
130. Lyne PD. 2002. Structure-based virtual screening: an overview. *Drug Discov Today* 7: 1047-55
131. Gramatica P. 2013. On the development and validation of QSAR models. *Methods Mol Biol* 930: 499-526
132. Kuz'min VE, Artemenko AG, Muratov EN. 2008. Hierarchical QSAR technology based on the Simplex representation of molecular structure. *J Comput Aided Mol Des* 22: 403-21
133. Kuz'min VE, Artemenko AG, Polischuk PG, Muratov EN, Hromov AI, Liahovskiy AV, Andronati SA, Makan SY. 2005. Hierarchic system of QSAR models (1D-4D) on the base of simplex representation of molecular structure. *Journal of Molecular Modeling* 11: 457-67
134. Kuz'min VE, Artemenko AG, Muratov EN. 2008. Hierarchical QSAR technology based on the Simplex representation of molecular structure. *Journal of Computer-Aided Molecular Design* 22: 403-21
135. Muratov EN, Artemenko AG, Varlamova EV, Polischuk PG, Lozitsky VP, Fedchuk AS, Lozitska RL, Gridina TL, Koroleva LS, Sil'nikov VN, Galabov AS, Makarov VA, Riabova OB, Wutzler P,

- Schmidtke M, Kuz'min VE. 2010. Per aspera ad astra: application of Simplex QSAR approach in antiviral research. *Future Medicinal Chemistry* 2: 1205-26
136. Laboratoire d'Infochimie UMR 7177 CNRS, Université de Strasbourg, 4, rue B. Pascal, Strasbourg 67000, France. . <http://infochim.u-strasbg.fr/>.
137. Bonachera F, Parent B, Barbosa F, Froloff N, Horvath D. 2006. Fuzzy tricentric pharmacophore fingerprints. 1. Topological fuzzy pharmacophore triplets and adapted molecular similarity scoring schemes. *Journal of Chemical Information and Modeling* 46: 2457-77
138. Bonachera F, Horvath D. 2008. Fuzzy tricentric pharmacophore fingerprints. 2. application of topological fuzzy pharmacophore triplets in quantitative structure-activity relationships. *Journal of Chemical Information and Modeling* 48: 409-25
139. Kuz'min VE, Artemenko AG, Polischuk PG, Muratov EN, Khromov AI, Liahovskiy AV, Andronati SA, Makan SY. 2005. Hierarchic System of QSAR Models (1D-4D) on the Base of Simplex Representation of Molecular Structure *Journal of Molecular Modelling* 11: 457-67
140. Solov'ev VP, Varnek A, Wipff G. 2000. Modeling of ion complexation and extraction using substructural molecular fragments. *J Chem Inf Comput Sci* 40: 847-58
141. Solov'ev V, Marcou G, Tsivadze A, Varnek A. 2012. Complexation of Mn<sup>2+</sup>, Fe<sup>2+</sup>, Y<sup>3+</sup>, La<sup>3+</sup>, Pb<sup>2+</sup>, and UO<sub>2</sub><sup>2+</sup> with Organic Ligands: QSPR Ensemble Modeling of Stability Constants. *Industrial & Engineering Chemistry Research* 51: 13482-9
142. Swamy MNST, K. 1981. Graphs, Networks, and Algorithms. *John Wiley & Sons: New York*
143. Tropsha A, Gramatica P, Gombar VK. 2003. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *Qsar & Combinatorial Science* 22: 69-77
144. Breiman L. 2001. Random forests. *Machine Learning* 45: 5-32
145. Polishchuk PG. 2010. CF. pp. A.V. Bogatsky Physico-Chemical Institute of NAS of Ukraine, Odessa, Ukraine
146. Polishchuk PG, Muratov EN, Artemenko AG, Kolumbin OG, Muratov NN, Kuz'min VE. 2009. Application of Random Forest Approach to QSAR Prediction of Aquatic Toxicity. *Journal of Chemical Information and Modeling* 49: 2481-8
147. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. 2003. Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences* 43: 1947-58
148. Kovdienko NA, Polishchuk PG, Muratov EN, Artemenko AG, Kuz'min VE, Gorb L, Hill F, Leszczynski J. 2010. Application of Random Forest and Multiple Linear Regression Techniques to QSPR Prediction of an Aqueous Solubility for Military Compounds. *Molecular Informatics* 29: 394-406
149. Ognichenko LN, Kuz'min VE, Gorb L, Hill FC, Artemenko AG, Polischuk PG, Leszczynski J. 2012. QSPR Prediction of Lipophilicity for Organic Compounds Using Random Forest Technique on the Basis of Simplex Representation of Molecular Structure. *Molecular Informatics* 31: 273-80
150. L. Breiman JHF, R. A. Olshen, C. J. Stone. 1984. Classification and Regression Trees. *Wadsworth, Belmont, CA*: 368
151. Kruskal JB, Jr. 1956. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *Proceedings of the American Mathematical Society* 7: 48-50
152. Sushko I, Novotarskyi S, Korner R, Pandey AK, Cherkasov A, Lo JZ, Gramatica P, Hansen K, Schroeter T, Muller KR, Xi LL, Liu HX, Yao XJ, Oberg T, Hormozdiari F, Dao PH, Sahinalp C, Todeschini R, Polishchuk P, Artemenko A, Kuz'min V, Martin TM, Young DM, Fourches D, Muratov E, Tropsha A, Baskin I, Horvath D, Marcou G, Muller C, Varnek A, Prokopenko VV, Tetko IV. 2010. Applicability Domains for Classification Problems: Benchmarking of Distance to Models for Ames Mutagenicity Set. *Journal of Chemical Information and Modeling* 50: 2094-111
153. Fawcett T. 2006. An introduction to ROC analysis. *Pattern Recognition Letters* 27: 861-74
154. Tetko IV, Solov'ev VP, Antonov AV, Yao XJ, Doucet JP, Fan BT, Hoonakker F, Fourches D, Jost P, Lachiche N, Varnek A. 2006. Benchmarking of linear and nonlinear approaches for quantitative structure-property relationship studies of metal complexation with ionophores. *Journal of Chemical Information and Modeling* 46: 808-19

155. Varnek A, Kireeva N, Tetko IV, Baskin II, Solov'ev VP. 2007. Exhaustive QSPR studies of a large diverse set of ionic liquids: How accurately can we predict melting points? *Journal of Chemical Information and Modeling* 47: 1111-22
156. Hansen K, Mika S, Schroeter T, Sutter A, ter Laak A, Steger-Hartmann T, Heinrich N, Muller KR. 2009. Benchmark Data Set for in Silico Prediction of Ames Mutagenicity. *Journal of Chemical Information and Modeling* 49: 2077-81
157. <http://infochim.u-strasbg.fr/cgi-bin/predictor.cgi?GeneralPropSelect=PhysProp>.
158. Muratov EN, Kuz'min VE, Artemenko AG, Kovdienko NA, Gorb L, Hill F, Leszczynski J. 2010. New QSPR equations for prediction of aqueous solubility for military compounds. *Chemosphere* 79: 887-90
159. Poroikov V, Filimonov D, Lagunin A, Glorizova T, Zakharov A. 2007. PASS: identification of probable targets and mechanisms of toxicity. *SAR QSAR Environ Res* 18: 101-10
160. Filimonov D.A. PVV. 2006. Prediction of biological activity spectra for organic compounds. *Russian Chemical Journal* 50: 66-75
161. Grant JA, Gallardo MA, Pickup BT. 1996. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *Journal of Computational Chemistry* 17: 1653-66
162. Zauhar RJ, Moyna G, Tian LF, Li ZJ, Welsh WJ. 2003. Shape signatures: A new approach to computer-aided ligand- and receptor-based drug design. *Journal of Medicinal Chemistry* 46: 5674-90
163. Ballester PJ, Finn PW, Richards WG. 2009. Ultrafast shape recognition: Evaluating a new ligand-based virtual screening technology. *Journal of Molecular Graphics & Modelling* 27: 836-45
164. Ballester PJ, Richards WG. 2007. Ultrafast shape recognition to search compound databases for similar molecular shapes. *Journal of Computational Chemistry* 28: 1711-23
165. Ebalunode JO, Zheng WF. 2010. Molecular Shape Technologies in Drug Discovery: Methods and Applications. *Current Topics in Medicinal Chemistry* 10: 669-79
166. ROCS, 3.1.2; OpenEye Scientific Software, Inc.: Santa Fe, NM, USA, 2010, [www.eyesopen.com](http://www.eyesopen.com).
167. Rush TS, Grant JA, Mosyak L, Nicholls A. 2005. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *Journal of Medicinal Chemistry* 48: 1489-95
168. Mills JEJ, Dean PM. 1996. Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *Journal of Computer-Aided Molecular Design* 10: 607-22
169. Hawkins PCD, Skillman AG, Nicholls A. 2007. Comparison of shape-matching and docking as virtual screening tools. *Journal of Medicinal Chemistry* 50: 74-82
170. Cheeseright T, Mackey M, Rose S, Vinter A. 2007. Molecular field technology applied to virtual screening and finding the bioactive conformation. *Expert Opinion on Drug Discovery* 2: 131-44
171. Cruciani G. 2005. Molecular Interaction Fields. Applications in Drug Discovery and ADME Prediction. 328
172. Klebe G, Abraham U, Mietzner T. 1994. Molecular Similarity Indexes in a Comparative-Analysis (Comsia) of Drug Molecules to Correlate and Predict Their Biological-Activity. *Journal of Medicinal Chemistry* 37: 4130-46
173. Vinter JG, Saunders MR. 1991. Molecular Modeling Approaches to Host-Guest Complexes. *Ciba Foundation Symposia* 158: 249-65
174. Apaya RP, Lucchese B, Price SL, Vinter JG. 1995. The Matching of Electrostatic Extrema - a Useful Method in Drug Design - a Study of Phosphodiesterase-III Inhibitors. *Journal of Computer-Aided Molecular Design* 9: 33-43
175. Vinter JG, Trollope KI. 1995. Multiconformational Composite Molecular-Potential Fields in the Analysis of Drug-Action .1. Methodology and First Evaluation Using 5-Ht and Histamine Action as Examples. *Journal of Computer-Aided Molecular Design* 9: 297-307
176. S. Rose AV. 2007. Molecular Field Technology and its Applications in Drug Discovery. *Innovations in pharmaceutical technology*

177. Cheeseright T, Mackey M, Rose S, Vinter A. 2006. Molecular field extrema as descriptors of biological activity: Definition and validation. *Journal of Chemical Information and Modeling* 46: 665-76
178. Wermuth G, Ganellin CR, Lindberg P, Mitscher LA. 1998. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure and Applied Chemistry* 70: 1129-43
179. Langer T, Wolber G. 2004. Pharmacophore definition and 3D searches. *Drug Discovery Today: Technologies* 1: 203-7
180. Wolber G, Seidel T, Bendix F, Langer T. 2008. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discovery Today* 13: 23-9
181. Leach AR, Gillet VJ, Lewis RA, Taylor R. 2010. Three-Dimensional Pharmacophore Methods in Drug Discovery. *Journal of Medicinal Chemistry* 53: 539-58
182. Sottriffer C. 2011. Principles, Challenges, and Practical Guidelines. *Virtual Screening* 48: 550
183. Langer TaH, R.D. 2006. Pharmacophores and pharmacophore searches. *Methods and principles in medicinal chemistry* 32
184. Cramer RD, 3rd, Patterson DE, Bunce JD. 1989. Recent advances in comparative molecular field analysis (CoMFA). *Prog Clin Biol Res* 291: 161-5
185. Wolber G, Dornhofer AA, Langer T. 2006. Efficient overlay of small organic molecules using 3D pharmacophores. *Journal of Computer-Aided Molecular Design* 20: 773-88
186. Wolber G, Langer T. 2005. LigandScout: 3-d pharmacophores derived from protein-bound Ligands and their use as virtual screening filters. *Journal of Chemical Information and Modeling* 45: 160-9
187. Wolber G, Seidel, T., Bendix, F., Ibis, G., Dornhofer, A.A., Biely, M., Adaktylos, P., and Kosara, R. 2010. LigandScout, Inte:Ligand GmbH, Vienna, Austria.
188. Halgren TA. 1996. Merck molecular force field .1. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry* 17: 490-519
189. Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, Stahl MT. 2010. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling* 50: 572-84
190. Hawkins PC, Nicholls A. Conformer Generation with OMEGA: Learning from the Data Set and the Analysis of Failures. *J Chem Inf Model*
191. Huang SY, Zou XQ. 2010. Advances and Challenges in Protein-Ligand Docking. *International Journal of Molecular Sciences* 11: 3016-34
192. Brooijmans N, Kuntz ID. 2003. Molecular recognition and docking algorithms. *Annual Review of Biophysics and Biomolecular Structure* 32: 335-73
193. Kitchen DB, Decornez H, Furr JR, Bajorath J. 2004. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews Drug Discovery* 3: 935-49
194. Carlson HA. 2002. Protein flexibility is an important component of structure-based drug discovery. *Current Pharmaceutical Design* 8: 1571-8
195. Carlson HA, McCammon JA. 2000. Accommodating protein flexibility in computational drug design. *Molecular Pharmacology* 57: 213-8
196. Teodoro ML, Kavraki LE. 2003. Conformational flexibility models for the receptor in structure based drug design. *Current Pharmaceutical Design* 9: 1635-48
197. Teague SJ. 2003. Implications of protein flexibility for drug discovery. *Nature Reviews Drug Discovery* 2: 527-41
198. Cozzini P, Kellogg GE, Spyrakis F, Abraham DJ, Costantino G, Emerson A, Fanelli F, Gohlke H, Kuhn LA, Morris GM, Orozco M, Pertinhez TA, Rizzi M, Sottriffer CA. 2008. Target Flexibility: An Emerging Consideration in Drug Discovery and Design. *Journal of Medicinal Chemistry* 51: 6237-55
199. Totrov M, Abagyan R. 2008. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Current Opinion in Structural Biology* 18: 178-84
200. Rognan D. 2006. Ligand Design for G Protein-Coupled Receptors. *Methods and principles in medicinal chemistry ; v. 30.*

201. Sousa SF, Fernandes PA, Ramos MJ. 2006. Protein-ligand docking: current status and future challenges. *Proteins* 65: 15-26
202. Korb O, Stutzle T, Exner TE. 2006. PLANTS: Application of ant colony optimization to structure-based drug design. *Ant Colony Optimization and Swarm Intelligence, Proceedings* 4150: 247-58
203. Korb O, Stutzle T, Exner TE. 2009. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *Journal of Chemical Information and Modeling* 49: 84-96
204. Kramer B, Rarey M, Lengauer T. 1999. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins-Structure Function and Genetics* 37: 228-41
205. Rarey M, Kramer B, Lengauer T. 1997. Multiple automatic base selection: Protein-ligand docking based on incremental construction without manual intervention. *Journal of Computer-Aided Molecular Design* 11: 369-84
206. Bohm HJ. 1994. The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein Ligand Complex of Known 3-Dimensional Structure. *Journal of Computer-Aided Molecular Design* 8: 243-56
207. Vilar S, Cozza G, Moro S. 2008. Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery. *Current Topics in Medicinal Chemistry* 8: 1555-72
208. Concato J, Feinstein AR. 1997. Monte Carlo methods in clinical research: applications in multivariable analysis. *J Investig Med* 45: 394-400
209. 2011. Molecular Operating Environment (MOE). *Chemical Computing Group Inc.* 2011.10
210. Harada T, Katada J, Tachiki A, Asari T, Iijima K, Uno I, Ojima I, Hayashi Y. 1997. Development of the new potent non-peptide gpIIb/IIIa antagonist NSL-95301 by utilizing combinatorial technique. *Bioorganic & Medicinal Chemistry Letters* 7: 209-12
211. Cheng S, Craig WS, Mullen D, Tschopp JF, Dixon D, Pierschbacher MD. 1994. Design and Synthesis of Novel Cyclic Rgd-Containing Peptides as Highly Potent and Selective Integrin Alpha(lib)Beta(3) Antagonists. *Journal of Medicinal Chemistry* 37: 1-8
212. Ojima I, Dong Q, Chakravarty S, Peerschke E, Hwang SM, Wong AS. 1995. Design, Synthesis and Sar of Rgd Peptide Hybrids as Highly Efficient Inhibitors of Platelet-Aggregation. *Bioorganic & Medicinal Chemistry Letters* 5: 1941-6
213. Stiliz HU, Guba W, Jablonka B, Just M, Klingler O, Konig W, Wehner V, Zoller G. 2001. Discovery of an orally active non-peptide fibrinogen receptor antagonist based on the hydantoin scaffold. *Journal of Medicinal Chemistry* 44: 1158-76
214. Zablocki JA, Rico JG, Garland RB, Rogers TE, Williams K, Schretzman LA, Rao SA, Bovy PR, Tjoeng FS, Lindmark RJ, Toth MV, Zupec ME, Mcmackins DE, Adams SP, Miyano M, Markos CS, Milton MN, Paulson S, Herin M, Jacqmin P, Nicholson NS, Panzerknodle SG, Haas NF, Page JD, Szalony JA, Taite BB, Salyers AK, King LW, Campion JG, Feigen LP. 1995. Potent in-Vitro and in-Vivo Inhibitors of Platelet-Aggregation Based Upon the Arg-Gly-Asp Sequence of Fibrinogen - (Aminobenzamidino)Succinyl (Abas) Series of Orally-Active Fibrinogen Receptor Antagonists. *Journal of Medicinal Chemistry* 38: 2378-94
215. Okumura K, Shimazaki T, Aoki Y, Yamashita H, Tanaka E, Banba S, Yazawa K, Kibayashi K, Banno H. 1998. New platelet fibrinogen receptor glycoprotein IIb-IIIa antagonists: Orally active series of N-alkylated amidines with a 6,6-bicyclic template. *Journal of Medicinal Chemistry* 41: 4036-52
216. Lemmen C, Lengauer T, Klebe G. 1998. FLEXX: A method for fast flexible ligand superposition. *Journal of Medicinal Chemistry* 41: 4502-20
217. Dixon SL, Smondyrev AM, Rao SN. 2006. PHASE: A novel approach to pharmacophore modeling and 3D database searching. *Chemical Biology & Drug Design* 67: 370-2
218. Miyashita M, Akamatsu M, Hayashi Y, Ueno T. 2000. Three-dimensional quantitative structure-activity relationship analyses of RGD mimetics as fibrinogen receptor antagonists. *Bioorganic & Medicinal Chemistry Letters* 10: 859-63
219. Yan YL, Li Y, Zhang SW, Ai CZ. 2011. Studies of tricyclic piperazine/piperidine furnished molecules as novel integrin alpha(v)beta(3)/alpha(IIb)beta(3) dual antagonists using 3D-QSAR and molecular docking. *Journal of Molecular Graphics & Modelling* 29: 747-62

220. Yang J, Zhan CY, Dong XC, Yang K, Wang FX. 2004. Interaction of human fibrinogen receptor (GPIIb-IIIa) with decorsin. *Acta Pharmacologica Sinica* 25: 1096-104
221. Feuston BP, Culberson JC, Hartman GD. 2003. Molecular model of the alpha(IIb)beta(3) integrin. *Journal of Medicinal Chemistry* 46: 5316-25
222. Yang J, Yao J, Chen J, Wang XN, Zhu TY, Chen LL, Chu P. 2009. Construction of drug screening cell model and application to new compounds inhibiting FITC-fibrinogen binding to CHO cells expressing human alpha IIb beta 3. *European Journal of Pharmacology* 618: 1-8
223. Pfaff M, Tangemann K, Muller B, Gurrath M, Muller G, Kessler H, Timpl R, Engel J. 1994. Selective Recognition of Cyclic Rgd Peptides of Nmr Defined Conformation by Alpha-II-Beta-3, Alpha-V-Beta-3, and Alpha-5-Beta-1 Integrins. *Journal of Biological Chemistry* 269: 20233-8
224. Xiong JP, Stehle T, Zhang RG, Joachimiak A, Frech M, Goodman SL, Arnaout MA. 2002. Crystal structure of the extracellular segment of integrin alpha V beta 3 in complex with an Arg-Gly-Asp ligand. *Science* 296: 151-5
225. Loftus JC, Otoole TE, Plow EF, Glass A, Frelinger AL, Ginsberg MH. 1990. A Beta-3 Integrin Mutation Abolishes Ligand-Binding and Alters Divalent-Cation Dependent Conformation. *Science* 249: 915-8
226. Springer TA, Zhu JH, Xiao T. 2008. Structural basis for distinctive recognition of fibrinogen gamma C peptide by the platelet integrin alpha(IIb)beta(3). *Journal of Cell Biology* 182: 791-800
227. Chatterjee J, Ovadia O, Zahn G, Marinelli L, Hoffman A, Gilon C, Kessler H. 2007. Multiple N-methylation by a designed approach enhances receptor selectivity. *J Med Chem* 50: 5878-81
228. Bollinger M, Manzenrieder F, Kolb R, Bochen A, Neubauer S, Marinelli L, Limongelli V, Novellino E, Moessmer G, Pell R, Lindner W, Fanous J, Hoffman A, Kessler H. 2012. Tailoring of integrin ligands: probing the charge capability of the metal ion-dependent adhesion site. *J Med Chem* 55: 871-82
229. Born GVR. 1962. Aggregation of Blood Platelets by Adenosine Diphosphate and its Reversal. *Nature* 194: 927-9
230. Xia Z, Wong T, Liu Q, KasirerFriede A, Brown E, Frojmovic MM. 1996. Optimally functional fluorescein isothiocyanate-labelled fibrinogen for quantitative studies of binding to activated platelets and platelet aggregation. *British Journal of Haematology* 93: 204-14
231. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP. 2012. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research* 40: D1100-D7
232. Mehrotra MM, Heath JA, Smyth MS, Pandey A, Rose JW, Seroogy JM, Volkots DL, Nannizzi-Alaimo L, Park GL, Lambing JL, Hollenbach SJ, Scarborough RM. 2004. Discovery of novel 2,8-diazaspiro[4.5]decanes as orally active glycoprotein IIb-IIIa antagonists. *Journal of Medicinal Chemistry* 47: 2037-61
233. Standartizer, 5.4; Chemaxon: Budapest, Hungary, [www.chemaxon.com/products.html](http://www.chemaxon.com/products.html).
234. *Instant JChem*, 5.4; Chemaxon: Budapest, Hungary, [www.chemaxon.com/products.html](http://www.chemaxon.com/products.html).
235. 2010. OEChem. Santa Fe, NM, USA: OpenEye Scientific Software, Inc.
236. Krysko AA, Samoylenko GV, Polishchuk PG, Andronati SA, Kabanova TA, Khristova TM, Kuz'min VE, Kabanov VM, Krysko OL, Varnek AA, Grygorash RY. 2011. RGD mimetics containing phthalimidine fragment as novel ligands of fibrinogen receptor. *Bioorg Med Chem Lett* 21: 5971-4
237. Krysko AA, Samoylenko GV, Polishchuk PG, Fonari MS, Kravtsov V, Andronati SA, Kabanova TA, Lipkowski J, Khristova TM, Kuz'min VE, Kabanov VM, Krysko OL, Varnek AA. Synthesis, biological evaluation, X-ray molecular structure and molecular docking studies of RGD mimetics containing 6-amino-2,3-dihydroisoindolin-1-one fragment as ligands of integrin alphaIIb beta3. *Bioorg Med Chem* 21: 4646-61
238. Rognan D. 2005. BioinfoDB: un inventaire de molécules commercialement disponibles à des fins de criblage biologique. *La Gazette du CINES* 1-4
239. Bosch X, Marrugat J, Sanchis J. 2010. Platelet glycoprotein IIb/IIIa blockers during percutaneous coronary intervention and as the initial medical treatment of non-ST segment elevation acute coronary syndromes. *Cochrane Database Syst Rev*: CD002130

240. Bougie DW, Wilker PR, Wuitschick ED, Curtis BR, Malik M, Levine S, Lind RN, Pereira J, Aster RH. 2002. Acute thrombocytopenia after treatment with tirofiban or eptifibatid is associated with antibodies specific for ligand-occupied GPIIb/IIIa. *Blood* 100: 2071-6
241. Scirica BM, Cannon CP, Cooper R, Aster RH, Brassard J, McCabe CH, Charlesworth A, Skene AM, Braunwald E. 2006. Drug-induced thrombocytopenia and thrombosis: Evidence from patients receiving an oral glycoprotein IIb/IIIa inhibitor in the Orbofiban in Patients with Unstable Coronary Syndromes-(OPUS-TIMI 16) trial. *Journal of Thrombosis and Thrombolysis* 22: 95-102
242. Bassler N, Loeffler C, Mangin P, Yuan YP, Schwarz M, Hagemeyer CE, Eisenhardt SU, Ahrens I, Bode C, Jackson SP, Peter K. 2007. A mechanistic model for paradoxical platelet activation by ligand-mimetic alpha(IIb)beta(3) (GPIIb/IIIa) antagonists. *Arteriosclerosis Thrombosis and Vascular Biology* 27: E9-E15
243. Jennings LK, Haga JH, Slack SM. 2000. Differential expression of a ligand induced binding site (LIBS) by GPIIb-IIIa ligand recognition peptides and parenteral antagonists. *Thrombosis and Haemostasis* 84: 1095-102
244. Negri A, Li J, Naini S, Coller BS, Filizola M. 2012. Structure-based virtual screening of small-molecule antagonists of platelet integrin alphaIIb beta3 that do not prime the receptor to bind ligand. *J Comput Aided Mol Des* 26: 1005-15
245. Hays SJ, Caprathe BW, Gilmore JL, Amin N, Emmerling MR, Michael W, Nadimpalli R, Nath R, Raser KJ, Stafford D, Watson D, Wang K, Jaen JC. 1998. 2-amino-4H-3,1-benzoxazin-4-ones as inhibitors of C1r serine protease. *J Med Chem* 41: 1060-7
246. Baek NN, Jang HR, Huh W, Kim YG, Kim DJ, Oh HY, Lee JE. 2012. The role of nafamostat mesylate in continuous renal replacement therapy among patients at high risk of bleeding. *Ren Fail* 34: 279-85
247. Fuse I, Higuchi W, Toba K, Aizawa Y. 1999. Inhibitory mechanism of human platelet aggregation by nafamostat mesilate. *Platelets* 10: 212-8
248. Hodohara K FY, Hosoda S, Yasunaga K. 1989. Inhibition of fibrinogen binding to ADP-stimulated human platelets by synthetic serine protease inhibitors. *Blood Vessels* 20: 213-9
249. Aoyama T, Ino Y, Ozeki M, Oda M, Sato T, Koshiyama Y, Suzuki S, Fujita M. 1984. Pharmacological Studies of Fut-175, Nafamostat Mesilate .1. Inhibition of Protease Activity in Invitro and Invivo Experiments. *Japanese Journal of Pharmacology* 35: 203-27
250. Paques EP, Romisch J. 1991. Comparative study on the in vitro effectiveness of antithrombotic agents. *Thromb Res* 64: 11-21
251. Tsuda Y, Nakahara T, Ueda K, Mori A, Sakamoto K, Ishii K. 2012. Effect of nafamostat on N-methyl-D-aspartate-induced retinal neuronal and capillary degeneration in rats. *Biol Pharm Bull* 35: 2209-13
252. Inman RD, Chiu B. 2012. Nafamostat mesylate, a serine protease inhibitor, demonstrates novel antimicrobial properties and effectiveness in Chlamydia-induced arthritis. *Arthritis Res Ther* 14: R150
253. Gocho T, Uwagawa T, Furukawa K, Haruki K, Fujiwara Y, Iwase R, Misawa T, Ohashi T, Yanaga K. 2013. Combination chemotherapy of serine protease inhibitor nafamostat mesilate with oxaliplatin targeting NF-kappaB activation for pancreatic cancer. *Cancer Lett*
254. Ohtake Y, Hirasawa H, Sugai T, Oda S, Shiga H, Matsuda K, Kitamura N. 1991. Nafamostat mesylate as anticoagulant in continuous hemofiltration and continuous hemodiafiltration. *Contrib Nephrol* 93: 215-7
255. Nakae H, Tajimi K. 2003. Pharmacokinetics of nafamostat mesilate during continuous hemodiafiltration with a polyacrylonitrile membrane. *Ther Apher Dial* 7: 483-5
256. Khasawneh FT, Huang JS, Turek JW, Le Breton GC. 2006. Differential mapping of the amino acids mediating agonist and antagonist coordination with the human thromboxane A2 receptor protein. *J Biol Chem* 281: 26951-65
257. Farmer PS. 1980. Bridging the gap between bioactive peptides and nonpeptides: some perspectives in design. *Drug Des.* 10: 119-43



258. Jin B, Hopfinger AJ. 1994. A Proposed Common Spatial Pharmacophore and the Corresponding Active Conformations of Some Txa(2) Receptor Antagonists. *Abstracts of Papers of the American Chemical Society* 208: 210-MEDI
259. Wei J, Liu YX, Wang SQ. 2009. 3D pharmacophore models for thromboxane A(2) receptor antagonists. *Journal of Molecular Modeling* 15: 1185-91
260. Ye Y, Liao Q, Wei J, Gao Q. 3D-QSAR study of corticotropin-releasing factor 1 antagonists and pharmacophore-based drug design. *Neurochem Int* 56: 107-17
261. Catalyst Tutorial. Available: [www.accelrys.com](http://www.accelrys.com). Accessed 2011 Aug 18.
262. Kawashima Y, Sato M, Yamamoto S, Shimazaki Y, Chiba Y, Satake M, Iwata C, Hatayama K. 1995. Structure-Activity Relationship Study of Txa(2) Receptor Antagonists - 4-[2-(4-Substituted Phenylsulfonylamino)Ethylthio]Phenoxyacetic Acids and Related-Compounds. *Chemical & Pharmaceutical Bulletin* 43: 1132-6
263. Kontogiorgis DHLaCA. 2002. Quantitative-Structure Activity Relationships on Thromboxane Receptor Antagonists. *Internet Electronic Journal of Molecular Design* Volume 1: 300-9
264. Kurup A. 2003. C-QSAR: a database of 18,000 QSARs and associated biological and physical data. *Journal of Computer-Aided Molecular Design* 17: 187-96
265. Kontogiorgis C, Hadjipavlou-Litina D. 2010. Thromboxane Synthase Inhibitors and Thromboxane A(2) Receptor Antagonists: A Quantitative Structure Activity Relationships (QSARs) Analysis. *Current Medicinal Chemistry* 17: 3162-214
266. Sairam KV, Sarma JA, Desiraju GR. 2003. 3D-QSAR studies of some [[1-aryl(or benzyl)-1-(benzenesulphonamido)methyl] phenyl] alkanolic acid derivatives as thromboxane A2 receptor antagonists. *Drug Des Discov* 18: 47-51
267. Hahn M, Rogers D. 1995. Receptor Surface Models .2. Application to Quantitative Structure-Activity-Relationships Studies. *Journal of Medicinal Chemistry* 38: 2091-102
268. Yamamoto Y, Kamiya K, Terao S. 1993. Modeling of Human Thromboxane-a(2) Receptor and Analysis of the Receptor Ligand Interaction. *Journal of Medicinal Chemistry* 36: 820-5
269. Strader CD, Sigal IS, Dixon RA. 1989. Structural basis of beta-adrenergic receptor function. *FASEB J* 3: 1825-32
270. Wouters J, Durant F, Masereel B. 1999. Antagonism of the TXA2 receptor by seratrodist: A structural approach. *Bioorganic & Medicinal Chemistry Letters* 9: 2867-70
271. Shiraishi M, Kato K, Terao S, Ashida Y, Terashita Z, Kito G. 1989. Quinones .4. Novel Eicosanoid Antagonists - Synthesis and Pharmacological Evaluation. *Journal of Medicinal Chemistry* 32: 2214-21
272. Ruan KH, Wu JX, So SP, Jenkins LA, Ruan CH. 2004. NMR structure of the thromboxane A(2) receptor ligand recognition pocket. *European Journal of Biochemistry* 271: 3006-16
273. Ruan KH, So SP, Wu JX, Li D, Huang A, Kung J. 2001. Solution structure of the second extracellular loop of human thromboxane A(2) receptor. *Faseb Journal* 15: A202-A
274. Wu HX, So SP, Ruan KH. 2003. Solution structure of the third extracellular loop of human thromboxane A(2) receptor. *Archives of Biochemistry and Biophysics* 414: 287-93
275. So SP, Wu JX, Huang GX, Huang AM, Li DW, Ruan KH. 2003. Identification of residues important for ligand binding of thromboxane A(2) receptor in the second extracellular loop using the NMR experiment-guided mutagenesis approach. *Journal of Biological Chemistry* 278: 10922-7

## **APPENDIX 1.**

### **Biological experiments for antagonists of integrin $\alpha_{IIb}\beta_3$**

## 1. Biological experiments for antagonists of integrin $\alpha_{IIb}\beta_3$

In this section, we describe experimental techniques used in PCI to investigate anti-aggregation activity and affinity for integrin  $\alpha_{IIb}\beta_3$  of the compounds theoretically designed in this work.

The anti-aggregation activity can be determined by platelet aggregometry test. It represents one of the most widely utilized test procedures for assessing platelet function. Platelet aggregometry can represent a diagnostic test procedure and primary hemostasis. There are two main types of instruments available [281]: (i) light transmission aggregometry (LTA), which represents the original methodology developed in the works of Born and O'Brien [234, 282], and (ii) whole blood aggregometry (WBA).

LTA measures light transmission through a test sample containing platelets in suspension that increases when platelets are aggregated by an agonist [283]. LTA is performed using a light transmission aggregometer, comprised of a light source, cuvette holder, a heater to provide a constant (37°C) sample temperature, and a photoelectric cell to measure the light beam after its passage through the platelet suspension [281]. LTA is a time consuming and technically challenging technique that is affected by many pre-analytical and analytical variables that must be carefully controlled. For example, the preparation of the platelet-rich-plasma (PRP) requires a centrifugation step and this not only takes time, but can also lead to test artifacts, as highlighted later. Hence, LTA should be performed only in highly specialized laboratories by experienced personnel [281]. The mostly used LTA test is Born's method [234] (described below), which was also used in PCI.

**In vitro inhibition of platelet aggregation in human platelet-rich plasma.** PRP was prepared by centrifugation of whole blood at 150 g for 10 min, and the platelet count was adjusted to  $1 \cdot 10^8$  platelets/mL with time matched platelet-poor plasma (PPP). PRP (250  $\mu$ l) was preincubated with 50  $\mu$ l of various concentrations of compounds to be tested, or saline, for 2 min at 37 °C prior to the addition of ADP (10  $\mu$ l). Platelet aggregation was measured by the change of light transmittance (PPP represents 100 %) under stirring conditions (1000 rpm) on a "THROMLITE-1006 A" aggregometer. The ability of the compounds to inhibit platelet aggregation was measured and the  $IC_{50}$  was determined as the concentration of the compound required to produce 50 % inhibition of the response to ADP [234].

Born's method had several issues which can cause poor reproducibility of results in different laboratories: different platelet counts in samples, which usually varied in range 200-300  $\times 10^9$ /L; changing of pH during storage of platelet rich plasma due to loss of CO<sub>2</sub> from the

sample; low stirring speed of the sample during analysis leads to decrease in aggregation slope curves, whereas too high speed can lead to damage of platelets; starting time and duration of the experiment is an important factor too, because platelet responses to agonists change in time [281].

The tests for affinity are necessary for understanding the mechanism of binding of fibrinogen to  $\alpha_{IIb}\beta_3$ . Until recently radioactive labeling of fibrinogen extensively has been used to detect the expression of  $\alpha_{IIb}\beta_3$  receptors[284], the affinity[285-286] and kinetics[287] of fibrinogen binding. However, radioactive fibrinogen studies can neither detect platelet subpopulation responses[288], nor dynamically monitor binding kinetics.

An alternative approach uses flow cytometry with fluorescein isothiocyanate (FITC), which conjugates antibodies specific to various domains of immobilized and/or soluble fibrinogen, such as 9F9[289], a polyclonal rabbit anti-human Fg[290] and anti-Fg-RIBS[291-292]. Using FITC-labeled Fg instead of radioactively labeled Fg saves complicated sample handling, and helps to avoid the problems of short storage period. With the help of flow cytometry, FITC-labeling of Fg also enables dynamic studies of Fg binding on platelets for determining  $K_D$  values, which is not feasible with the anti-FgmAb 9F9, as well as for determining kinetics of binding[293]. Addition of FITC-Fg to PRP enables direct studies of Fg binding to platelets in native plasma, without the complications of platelet isolation, and can readily be extended to studies of diluted whole blood as previously reported for studies with labeled antibodies [289-290]. This approach was used in PCI to measure the affinity for  $\alpha_{IIb}\beta_3$  receptor.

#### **In vitro inhibition of FITC-labeled fibrinogen binding to activated human platelets.**

Fresh platelet concentrate was centrifuged (900g, 15 min), and the platelet pellet was washed at pH 5.6 in the presence of PGE<sub>1</sub> and finally resuspended carefully in Tyrode's buffer, containing BSA, pH 7.4. Platelet suspension was incubated for 30 min with peptidomimetics at concentrations in the range from 0.001 nM to 1  $\mu$ M at 37 °C. FITC-Fg was added after addition of agonist (ADP, 2  $\mu$ M). Following 60 min incubation at room temperature (under protection with the light), the reaction mixtures were layered onto a 20 % sucrose cushion and centrifuged at 6000 rpm for 15 min. The tips of each tube were clipped off, and the platelet pellet, containing the bound fibrinogen, solubilized with a 3 % SDS solution. The bound FITC-Fg was calculated spectrofluorimetrically [235].

As well as Born's method this approach also have some drawbacks. Thus a high interindividual variability of  $\alpha_{IIb}\beta_3$  population levels on platelets, which can vary in range 10000-44000 fibrinogen binding sites per platelet [235, 294], can lead to significant difference in dissociation constants of fibrinogen ( $K_D = 70$ -255 nM) [295] and differences in affinity values of

antagonists of fibrinogen receptors (for example, clinically relevant concentration of tirofiban caused inhibition of fibrinogen binding ranging from 17% to 88% [296]).

## Reference

1. Favalaro EJ, Lippi G, Franchini M. 2010. Contemporary platelet function testing. *Clinical Chemistry and Laboratory Medicine* 48: 579-98
2. O'Brien J. 1962. Platelet aggregation: Part II Some results from a new method of study. *J Clin Pathol.* 15(5): 452-5
3. Born GVR. 1962. Aggregation of Blood Platelets by Adenosine Diphosphate and its Reversal. *Nature* 194: 927-9
4. Cattaneo M. 2009. Light Transmission Aggregometry and ATP Release for the Diagnostic Assessment of Platelet Function. *Seminars in Thrombosis and Hemostasis* 35: 158-67
5. Abrams CS, Ellison N, Budzynski AZ, Shattil SJ. 1990. Direct Detection of Activated Platelets and Platelet-Derived Microparticles in Humans. *Blood* 75: 128-38
6. Klopogge E, Akkerman JWN. 1986. Platelet-Activating-Factor (Paf-Acether) Induces High-Affinity and Low-Affinity Binding of Fibrinogen to Human-Platelets Via Independent Mechanisms. *Biochemical Journal* 240: 403-12
7. Peerschke EI, Zucker MB, Grant RA, Egan JJ, Johnson MM. 1980. Correlation between Fibrinogen Binding to Human-Platelets and Platelet Aggregability. *Blood* 55: 841-7
8. Klopogge E, Mommersteeg M, Akkerman JWN. 1986. Kinetics of Platelet-Activating-Factor 1-O-Alkyl-2-Acetyl-Sn-Glycero-3-Phosphocholine-Induced Fibrinogen Binding to Human-Platelets. *Journal of Biological Chemistry* 261: 1071-6
9. Frojmovic MM, Mooney RF, Wong T. 1994. Dynamics of Platelet Glycoprotein lib-IIa Receptor Expression and Fibrinogen Binding .1. Quantal Activation of Platelet Subpopulations Varies with Adenosine-Diphosphate Concentration. *Biophysical Journal* 67: 2060-8
10. Shattil SJ, Budzynski A, Scrutton MC. 1989. Epinephrine Induces Platelet Fibrinogen Receptor Expression, Fibrinogen Binding, and Aggregation in Whole-Blood in the Absence of Other Excitatory Agonists. *Blood* 73: 150-8
11. Warkentin TE, Powling MJ, Hardisty RM. 1990. Measurement of Fibrinogen Binding to Platelets in Whole-Blood by Flow-Cytometry - a Micromethod for the Detection of Platelet Activation. *British Journal of Haematology* 76: 387-94
12. Abrams CS, Ruggeri ZM, Taub R, Hoxie JA, Nagaswami C, Weisel JW, Shattil SJ. 1992. Antiidiotypic Antibodies against an Antibody to the Platelet Glycoprotein (Gp) lib-IIa Complex Mimic Gp lib-IIa by Recognizing Fibrinogen. *Journal of Biological Chemistry* 267: 2775-85
13. Zamarron C, Ginsberg MH, Plow EF. 1991. A Receptor-Induced Binding-Site in Fibrinogen Elicited by Its Interaction with Platelet Membrane Glycoprotein-lib-IIa. *Journal of Biological Chemistry* 266: 16193-9
14. Xia ZM, Frojmovic MM. 1994. Aggregation Efficiency of Activated Normal or Fixed Platelets in a Simple Shear Field - Effect of Shear End Fibrinogen Occupancy. *Biophysical Journal* 66: 2190-201
15. Xia Z, Wong T, Liu Q, Kasirer-Friede A, Brown E, Frojmovic MM. 1996. Optimally functional fluorescein isothiocyanate-labelled fibrinogen for quantitative studies of binding to activated platelets and platelet aggregation. *British Journal of Haematology* 93: 204-14
16. Bennett JS, Vilaire G. 1979. Exposure of Platelet Fibrinogen Receptors by Adp and Epinephrine. *Journal of Clinical Investigation* 64: 1393-401
17. Xia Z, Wong T, Liu Q, Kasirer-Friede A, Brown E, Frojmovic MM. 1996. Optimally functional fluorescein isothiocyanate-labelled fibrinogen for quantitative studies of

- binding to activated platelets and platelet aggregation. *British Journal of Haematology* 93: 204-14
18. Holmes MB, Sobel BE, Schneider DJ. 1999. Variable responses to inhibition of fibrinogen binding induced by tirofiban and eptifibatide in blood from healthy subjects. *The American Journal of Cardiology* 84: 203-7

## **APPENDIX 2.**

### **Publications**





Contents lists available at ScienceDirect

## Bioorganic &amp; Medicinal Chemistry Letters

journal homepage: [www.elsevier.com/locate/bmcl](http://www.elsevier.com/locate/bmcl)

## RGD mimetics containing phthalimidine fragment as novel ligands of fibrinogen receptor

Andrei A. Krysko<sup>a,\*</sup>, Georgiy V. Samoylenko<sup>a</sup>, Pavel G. Polishchuk<sup>a,b</sup>, Sergei A. Andronati<sup>a</sup>, Tatyana A. Kabanova<sup>a</sup>, Tetiana M. Khristova<sup>a,b</sup>, Victor E. Kuz'min<sup>a</sup>, Vladimir M. Kabanov<sup>a</sup>, Olga L. Krysko<sup>a</sup>, Alexandre A. Varnek<sup>b</sup>, Ruslan Ya. Grygorash<sup>a</sup>

<sup>a</sup>A.V. Bogatsky Physico-Chemical Institute of the National Academy of Sciences of Ukraine, Lustdorfskaya doroga 86, Odessa 65080, Ukraine

<sup>b</sup>Laboratory of Chemoinformatics, University of Strasbourg, 1, rue B. Pascal, Strasbourg 67000, France

## ARTICLE INFO

## Article history:

Received 19 May 2011

Revised 15 July 2011

Accepted 17 July 2011

Available online 26 July 2011

## Keywords:

Fibrinogen receptor antagonists

 $\alpha_{IIb}\beta_3$ 

RGD mimetics

Phthalimidine

Platelet aggregation

## ABSTRACT

The novel RGD mimetics with phthalimidine central fragment were synthesized with the use of 4-piperidine-4-yl-butyric, 4-piperidine-4-yl-benzoic, 4-piperazine-4-yl-benzoic and 1,2,3,4-tetrahydroisoquinoline-7-carboxylic acids as surrogates of Arg motif. The synthesized compounds potently inhibited platelet aggregation in vitro and blocked FITC-Fg binding to  $\alpha_{IIb}\beta_3$  integrin in a suspension of washed human platelets. The key  $\alpha_{IIb}\beta_3$  protein–ligand interactions were determined in docking experiments.

© 2011 Elsevier Ltd. All rights reserved.

Over the last twenty years, the fibrinogen receptor GP IIb/IIIa (or integrin  $\alpha_{IIb}\beta_3$ ), has attracted a considerable attention as a promising therapeutic target and its antagonists were applied for treatment of thrombotic disorders such as unstable angina, myocardial infarction, ischemic disease, atherosclerosis, and stroke.<sup>1</sup> In the majority of cases, the design of  $\alpha_{IIb}\beta_3$  antagonists has been based on the modeling of Arg-Gly-Asp (RGD) sequence.<sup>1b,2</sup> The main binding sites of RGD sequence are  $\delta$ -guanidine of arginine and  $\beta$ -carboxylic group of aspartyl. Previously, to create non-peptide fibrinogen receptor antagonists mimicking RGD sequence, *p*-benzamide, piperidine, isoindoline, tetrahydroisoquinoline<sup>1b,2,3</sup> and  $\beta$ -alanine<sup>4</sup> containing fragments were successfully utilised as bioisosteres of arginine and Asp moieties correspondingly. This approach led the development of Tirofiban (Aggrastat<sup>®</sup>) approved for treatment of patients with unstable angina.<sup>5</sup> Numerous bicyclic scaffolds have been also used as RGD peptidomimetic blocks including indole,<sup>6</sup> 3,4-dihydro-2*H*-benzopyran,<sup>7</sup> tetrahydronaphthalene,<sup>7b</sup> 1,2,3,4-tetrahydroisoquinoline,<sup>7c</sup> 3-oxo-1,4-benzodiazepine,<sup>8</sup> 3,4-dihydro-2*H*-1,4-benzoxazine,<sup>9</sup> benzimidazole, benzoxazole,<sup>10</sup> 1,2,4-triazolo[3,4-*a*]pyridine,<sup>11</sup> etc.<sup>1b</sup>

Phthalimidines (2,3-dihydroisoindol-1-one) exhibit a wide spectrum of biological activities, for example, antagonism of

serotonin 5-HT<sup>12a-c</sup> and dopamine D<sup>12c</sup> receptors, inhibition of TNF- $\alpha$ <sup>12d</sup> and thromboxane A<sub>2</sub><sup>12e</sup> as well as documented antifungal and antibacterial activity.<sup>12f</sup> Indobufen, a phthalimidine derivative, is a platelet aggregation inhibitor.<sup>12g</sup> The antagonist of  $\alpha_{IIb}\beta_3$  **L-709,780** containing phthalimidine fragment inhibits ADP-induced platelet aggregation of human gel-filtered platelets with an IC<sub>50</sub> of 0.025  $\mu$ M.<sup>12h</sup>

This Letter describes the synthesis of new RGD mimetics containing a phthalimidine fragment and the study of their anti-aggregative properties. Furthermore, we demonstrate the utility of 4-piperidine-4-yl-butyric, 4-piperidine-4-yl-benzoic, 4-piperazine-4-yl-benzoic and 1,2,3,4-tetrahydroisoquinoline-7-carboxylic acid residues as Arg surrogates for RGD mimetic design.

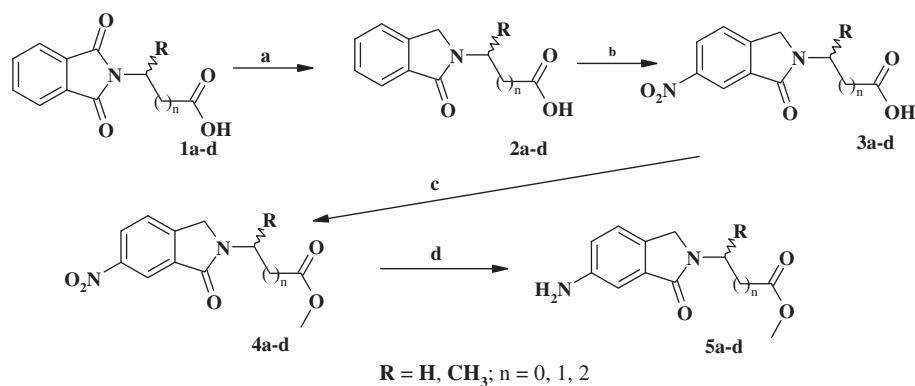
Initial Boc derivatives of 4-piperidine-4-yl-butyric (**6a**),<sup>13</sup> 4-piperidine-4-yl-benzoic (**6b**),<sup>14</sup> 4-piperazine-4-yl-benzoic (**6c**)<sup>15</sup> and 1,2,3,4-tetrahydroisoquinoline-7-carboxylic (**6d**)<sup>16</sup> acids have been synthesized using previously published methods.

The synthesis of aminophthalimidine building blocks **5a–d** is shown in Scheme 1 and the synthesis of target RGD mimetics **8a–i** is presented in Scheme 2. The central step of formation of phthalimidines **5a–d** is the reduction of phthalimides **1a–d** using zinc amalgam. The nitro derivatives **3a–d** obtained by nitration of the compounds **2a–d** were further reduced by H<sub>2</sub>/Pd(C).

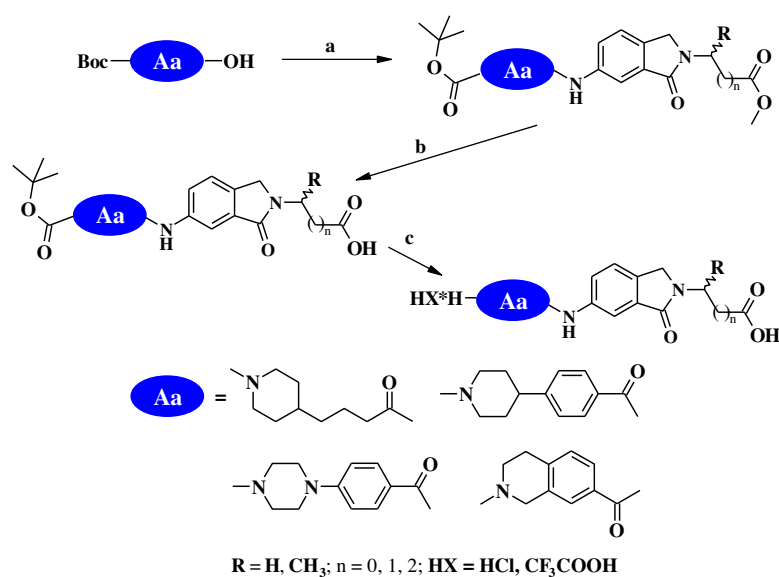
Condensation of acids **6a–c** with amines **5a–d** was carried out using HATU. Subsequent saponification of ester groups of compounds **7a–n** and elimination of Boc-protective groups yielded

\* Corresponding author.

E-mail addresses: [peptides@paco.net](mailto:peptides@paco.net), [peptides@physchem.od.ua](mailto:peptides@physchem.od.ua) (A.A. Krysko).



**Scheme 1.** Synthesis of aminophthalimidines. Reagents and conditions: (a) Zn (Hg), HCl, reflux, 4 h, 65–70%; (b) HNO<sub>3</sub>, H<sub>2</sub>SO<sub>4</sub>, –5 °C, 6 h, 54–65%; (c) MeOH (H<sub>2</sub>SO<sub>4</sub>), reflux, 3 h, 74–87%; (d) H<sub>2</sub>/Pd (C), MeOH, room temperature, 7 h, 82–92%.



**Scheme 2.** Synthesis of RGD mimetics. Reagents and conditions: (a) NEt<sub>3</sub>, HATU, **5a–d**, room temperature, overnight, 46–58%; (b) (i) 1 M NaOH, H<sub>2</sub>O, MeOH, room temperature, overnight; (ii) 1 M HCl, 73–81%; (c) CH<sub>2</sub>Cl<sub>2</sub>, HCl gas, 0 °C, 1 h or CH<sub>2</sub>Cl<sub>2</sub>, TFA, 0 °C, 2 h, 92–96%.

the target mimetics **9a–n**. Mimetics **9h–k** were obtained only as racemic mixtures in order to reveal potent compounds and to determinate general characteristics of structure–activity relationships.

Biological activity was assessed *in vitro* by measuring the ability of synthesized compounds to inhibit the binding of fluorescein-isothiocyanate-labeled fibrinogen (FITC-Fg)<sup>17</sup> to  $\alpha_{\text{IIb}}\beta_3$  (in a suspension of human washed platelets).<sup>18</sup> Functional activity was determined by measuring the inhibition of ADP induced platelet aggregation in human platelet-rich plasma (PRP) by Born's method.<sup>19</sup>

Experimental data (Table 1) demonstrate that mimetics **9** are potent inhibitors of FITC-Fg binding to  $\alpha_{\text{IIb}}\beta_3$  in activated platelets. We found that peptidomimetics **9d–k** with the  $\beta$ -alanine fragment exhibited higher antiaggregative activities compared to compounds **9a–c** and **9l–n**. This suggests that  $\beta$ -amino acids might be the optimal blocks for the  $\alpha_{\text{IIb}}\beta_3$  antagonists design. The compounds **9f** and **9g** displayed the highest antiaggregative activity among the synthesized mimetics with unsubstituted  $\beta$ -alanine as the Asp bioisostere (Table 1), whereas the antiaggregative activity of **9g** exceeds that of **9d** nearly fivefold, and that of **9e** nearly ninefold.

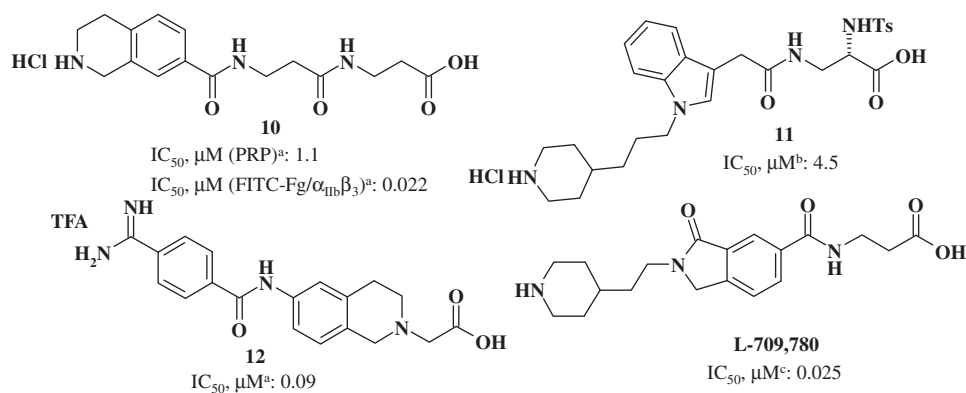
Comparison of antiaggregative properties of the mimetic **9f** with the previously described compound **10** (Fig. 1) containing

fragments of 1,2,3,4-tetrahydroisoquinoline-7-carboxylic acid and  $\beta$ -alanine, demonstrated that these two antagonists inhibit platelet aggregation with similar potencies. The affinity of the mimetic **9g** for  $\alpha_{\text{IIb}}\beta_3$  on activated platelets was slightly higher than the affinity of the mimetic **10**. The antiaggregative activity of **9f** was found to be twice higher than that of **9g**. Replacing  $\beta$ -alanine with  $\beta$ -methyl- $\beta$ -alanine leads to compounds **9h–k**. Their antiaggregative properties were generally similar to those of unsubstituted ones, with the only exception of the mimetic **9k**. This compound possessed the highest antiaggregative activity among all compounds synthesized in this study. Furthermore, antiaggregative activity of **9k** is comparable to those of compounds **11** and **12** containing bicyclic cores reported in the literature, but in the case of **L-709,780** the comparison is not valid because of the difference in biomaterials used in the assays.

Ligand to protein docking study for all synthesized compounds has been performed with the MOE<sup>20</sup> program. The structure of  $\alpha_{\text{IIb}}\beta_3$  receptor-Tirofiban complex (**2VDM**) was extracted from the Protein Data Bank.<sup>21</sup> For docking, all water molecules were removed from the binding pocket of **2VDM** with the exception of two molecules coordinated with the metal ion-dependent adhesive site (MIDAS) of  $\alpha_{\text{IIb}}\beta_3$  receptor – Mg<sup>2+</sup>, which is involved in the interaction with carboxyl group of Tirofiban. Then, structure of the complex was optimized using the MMFF94x force field. For initial ligand Tirofiban,

**Table 1**  
Biological properties of RGD mimetics **9** and RGDS peptide

Compound	HX	Aa	n	R	IC <sub>50</sub> , μM (PRP) <sup>a</sup>	IC <sub>50</sub> , μM (FITC-Fg/α <sub>IIb</sub> β <sub>3</sub> ) <sup>b</sup>
<b>9a</b>	TFA		0	H	66.0 ± 9.0	–
<b>9b</b>	TFA		0	H	24.0 ± 3.0	0.27 ± 0.06
<b>9c</b>	HCl		0	H	120.0 ± 20.0	1.2 ± 0.1
<b>9d</b>	TFA		1	H	5.9 ± 0.6	0.0055 ± 0.009
<b>9e</b>	TFA		1	H	9.6 ± 1.9	0.0068 ± 0.0012
<b>9f</b>	TFA		1	H	0.54 ± 0.06	–
<b>9g</b>	HCl		1	H	1.1 ± 0.1	0.0065 ± 0.0005
<b>9h</b>	HCl		1	CH <sub>3</sub>	5.4 ± 1.0	0.35 ± 0.03
<b>9i</b>	HCl		1	CH <sub>3</sub>	6.2 ± 1.2	–
<b>9j</b>	HCl		1	CH <sub>3</sub>	3.74 ± 0.51	0.037 ± 0.08
<b>9k</b>	HCl		1	CH <sub>3</sub>	0.086 ± 0.007	0.0065 ± 0.0012
<b>9l</b>	TFA		2	H	51.0 ± 30.0	–
<b>9m</b>	TFA		2	H	410.0 ± 60.0	–
<b>9n</b>	HCl		2	H	330.0 ± 50.0	–
	RGDS				31.0 ± 2.0	13.0 ± 1.6

<sup>a</sup> Concentration required to reduce ADP-induced human platelet aggregation response by 50%.<sup>b</sup> Concentration required to reduce binding of FITC-Fg to α<sub>IIb</sub>β<sub>3</sub> in suspensions of washed human platelets by 50%. The IC<sub>50</sub> values are expressed as averages of at least two determinations.**Figure 1.** Structure and in vitro activities of α<sub>IIb</sub>β<sub>3</sub> antagonists **10**,<sup>16</sup> **11**,<sup>6</sup> **12**,<sup>7c</sup> and **L-709,780**.<sup>12h</sup> <sup>a</sup>See Table 1; <sup>b</sup>inhibition of guinea-pig PRP aggregation induced by collagen; <sup>c</sup>inhibition of ADP-induced platelet aggregation of human gel-filtered platelets.

which was a part of **2VDM**, re-docking had been carried out with good result (RMSD = 0.6 Å). Thus α<sub>IIb</sub>β<sub>3</sub> cavity was prepared, and all further studies of compounds **9** docking were carried out for this binding site of the receptor.

The docking studies have revealed some general patterns for binding of mimetics **9** to their receptor α<sub>IIb</sub>β<sub>3</sub>. It has been shown that the nitrogen atom of Arg-isosteres (the fragments of 4-piperidine-4-yl-butyrin, 4-piperidine-4-yl-benzoic, 4-piperazine-4-yl-benzoic

and 1,2,3,4-tetrahydroisoquinoline-7-carboxylic acids) interacts with two amino acid residues of α chain of fibrinogen receptor, namely with carboxyl group of D224 side chain and S225 amide bond. Carboxyl group of mimetics **9** is involved in coordination sphere of Mg<sup>2+</sup>, and also interacts with Y122 amide bond and amide group of N215 side chain incorporated in the α<sub>IIb</sub>β<sub>3</sub> β-chain. These interactions are illustrated in Figure 2 using the complex of **9g** as an example.

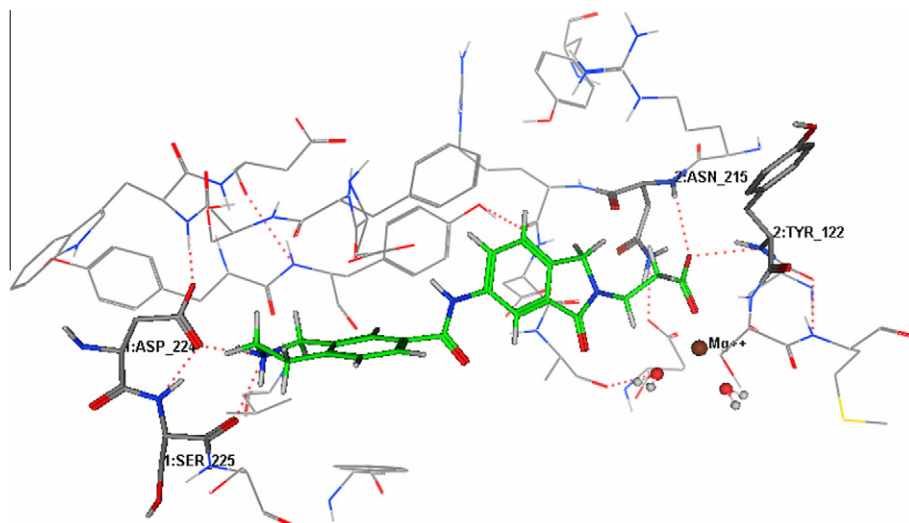


Figure 2. Binding of mimetic **9g** to the  $\alpha_{IIb}\beta_3$  receptor observed in docking experiments.

Docking studies revealed that *S*-enantiomers of compounds **9i** and **9j** bind to  $\alpha_{IIb}\beta_3$  stronger than corresponding *R*-enantiomers. In the case of compounds **9h** and **9k**, no significant differences in values of scoring function for *S* and *R*-enantiomers was observed. Detailed investigation of the relationship between the stereochemistry of the mimetics **9** and their affinity for  $\alpha_{IIb}\beta_3$  will be the purpose of future studies.

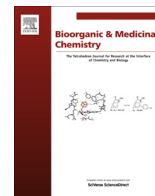
To summarize, for a series of 4-piperidine-4-yl-butyric, 4-piperidine-4-yl-benzoic and 1,2,3,4-tetrahydroisoquinoline-7-carboxylic acids derivatives, no significant influence of Arg isostere structure on affinity for  $\alpha_{IIb}\beta_3$  was observed. At the same time, the compounds **9d–k** have demonstrated a high affinity for  $\alpha_{IIb}\beta_3$  and acceptable antiaggregative activity. A combination of tetrahydroisoquinoline and  $\beta$ -methyl- $\beta$ -alanine as Arg and Asp bioisosteres, respectively, within the given phthalimidine series, leads to the best prospective inhibitor of the platelet aggregation. It has been shown by docking studies that  $\alpha$ :D224,  $\alpha$ :S225 and  $\beta$ :Y22,  $\beta$ :N215 residues of  $\alpha_{IIb}\beta_3$  integrin play the key role in binding of the mimetics **9** to the receptor.

## Acknowledgments

The author thanks Dr. M. Kopanitsa (Cambridge, United Kingdom) for helpful discussions and the proof reading of the manuscript.

## References and notes

- (a) Tsikaris, V. J. *Pept. Sci.* **2004**, *10*, 589; (b) Andronati, S. A.; Karaseva, T. L.; Krysko, A. A. *Curr. Med. Chem.* **2004**, *11*, 1183.
- Perdih, A.; Sollner Dolenc, M. *Curr. Med. Chem.* **2010**, *17*, 2371.
- (a) Masic, L. P.; Kikelj, D. *Tetrahedron* **2001**, *57*, 7073–7105; (b) Masic, L. P. *Curr. Med. Chem.* **2006**, *13*, 3627; (c) Krysko, A. A.; Chugunov, B. M.; Malovichko, O. L.; Andronati, S. A.; Kabanova, T. A.; Karaseva, T. L.; Kiriyak, A. V. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5533; (d) Andronati, S. A.; Krysko, A. A.; Chugunov, B. M.; Kabanova, T. A.; Artemenko, A. G. *Russ. J. Org. Chem.* **2006**, *42*, 1174; (e) Malovichko, O. L.; Petrus, A. S.; Krysko, A. A.; Kabanova, T. A.; Andronati, S. A.; Karaseva, T. L.; Kiriyak, A. V. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5294; (f) Krysko, A. A.; Krysko, O. L.; Kabanova, T. A.; Andronati, S. A.; Kabanov, V. M. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 4444.
- Stefanic, P.; Dolenc, M. S. *Curr. Med. Chem.* **2004**, *11*, 945.
- (a) Hartman, G. D.; Egbertson, M. S.; Halczenko, W.; Laswell, W. L.; Duggan, M. E.; Smith, R. L.; Naylor, A. M.; Manno, P. D.; Lynch, R. J.; Zhang, G.; Chang, C. T.-C.; Gould, R. J. *J. Med. Chem.* **1992**, *35*, 4640; (b) Egbertson, M. S.; Chang, C. T.-C.; Duggan, M. E.; Gould, R. J.; Halczenko, W.; Hartman, G. D.; Laswell, W. L.; Lynch, R. J.; Lynch, R. J.; Manno, P. D.; Naylor, A. M.; Prugh, J. D.; Ramjit, D. R.; Sitko, G. R.; Smith, R. S.; Turchi, L. M.; Zhang, G. *J. Med. Chem.* **1994**, *37*, 2537; (c) Cook, J. J.; Bednar, B.; Lynch, J. J.; Gould, R. J.; Egbertson, M. S.; Halczenko, W.; Duggan, M. E.; Hartman, G. D.; Lo, M.; Murphy, G. M.; Deckelbaum, L. I.; Sax, F. L.; Barr, E. *Cardiovasc. Drug Rev.* **1999**, *17*, 199.
- Grumel, V.; Mérou, J.-Y.; Lesur, B.; Giboulot, T.; Frydman, A.; Guillaumet, G. *Eur. J. Med. Chem.* **2002**, *37*, 45.
- (a) Anderluh, M.; Marc, G.; Dolenc, M. S. *Synth. Commun.* **2005**, *35*, 1461; (b) Okumura, K.; Shimazaki, T.; Aoki, Y.; Yamashita, H.; Tanaka, E.; Banba, S.; Yazawa, K.; Kibayashi, K.; Banno, H. *J. Med. Chem.* **1998**, *41*, 4036; (c) Fisher, M. J.; Arfstan, A. E.; Giese, U.; Gunn, B. P.; Harms, C.; Khau, S. V.; Kinnick, M. D.; Lindstrom, T. D.; Martinelli, M. J.; Mest, H.-J.; Mohr, M. J.; Morin, J. M., Jr.; Mullaney, J. T.; Nunes, A.; Paal, M.; Rapp, A.; Rühler, G.; Ruterbories, K. J.; Sall, D. J.; Scarborough, R. M.; Schotten, T.; Sommer, B.; Stenzel, W.; Towner, R. D.; Um, S. L.; Utterback, B. G.; Vasileff, R. T.; Vöelkers, S.; Wyss, V. L.; Jakubowski, J. A. *J. Med. Chem.* **1999**, *42*, 4875.
- (a) Ku, T. W.; Ali, F. E.; Bondinell, W. E.; Erhard, K. F.; Huffman, W. F.; Venslavsky, J. W.; Yuan, C. C.-K. *Tetrahedron Lett.* **1997**, *38*, 3131; (b) Atkins, R. J.; Banks, A.; Bellingham, R. K.; Breen, G. F.; Carey, J. S.; Etridge, S. K.; Hayes, J. F.; Hussain, N.; Morgan, D. O.; Oxley, P.; Passey, S. C.; Walsgrove, T. C.; Wells, A. S. *Org. Process Res. Dev.* **2003**, *7*, 663.
- (a) Anderluh, M.; Cesar, J.; Štefanič, P.; Kikelj, D.; Janeš, D.; Murn, J.; Nadrah, K.; Tominc, M.; Addicks, E.; Giannis, A.; Kigek, M.; Dolenc, M. S. *Eur. J. Med. Chem.* **2005**, *40*, 25; (b) Anderluh, P. Š.; Anderluh, M.; Ila, J.; Mravljak, J.; Dolenc, M. S.; Stegnar, M.; Kikelj, D. *J. Med. Chem.* **2005**, *48*, 3110.
- Xue, C.-B.; Rafalski, M.; Roderick, J.; Eyerann, C. J.; Mousa, S.; Olson, R. E.; DeGrado, W. F. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 339.
- Lawson, E. C.; Hoekstra, W. J.; Addo, M. F.; Andrade-Gordon, P.; Damiano, B. P.; Kauffman, J. A.; Mitchell, J. A.; Maryanoff, B. E. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2619.
- (a) Zhuang, Z.-P.; Kung, M.-P.; Mu, M.; Kung, H. F. *J. Med. Chem.* **1998**, *41*, 157; (b) Hamprecht, D.; Fabrizio, M.; Tedesco, G.; Checchia, A.; Donati, D.; Petrone, M.; Terreni, S.; Wood, M. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 428; (c) Norman, M. H.; Rigdon, G. C.; Navas, F.; Cooper, B. R. *J. Med. Chem.* **1994**, *37*, 2552; (d) Luo, W.; Yu, Q.-S.; Salcedo, I.; Holloway, H. W.; Lahiri, D. K.; Brossi, A.; Tweedie, D.; Greig, N. H. *Bioorg. Med. Chem.* **2011**, *19*, 3965; (e) Kato, Y.; Ebiike, H.; Achiwa, K.; Ashizawa, N.; Kurihara, T.; Kobayashi, F. *Chem. Pharm. Bull.* **1990**, *38*, 2060; (f) Cai, S.-L.; Chen, Y.; Sun, W.-X.; Li, H.; Chen, Y.; Yuan, S.-S. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 5649; (g) Cimminiello, C.; Milani, M.; Uberti, T.; Arpaia, G.; Bonfardeci, G. *J. Int. Med. Res.* **1989**, *17*, 514; (h) Brashear, K. M.; Cook, J. J.; Bednar, B.; Bednar, R. A.; Gould, R. J.; Halczenko, W.; Holahan, M. A.; Lynch, R. J.; Hartman, G. D.; Hutchinson, J. H. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2793.
- Klein, S. I.; Molino, B. F.; Czekaj, M.; Gardner, C. J.; Chu, V.; Brown, K.; Sabatino, R. D.; Bostwick, J. S.; Kasiewski, C.; Bentley, R.; Windisch, V.; Perrone, M.; Dunwiddie, C. T.; Leadley, R. J. *J. Med. Chem.* **1998**, *41*, 2492.
- Frank, H.; Guenter, L.; Volkhard, A.; Helmut, P.; Thomas, M.; Johannes, W.; Brian, G. Eur. Patent 0604800, 1993.
- Duggan, M. E.; Egbertson, M. S.; Hartman, G. D.; Young, S. D.; Ihle, N. C. US Patent 5854245, 1998.
- Malovichko, O. L.; Krysko, A. A.; Kabanova, T. A.; Andronati, S. A.; Grishkovets, V. I.; Kachala, V. V.; Panov, D. A. *Med. Chem.* **2009**, *5*, 158.
- Hantgan, R. *Biochim. Biophys. Acta* **1987**, *927*, 55.
- Xia, Z.; Wong, T.; Liu, Q.; Kasirer-Friede, A.; Brown, E.; Frojmnivc, M. M. *Br. J. Haematol.* **1996**, *93*, 204.
- Born, G. V. R. *Nature* **1962**, *194*, 927.
- Molecular Operating Environment (MOE)*, 2010.10; Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2010.
- <http://dx.doi.org/10.2210/pdb2vdm/pdb>; (b) Xiao, T.; Takagi, J.; Collier, B. S.; Wang, J.-H.; Springer, T. A. *Nature* **2004**, *432*, 59.



# Synthesis, biological evaluation, X-ray molecular structure and molecular docking studies of RGD mimetics containing 6-amino-2,3-dihydroisoindolin-1-one fragment as ligands of integrin $\alpha_{IIb}\beta_3$ <sup>☆</sup>



Andrei A. Krysko <sup>a,\*</sup>, Georgiy V. Samoilenko <sup>a</sup>, Pavel G. Polishchuk <sup>a,b</sup>, Marina S. Fonari <sup>c</sup>, Victor Ch. Kravtsov <sup>c</sup>, Sergei A. Andronati <sup>a</sup>, Tatyana A. Kabanova <sup>a</sup>, Janusz Lipkowski <sup>d</sup>, Tetiana M. Khristova <sup>a,b</sup>, Victor E. Kuz'min <sup>a</sup>, Vladimir M. Kabanov <sup>a</sup>, Olga L. Krysko <sup>a</sup>, Alexandre A. Varnek <sup>b</sup>

<sup>a</sup>A.V. Bogatsky Physico-Chemical Institute of the National Academy of Sciences of Ukraine, Lustdorfskaya Doroga 86, Odessa 65080, Ukraine

<sup>b</sup>Laboratory of Chemoinformatics, University of Strasbourg, 1 Rue B. Pascal, Strasbourg 67000, France

<sup>c</sup>Institute of Applied Physics, Academy of Sciences of Moldova, Academiei 5, Chisinau 2028, Republic of Moldova

<sup>d</sup>Institute of Physical Chemistry, Polish Academy of Sciences, Kasprzaka 44, 01 224 Warsaw, Poland

## ARTICLE INFO

### Article history:

Received 31 January 2013

Revised 1 May 2013

Accepted 13 May 2013

Available online 22 May 2013

### Keywords:

Fibrinogen receptor antagonists

$\alpha_{IIb}\beta_3$

RGD mimetics

Phthalimidine

Platelet aggregation

## ABSTRACT

A series of novel RGD mimetics containing phthalimidine fragment was designed and synthesized. Their antiaggregative activity determined by Born's method was shown to be due to inhibition of fibrinogen binding to  $\alpha_{IIb}\beta_3$ . Molecular docking of RGD mimetics to  $\alpha_{IIb}\beta_3$  receptor showed the key interactions in this complex, and also some correlations have been observed between values of biological activity and docking scores. The single crystal X-ray data were obtained for five mimetics.

© 2013 The Authors. Published by Elsevier Ltd. All rights reserved.

## 1. Introduction

Thrombosis is the most important pathological process underlying many cardiovascular diseases, which are responsible for elevated mortality worldwide.<sup>1</sup> The use of ADP and thromboxane inhibition does not preclude the formation of thrombus, because platelets can still be activated by other mechanisms. It gave a stimulus to the development of an entirely separate class of antiplatelet drugs - fibrinogen receptor (integrin  $\alpha_{IIb}\beta_3$ ) antagonists. During the platelet activation process, the surface of the platelet transforms its shape to expose the fibrinogen receptors. These receptors bind to fibrinogen and Von Willebrand factor, resulting in clot formation and clot adherence, respectively.<sup>2</sup> Binding of fibrinogen to  $\alpha_{IIb}\beta_3$  on platelets is responsible for securing aggregated platelets to one another. Thus, blocking these receptors prevents platelet

aggregation regardless of the activation pathway.<sup>3,4</sup> Inhibitors of  $\alpha_{IIb}\beta_3$  have proven useful in reducing the risk of periprocedural MI and urgent target vessel revascularization during catheterization and have claimed a place in therapy for these indications.<sup>5</sup> There exist three different  $\alpha_{IIb}\beta_3$  inhibitors approved for use: abciximab (Eli Lilly & Company, Indianapolis, Indiana), eptifibatid (Schering-Plough, Kenilworth, New Jersey), and tirofiban (Medicure, Winnipeg, Canada).<sup>6</sup> However, the current agents have several limitations, including the need for intravenous infusion and, most important, the induction of thrombocytopenia in some patients.<sup>7</sup> These limitations of above mentioned drugs give evidence of need in new  $\alpha_{IIb}\beta_3$  inhibitors.

Although, more frequently the design of  $\alpha_{IIb}\beta_3$  antagonists is based on the mimicking of Arg–Gly–Asp (RGD) sequence, alternative approach represents the use of dodecapeptide sequence.<sup>8,9</sup> The main binding sites of RGD sequence are  $\delta$ -guanidine of arginine and  $\beta$ -carboxylic group of aspartyl, and for the dodecapeptide sequence - lysine amino group and aspartyl carboxylic group, correspondingly. Similarity of these two approaches is evident.

In our recent publication, the use of phthalimidine scaffolds for designing potent integrin  $\alpha_{IIb}\beta_3$  antagonists has been demon-

<sup>☆</sup> This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

\* Corresponding author. Tel.: +380 308663041; fax: +380 482652012.

E-mail addresses: [peptides@paco.net](mailto:peptides@paco.net), [peptides@physchem.od.ua](mailto:peptides@physchem.od.ua) (A.A. Krysko).

strated.<sup>10</sup> Herein, we describe the optimization of the novel Arg surrogates for obtaining RGD mimetics active inhibitors.

This article describes the synthesis of new RGD mimetics containing phthalimidine (2,3-dihydroisoindol-1-one) fragment and study of their antiaggregative properties. We also discuss the possibility to use the residues of 4-piperidineacetic, 4-piperidine-4-yl-butyric, 4-piperidine-4-yl-benzoic, 4-piperazine-1-yl-benzoic, 1,2,3,4-tetrahydroisoquinoline-7-carboxylic, and 3-piperazine-1-yl-benzoic acids as Arg surrogates for RGD mimetics design.

## 2. Results and discussion

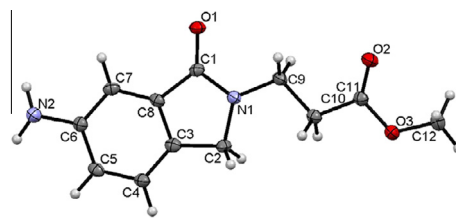
### 2.1. Synthesis

Description of the synthesis of 6-amino-2,3-dihydroisoindolin-1-one building blocks **5a–d** was given in a previous report (Scheme 1). The key step was the formation of the 6-amino-2,3-dihydroisoindolin-1-one building blocks **5a–d** by reduction of phthalimides **1a–d** using zinc amalgam. As a result of nitration of the compounds **2a–d**, there were obtained nitro-derivatives **3a–d** followed by their reduction by H<sub>2</sub>/Pd(C). Regioselectivity of nitration reaction is expected. The X-ray analysis data obtained for the compound **5b** is a confirmation.

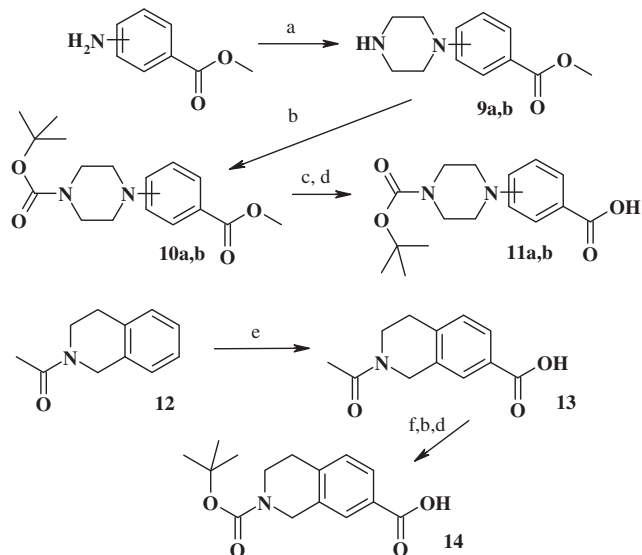
The view of molecular structure of **5b** is shown in Figure 1. The molecule has a rather flat skeleton as evidenced by the dihedral angle of 8.8° between the planes of almost coplanar nonhydrogen atoms of methylpropionate chain and phthalimide core. The geometric parameters of planar aminophthalimide core are in line with the related 3-(1-oxo-1,3-dihydroisoindol-2-yl)propionic acid,<sup>11</sup> but these molecules differ by the molecular shape as far as the latter one has an angular conformation.

Homologation of Boc protected piperidine-4-carboxylic acid gave the 1-Boc-4-piperidineacetic acid (**6**), by a similar method reported in the literature.<sup>12</sup> Initial Boc derivatives of the 4-piperidine-4-yl-butyric (**7**)<sup>13</sup> and 4-piperidine-4-yl-benzoic (**8**)<sup>14</sup> acids have been synthesized using previously published methods. Isomeric methyl esters of the 4-piperazine-4-yl-benzoic (**9a**)<sup>15</sup> and 3-piperazine-4-yl-benzoic (**9b**) acids were made using methyl esters of aminobenzoic acids and bis-(2-chloroethyl)amine. Boc-protection of the compounds **9a** and **9b** with Boc<sub>2</sub>O resulted in the methyl esters **10a** and **10b**, respectively. Subsequent saponification of ester groups of compounds **10a** and **10b** yielded the target acids **11a,b** (Scheme 2).

Precursor compound **14** was prepared in three steps in contrast to our previously published route for four-steps synthesis.<sup>16</sup> Acylation of 2-acetyl-1,2,3,4-tetrahydroisoquinoline (**12**) with oxalyl chloride at the presence of aluminum chloride gave 1-acetyl-1,2,3,4-tetrahydroisoquinoline-7-carboxylic acid (**13**). Further



**Figure 1.** View of **5b**. ORTEP diagram with the numbering scheme. The thermal ellipsoids are of 50% probability level, with hydrogens as spheres of arbitrary diameter.

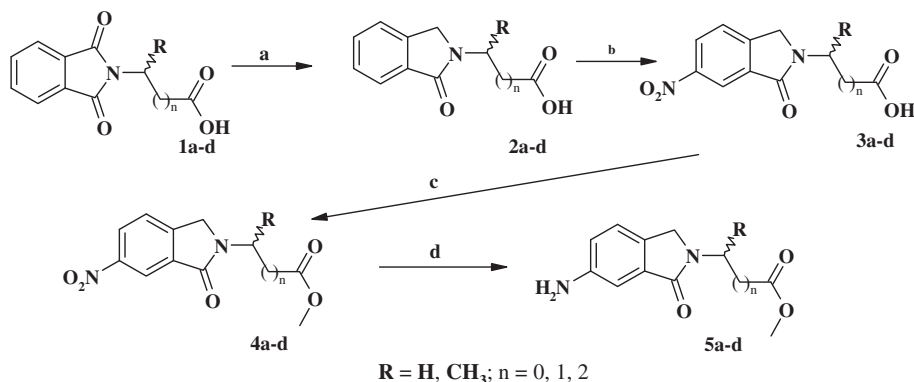


**Scheme 2.** Synthesis of the acids **11** and **14**. Reagents: (a) HCl·NH(CH<sub>2</sub>CH<sub>2</sub>Cl)<sub>2</sub>, K<sub>2</sub>CO<sub>3</sub>, BuOH, reflux, ≈177 h; (b) Boc<sub>2</sub>O, room temperature, overnight; (c) NaOH, H<sub>2</sub>O 40 °C, 6 h; (d) HCl, H<sub>2</sub>O; (e) (COCl)<sub>2</sub>, AlCl<sub>3</sub>; (f) NaOH, H<sub>2</sub>O.

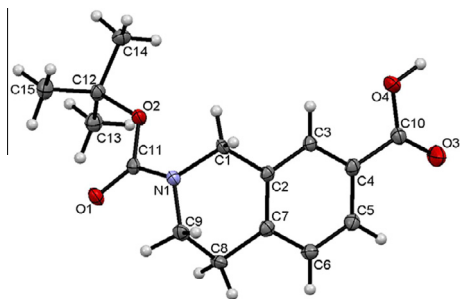
removal of the acetyl group and reaction with Boc<sub>2</sub>O resulted in the acid **14** (Scheme 2). The structure of compound **14** has been confirmed by data of X-ray analysis (Fig. 2).

The molecular structure **15** is shown in Figure 2. The molecule has an angular shape with the aza-cycle of the 1,2,3,4-tetrahydroisoquinoline core being in a half-chair conformation with the N1 atom deviating at  $-0.574(3)$  Å from the mean plane of cyclic C-atoms, the rms deviation of fitted carbon atoms being 0.056 Å.

Condensation of acids **6–8**, **11a,b** and **14** with amines **5a–d** has been conducted using the HBTU or HATU (Scheme 3). Subsequent saponification of ester groups of the compounds **15a–u** and



**Scheme 1.** Synthesis of aminophthalimidines. Reagents: (a) Zn (Hg), HCl, reflux, 4 h; (b) HNO<sub>3</sub>, H<sub>2</sub>SO<sub>4</sub>,  $-5$  °C, 6 h; (c) MeOH (H<sub>2</sub>SO<sub>4</sub>), reflux, 3 h; (d) H<sub>2</sub>/Pd (C), MeOH, room temperature, 7 h.



**Figure 2.** View of **15**. ORTEP diagram with the numbering scheme. The thermal ellipsoids are of 50% probability level, with hydrogens as spheres of arbitrary diameter.

elimination of Boc-protective groups yielded the target mimetics **17a–u**. Mimetics **17l–q** were obtained only as racemic mixtures in order to reveal potent compounds and to determinate general characteristics of structure–activity relationships.

## 2.2. X-ray structure

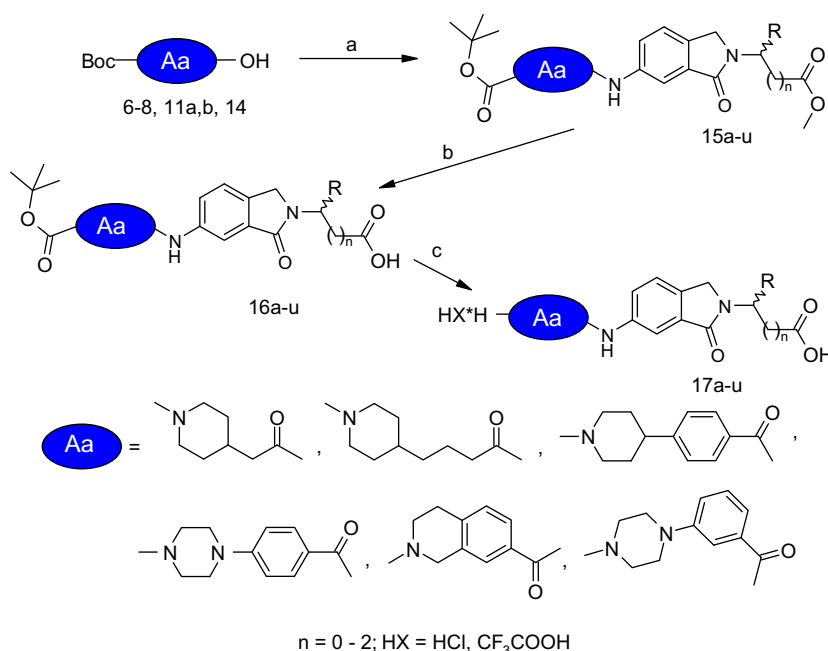
Fortunately, single crystals of target compounds were grown from water, and X-ray data were obtained for five mimetics, **17c,h,i,j,t**. They were fixed in the solid crystalline state as either trifluoroacetate or chloride salt with the compositions **17c,h** and **17i,j,t**·H<sub>2</sub>O. After 2 h of boiling of aqueous solutions of compounds **17c** and **17h** and then their cooling to room temperature, single crystals were grown. They were identified as the hydrates of zwitterionic forms, **17c**·TfaOH·5H<sub>2</sub>O and **17h**·TfaOH·2H<sub>2</sub>O. The ORTEP drawings for the formula units are shown in Figure 3. All the mimetics molecules bear three planar fragments, the benzoyl ring, the 1-oxo-1,3-dihydroisoindolyl core with the equivalent bond lengths being similar within themselves, in the parent **5b** and in the closely related 3-(1-oxo-1,3-dihydroisoindol-2-yl)propionic acid,<sup>11</sup> and the amide bridge between these moieties. The amide bridge is always fixed in the *syn*-conformation with regard to the

phthalimidine core and anti-conformation with regard to the benzoyl moieties as the torsion angles O2–C9–N2–C6 and C10–C9–N2–C6 indicate falling in the ranges 0.6(8)–8.7(9)° and 171.1(6)–179(1)°, respectively (Table 1). The molecules differ by mutual orientations of the above-mentioned planar molecular moieties, by the orientation of piperidine or piperazine fragment with respect to the benzoyl moieties, and carboxylic tails with respect to the indolyl core. Moreover, the different mutual arrangement (*cis*- or *trans*-) of the amide and indolyl carbonyl oxygen atoms O1 and O2 should be mentioned. The conformational flexibility and essential variation of cross-molecule distance between the distal cyclic N and carboxylic O-atoms in the wide range 16.297–21.434 Å, Table 1, indicate the possibility to fine tuning substrate-receptor interactions.

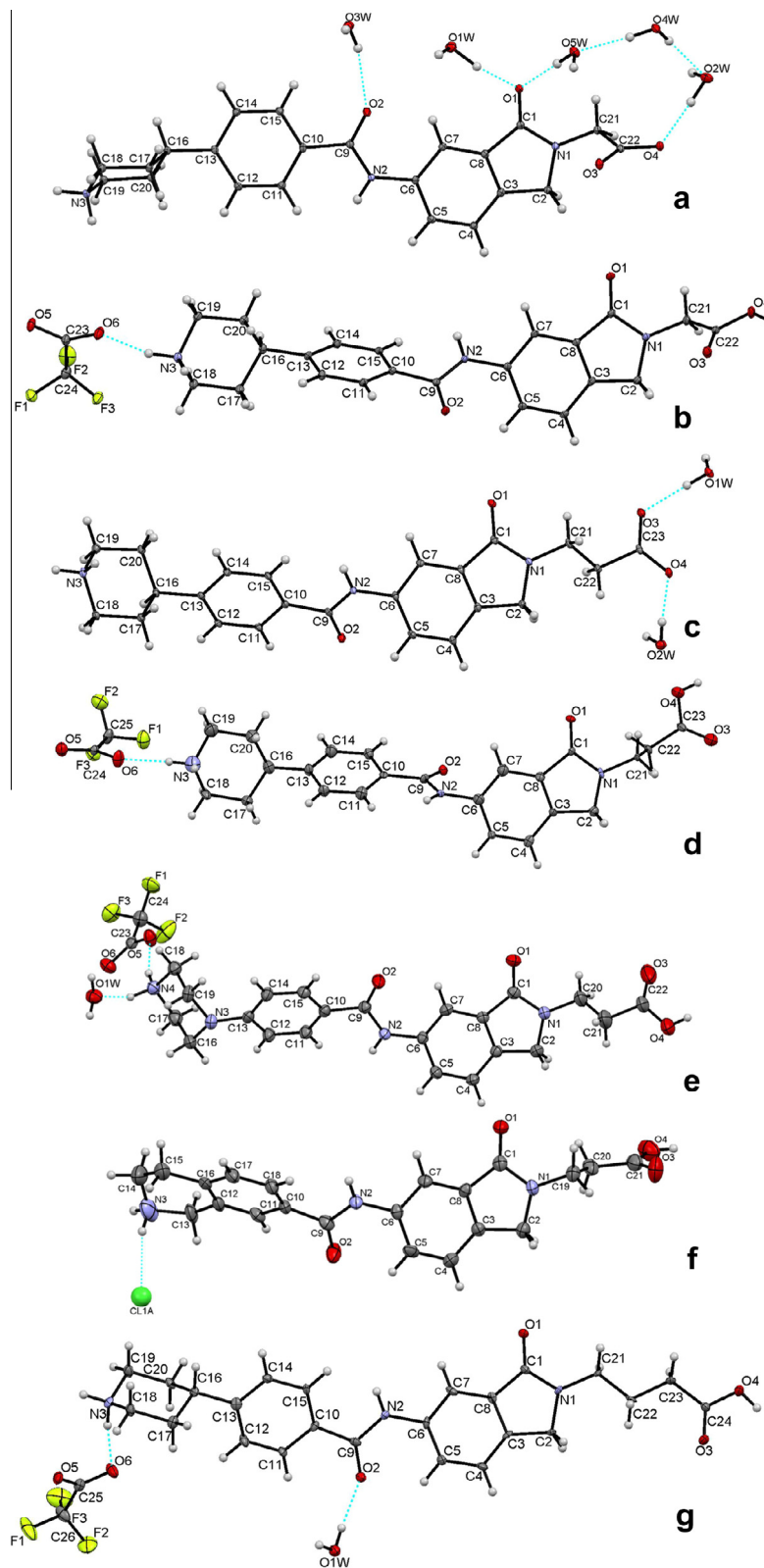
## 2.3. In vitro biology

Functional activity was determined by measuring the inhibition of ADP induced platelet aggregation in human platelet-rich plasma (PRP) by Born's method.<sup>17</sup> Mode of action for some compounds was subsequently revealed in vitro by measuring the ability of compounds to inhibit the binding of fluoresceinisothiocyanate-labeled fibrinogen (FITC-Fg)<sup>18</sup> to  $\alpha_{IIb}\beta_3$  (in a suspension of human washed platelets).<sup>19</sup> Experimental data (Table 2) evidently show high affinities of the compounds **17** for  $\alpha_{IIb}\beta_3$ . RGDS peptide and Tirofiban were used as standard inhibitors.

We first investigated the SAR of compounds **17a–e**, containing a fragment of (1-oxo-1,3-dihydroisoindol-2-yl)acetic acid. These compounds showed low in vitro antiaggregative activity, except for the leader of the group, mimetic **17d** containing 4-piperazine-4-yl-benzoyl, as Arg isostere. Homologation of C-terminal template resulted in increased activity for five compounds **17f–j** containing the same surrogates of Arg. Here again was the leading mimetic **17i**, containing a residue of 4-piperazine-4-yl-benzoic acid. Leading indicators of antiaggregative activity of compounds containing 4-piperazine-4-yl-benzoyl, forced us to test 3-piperazine-4-yl-benzoyl derivative **17k**. It was found that the compound **17k** was five times less active than its isomer **17i**. Further increase of C-terminal



**Scheme 3.** Synthesis of RGD mimetics. Reagents: (a) NEt<sub>3</sub>, HBTU, **5a–d**, room temperature, overnight; (b) 1 M NaOH, H<sub>2</sub>O, MeOH, room temperature, overnight, 1 M HCl; (c) CH<sub>2</sub>Cl<sub>2</sub>, HCl gas, 0 °C, 1 h or CH<sub>2</sub>Cl<sub>2</sub>, TFA, 0 °C, 2 h.



**Figure 3.** ORTEP diagram with the numbering scheme for **17c**-TfaOH·5H<sub>2</sub>O (a), **17c** (b), **17h**-TfaOH·2H<sub>2</sub>O (c), **17h** (d), **17i**·H<sub>2</sub>O (e), **17j**·H<sub>2</sub>O (f), and **18t**·H<sub>2</sub>O (g). The thermal ellipsoids are of 30% probability level, with hydrogens as spheres of arbitrary diameter. Only major positions are shown for the disordered fragments.

residue length by another one carbon atom led to a decrease in aggregative properties. Surprise for us was the value of IC<sub>50</sub> for compound **17r**. Mimetic **17r** was more active than the mimetic

**17b**, these two compounds have the same number of atoms between the N- and C-terminals. Finally, replacing C-terminal fragment β-alanine with β-methyl-β-alanine led to compounds **17l–p**.



**Table 1**  
Main geometrical parameters for **17** entities in studied compounds

Compounds	Geometric parameter						
	<sup>a</sup> $\phi_1$ (°)	<sup>b</sup> $\phi_2$ (°)	<sup>c</sup> $\phi_1$ (°)	<sup>d</sup> $\phi_2$ (°)	<sup>e</sup> $\phi_3$ (°)	<sup>f</sup> $\phi_4$ (°)	<sup>g</sup> $r$ (Å)
<b>17c</b> -TfaOH·5H <sub>2</sub> O	−2.5(3)	176.8(2)	84.87(5)	26.51(8)	7.48(6)	79.1(2)	16.466(2) 18.377(2)
<b>17c</b>	4.4(6)	−174.3(3)	87.5(1)	23.7(2)	39.7(1)	59.7(3)	17.256(5) 18.718(5)
<b>17h</b> -TfaOH·2H <sub>2</sub> O	−1.7(4)	175.6(3)	70.53(9)	4.9(2)	45.1(1)	10.3(3)	18.913(3) 19.6857(3)
<b>17h</b>	3(3)	179(1)	61.1(7)	15.3(9)	65.3(4)	9.8(9)	18.05(3) 19.10(3)
<b>17i</b> -H <sub>2</sub> O	2.2(4)	−175.3(2)	56.06(8)	22.3(1)	17.3(1)	8.6(6)	16.297(4) 16.885(4)
<b>17j</b> -H <sub>2</sub> O	−0.6(8)	−179.5(4)	3.8(3)	29.9(2)	30.9(2)	75.8(2)	16.803(7) 17.091(7)
<b>17t</b> -H <sub>2</sub> O	−8.7(9)	171.1(6)	80.2(2)	16.6(3)	15.8(3)	15.5(1)	20.235(7) 21.434(7)

<sup>a</sup> Torsion angle O2–C9–N2–C6.

<sup>b</sup> Torsion angle C10–C9–N2–C6.

<sup>c</sup> Dihedral angle between the average planes through the terminal piperidine/piperazine ring and phenyl ring, C10/C11/C12/C13/C14/C15.

<sup>d</sup> Dihedral angle between the average planes through the phenyl ring, C10/C11/C12/C13/C14/C15 and amide bridge, C10/C9/O2/N2/C6.

<sup>e</sup> Dihedral angle between the average planes through the amide bridge, C10/C9/O2/N2/C6 and phthalimidine core, C1 > C8/N1/O1.

<sup>f</sup> Dihedral angle between the average planes through the phthalimidine core, C1 > C6/N1/O1 and carboxylic group, defined by C,O,O atoms.

<sup>g</sup> Cross-molecule distance between the cyclic N and carboxylic O-atoms.

Effect of CH<sub>3</sub> group is ambiguous. In two cases, antiaggregative activity observed for mimetics containing  $\beta$ -methyl- $\beta$ -alanine fragment (**17o** and **17q**) was lower than for unsubstituted ones (**17i** and **17k**). As for methylated/unmethylated pairs **17m/18g** and **17n/17h**, their antiaggregative activities were nearly equal. Positive effect of methyl group was observed only for compounds **17l** and **17p** compared to their unsubstituted analogs **17f** and **17j**, correspondingly. Compound **17p** possessed the highest antiaggregative activity among all compounds synthesized in this study.

## 2.4. Molecular docking

Molecular docking studies have been performed in order to give a microscopic insight into experimentally observed structure–property relationship. FlexX tool from LeadIT package has been used.<sup>20</sup> The X-ray structure of complex of Tirofiban with headpiece of  $\alpha_{IIb}\beta_3$  integrin (2VDM) has been taken from Protein Data Bank.<sup>21</sup> Structure of the pocket has been prepared and four water molecules have been kept in the cavity: two water molecules, coordinated with Mg<sup>2+</sup> metal ion-dependent adhesive site (MIDAS) of the  $\alpha_{IIb}\beta_3$  integrin, and two water molecules coordinated with Asp232 residue of the  $\alpha_{IIb}$ -subunit. Investigated compounds have been docked in the prepared pocket. Because the interaction of ligands of  $\alpha_{IIb}\beta_3$  integrin with  $\alpha$ Asp224 and MIDAS are essential for compound binding and activity, only best poses which follow these requirements have been taken for further analysis.<sup>21</sup>

A reasonable correlation between affinity for  $\alpha_{IIb}\beta_3$  and docking score has been observed ( $R_{\text{Spearman}} = 0.72$ , Fig. 4) for a small subset of compounds for which the activity values were available. Antiaggregative activity poorly correlates with docking score ( $R_{\text{Spearman}} = 0.35$ ). However, exclusion of one outlier (compound **17u**) significantly improves the correlation ( $R_{\text{Spearman}} = 0.52$ , Fig. 4).

Discovered correlations suggest that docking score can be used for elucidation of antiaggregative activity of the compounds. Thus, higher antiaggregative potency of **17f–j** compared to **17a–e** could be explained by the fact that the former form additional H-bonds to OH-group of Tyr190 or to water molecules connected with  $\alpha$ Asp232, whereas the latter cannot form hydrogen bonds to OH-group of  $\alpha$ Tyr190 due to the steric reasons (Fig. 5, compounds **17e** and **17j**).

Compounds **17s,t** resulted from the increase of length of the Asp-mimetic part of **17g,h**, are too big compared to the cavity and, therefore they do not interact efficiently with the protein residues (Fig. 5, compounds **17h** and **17t**). The difference between activity of the compounds **17b** and **17r**, which have the same

topological length between Asp- and Arg-mimetic parts, can be explained by different protein–ligand interaction patterns. The carboxylic acid group of **18r** is buried deeper in the receptor pocket and interacts with  $\beta$ Asn215 as well as with MIDAS, and carbonyl group of 2,3-dihydroisindol-1-one moiety forms H-bond with OH-group of  $\alpha$ Tyr190. On the other hand, carboxylic acid group of **17b** interacts only with MIDAS, whereas carbonyl group of 4-piperidin-4-yl-butanoyl forms H-bond with water molecules coordinated to  $\alpha$ Asp232 (Fig. 5, compounds **17b** and **17r**). Comparison of the docking pose of **17i** with that of **17d** shows that **17i** is characterized by bent conformation whereas **18d** adopts almost linear form. Furthermore, **17i** is characterized by higher docking score value compared to **17d** (38.6 vs 36.4).

## 3. Conclusion

This paper is devoted to rational design of the  $\alpha_{IIb}\beta_3$  integrin antagonists. It has been demonstrated that RGD mimetics containing 2,3-dihydroisindol-1-one can be efficiently used as  $\alpha_{IIb}\beta_3$  antagonists and platelet aggregation inhibitors through appropriate structural modulation. In particular, modification of the Arg surrogates represents an attractive way of optimizing  $\alpha_{IIb}\beta_3$  integrin ligands. It can be summarized that the most preferable structural features for high antiaggregative activity are tetrahydroisoquinoline and  $\beta$ -methyl- $\beta$ -alanine as Arg and Asp bioisosteres, respectively. Combination of these fragments within the given phthalimidine series leads to the best prospective inhibitor of platelet aggregation. The most active compounds identified display favorable biological properties which are promising for further development of antithrombotics.

## 4. Experiment

### 4.1. Chemistry

Melting points are uncorrected. <sup>1</sup>H and <sup>13</sup>C NMR spectra were recorded on Bruker AVANCE-II-400 (at 400 MHz for proton), or Bruker Avance DRX 500 (at 500 MHz for proton and 125 MHz for carbon) spectrometers with chemical shifts in ppm with the internal TMS as a standard. Electron ionization (EI) and fast-atom bombardment (FAB) mass spectra were recorded on a VG Analytical VG 70-70EQ instrument. FAB spectra were performed equipped with an argon primary atom beam, and an *m*-nitrobenzyl alcohol matrix was utilized. High-resolution mass spectra (HRMS) were recorded under FAB conditions. The purity was measured by HPLC

**Table 2**  
Biological properties of RGD mimetics **17**, RGDs peptide and Tirofiban

Compounds	HX	Aa	n	R	IC <sub>50</sub> , (PRP) <sup>a</sup> (μM)	IC <sub>50</sub> , (FITC-Fg/α <sub>IIb</sub> β <sub>3</sub> ) <sup>b</sup> (μM)
<b>17a</b>	TFA		0	H	51 ± 11.0	--
<b>17b</b>	TFA		0	H	66.0 ± 9.0	--
<b>17c</b>	TFA		0	H	24.0 ± 3.0	0.27 ± 0.06
<b>17d</b>	TFA		0	H	3.3 ± 0.5	--
<b>17e</b>	HCl		0	H	120.0 ± 20.0	1.2 ± 0.1
<b>17f</b>	TFA		1	H	7.4 ± 1.4	--
<b>17g</b>	TFA		1	H	5.9 ± 0.6	0.0055 ± 0.009
<b>17h</b>	TFA		1	H	9.6 ± 1.9	0.0068 ± 0.0012
<b>17i</b>	TFA		1	H	0.54 ± 0.06	--
<b>17j</b>	HCl		1	H	1.1 ± 0.1	0.0065 ± 0.0005
<b>17k</b>	TFA		1	H	2.7 ± 0.5	--
<b>17l</b>	TFA		1	CH <sub>3</sub>	2.7 ± 0.6	--
<b>17m</b>	HCl		1	CH <sub>3</sub>	5.4 ± 1.0	0.35 ± 0.03
<b>17n</b>	HCl		1	CH <sub>3</sub>	6.2 ± 1.2	--
<b>17o</b>	HCl		1	CH <sub>3</sub>	3.74 ± 0.51	0.037 ± 0.08
<b>17p</b>	HCl		1	CH <sub>3</sub>	0.086 ± 0.007	0.0065 ± 0.0012
<b>17q</b>	HCl		1	CH <sub>3</sub>	62.0 ± 5.0	--
<b>17r</b>	TFA		2	H	1.8 ± 0.3	--
<b>17s</b>	TFA		2	H	51.0 ± 3.0	--
<b>17t</b>	TFA		2	H	410.0 ± 60.0	--
<b>17u</b>	HCl		2	H	330.0 ± 50.0	--
		RGDS			31.0 ± 2.0	13.0 ± 1.6
		Tirofiban			0.032 ± 0.004	0.0024 ± 0.0004

<sup>a</sup> Concentration required to reduce ADP-induced human platelet aggregation response by 50%. The IC<sub>50</sub> values are expressed as the average of at least three determinations. The average error for the IC<sub>50</sub> determinations was 15%.

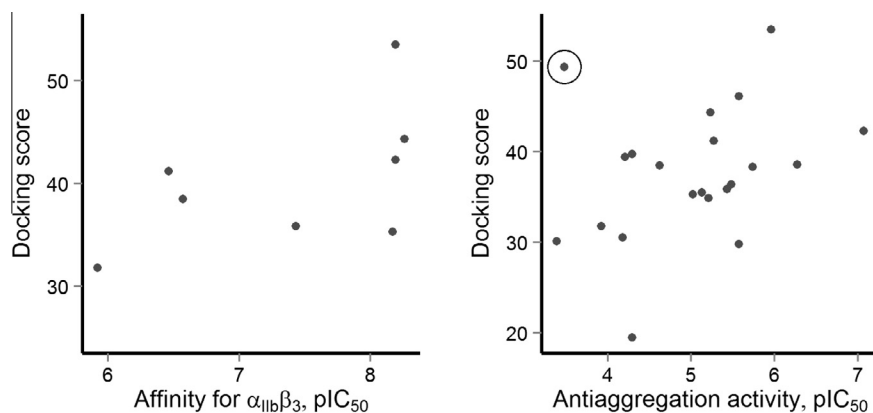
<sup>b</sup> Concentration required to reduce binding of FITC-Fg to α<sub>IIb</sub>β<sub>3</sub> on the suspension of washed human platelets by 50%. The IC<sub>50</sub> values are expressed as the average of at least three determinations. The average error for the IC<sub>50</sub> determinations was 15%.

conducted on an Shimadzu system (System Controller CBM-20A, two pumps LC-8A and Photo-diode Array detector SPD-M20A) using a Hypersil GOLD 3 μm (4.6 × 150 mm) or Hypersil GOLD aQ 3 μm (4.6 × 150 mm) column. The progress of reactions was monitored by TLC (silica gel 60 F254, Merck).

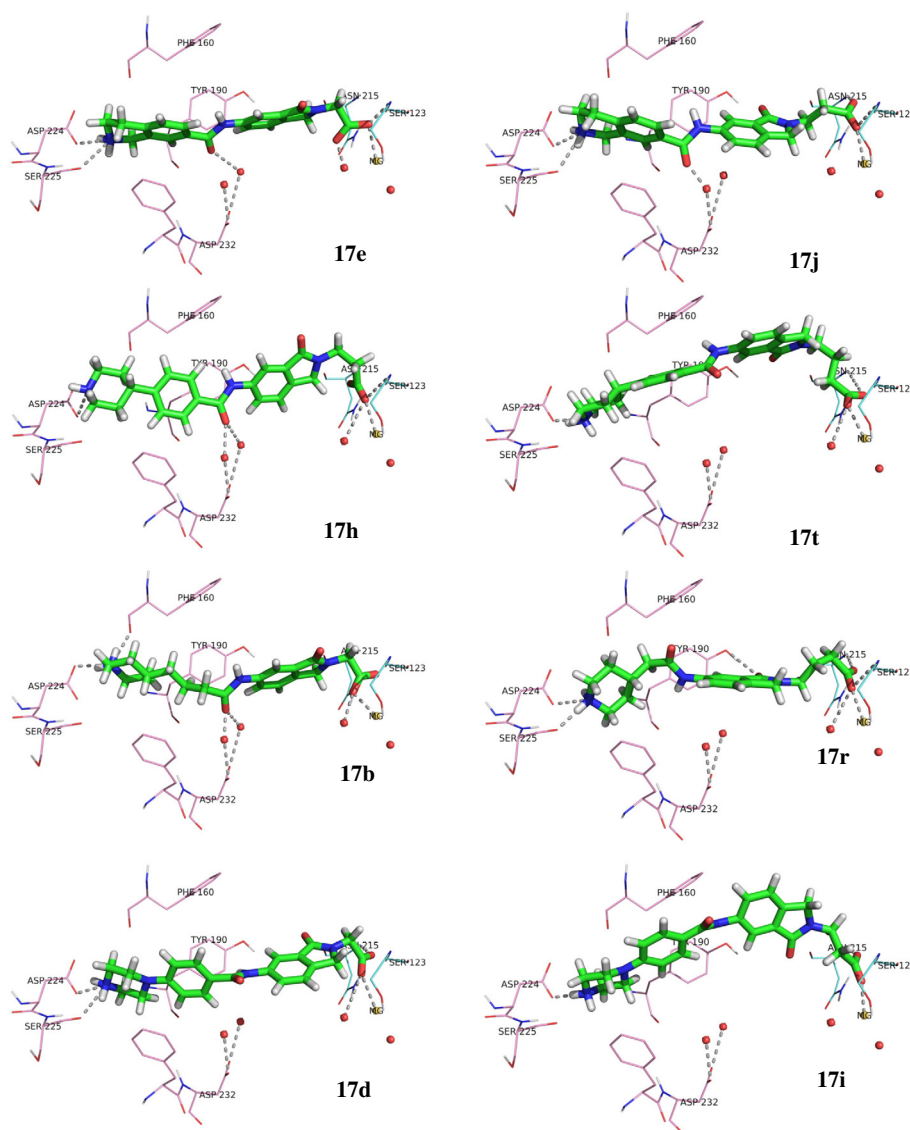
The compound **11a** was prepared from 4-aminobenzoic acid methyl ester by the procedure. The method of synthesis of **11b** from 3-aminobenzoic acid methyl ester not significantly differ from the procedure for compound **11a**. The acids **7**, **8** and **14** have been synthesized using previously published method.

#### 4.1.1. General procedure for a preparation of compound **2**

24.5 g of mossy zinc was amalgamated with 1 g of metallic mercury and 1 M HCl solution (100 ml) in water. The suspension was shaken for 5 min, and then the aqueous layer was discarded. The zinc was covered with 24.5 ml of concentrated HCl and to this was added 0.0488 mol of compound **1**. The mixture was heated to boiling for 0.5 h. At this stage, it is necessary to keep the reaction temperature under control (at rapid foaming, heating was terminated and reaction mixture was shortly cooled externally). After all the compound **1** was dissolved the mixture was boiled under re-



**Figure 4.** Plots of affinity for  $\alpha_{IIb}\beta_3$  of **17c,e,g,h,j,m,o,p** and antiaggregative activity of **17a–u** against docking score.



**Figure 5.** Docking poses and interactions of compounds **17b,d,e,h–j,r,t** inside the  $\alpha_{IIb}\beta_3$  receptor cavity.

flux for 4 h. The hot solution was decanted from undissolved zinc and allowed to cool. A white crystalline solid precipitated and was collected. Recrystallization from water to give pure compound **2**.

**4.1.1.1. (1-Oxo-1,3-dihydroisoindol-2-yl)acetic acid (2a).** Mp = 177–179.5 °C;  $^1\text{H}$  NMR  $\delta$  (500 MHz,  $d_6$ -DMSO) 4.30 (s, 2H), 4.53 (s, 2H), 7.50–7.53 (m, 1H), 7.56–7.58 (m, 2H), 7.68 (d,  $J$  = 7.4 Hz, 1H), 12.93 (s, 1H); MS (EI)  $m/z$ : 191.

**4.1.1.2. 3-(1-Oxo-1,3-dihydroisoindol-2-yl)propionic acid (2b).** Mp = 113–116 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 2.62 (t,  $J$  = 7.0 Hz, 2H), 3.74 (t,  $J$  = 7.0 Hz, 2H), 4.50 (s, 2H), 7.47–7.50 (m, 1H), 7.60 (d,  $J$  = 3.6 Hz, 2H), 7.68 (d,  $J$  = 7.3 Hz, 1H), 12.34 (br s, 1H); MS (EI)  $m/z$ : 205.

**4.1.1.3. 3-(1-Oxo-1,3-dihydroisoindol-2-yl)butyric acid (2c).** Mp = 205.5–206 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.28 (d,  $J$  = 6.6 Hz, 3H), 2.64 (ddd,  $J$  = 31.4, 15.3, 7.3 Hz, 2H), 4.46 (dd,  $J$  = 28.4, 17.6 Hz, 2H), 4.62 (dt,  $J$  = 13.9, 6.6 Hz, 1H), 7.47–7.50 (m, 1H), 7.60–7.61 (m, 2H), 7.67 (d,  $J$  = 7.6 Hz, 1H), 12.26 (br s, 1H); MS (EI)  $m/z$ : 219.

**4.1.1.4. 4-(1-Oxo-1,3-dihydroisoindol-2-yl)butyric acid (2d).** Mp = 125–127 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 1.8 (dt,  $J$  = 14.2, 7.2 Hz, 2H) 2.2 (t,  $J$  = 7.3 Hz, 2H) 3.5 (t,  $J$  = 7.0 Hz, 2H) 7.5 (m,  $J$  = 7.9, 7.9, 4.0 Hz, 1H) 7.6 (m,  $J$  = 3.9 Hz, 2H) 7.7 (d,  $J$  = 7.6 Hz, 1H) 12.1 (br s, 1H); MS (EI)  $m/z$ : 219.

#### 4.1.2. General procedure for a preparation of compound 3

Compound 2 (0.0252 mol) was dissolved in concentrated  $\text{H}_2\text{SO}_4$  (37.7 ml) at 0 °C, and to this solution was added  $\text{HNO}_3$  (12.7 ml,  $d$  = 1.5) at –10 °C. The mixture was stirred for 6 h at –5 °C. The reaction mixture was poured onto crushed ice (200 g). A crystalline solid precipitated and was collected. The filtered product was recrystallized from methanol.

**4.1.2.2. (6-Nitro-1-oxo-1,3-dihydroisoindol-2-yl)acetic acid (3a)** Mp = 222–223 °C;  $^1\text{H NMR } \delta$  (400 MHz, DMSO- $d_6$ ) 4.33 (s, 2H), 4.68 (s, 2H), 7.91 (d,  $J$  = 8.3 Hz, 1H), 8.37 (d,  $J$  = 2.1 Hz, 1H), 8.48 (dd,  $J$  = 2.1 Hz, 8.3 Hz, 1H); MS (EI)  $m/z$ : 219.

**4.1.2.3. 3-(6-Nitro-1-oxo-1,3-dihydroisoindol-2-yl)propionic acid (3b).** Mp = 196.5–197 °C;  $^1\text{H NMR } \delta$  (400 MHz, DMSO- $d_6$ ) 2.65 (t,  $J$  = 7.0 Hz, 2H), 3.77 (t,  $J$  = 7.0 Hz, 2H), 4.67 (s, 2H), 7.90 (d,  $J$  = 8.3 Hz, 1H), 8.33 (d,  $J$  = 2.1 Hz, 1H), 8.45 (dd,  $J$  = 2.1 Hz,  $J$  = 8.3 Hz, 1H); MS (EI)  $m/z$ : 250.

**4.1.2.4. 3-(6-Nitro-1-oxo-1,3-dihydroisoindol-2-yl)butyric acid (3c).** Mp = 194–195 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.31 (d,  $J$  = 6.8 Hz, 3H), 2.67 (ddd,  $J$  = 36.9, 15.5, 7.3 Hz, 2H), 4.58–4.68 (m, 3H), 7.90 (d,  $J$  = 8.3 Hz, 2H), 8.33 (d,  $J$  = 1.7 Hz, 1H), 8.46 (dd,  $J$  = 8.3, 1.7 Hz, 1H), 12.31 (s, 1H); MS (EI)  $m/z$ : 264.

**4.1.2.5. 4-(6-Nitro-1-oxo-1,3-dihydroisoindol-2-yl)butyric acid (3d).** Mp = 153–155 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO)  $\delta$  ppm 1.87 (dt,  $J$  = 13.8, 7.0 Hz, 2H) 2.28 (t,  $J$  = 7.0 Hz, 2H) 3.59 (t,  $J$  = 7.0 Hz, 2H) 4.65 (s, 2H) 7.90 (d,  $J$  = 8.3 Hz, 1H) 8.34 (s, 1H) 8.46 (d,  $J$  = 8.0 Hz, 1H) 12.07 (s, 1H); MS (EI)  $m/z$ : 264.

#### 4.1.3. General procedure for a preparation of compound 4

Compound 3 (0.0169 mol) was dissolved in methanol (50 ml), and to this solution was added concentrated  $\text{H}_2\text{SO}_4$  (0.5 ml). The reaction solution was refluxed for 3 h. The solvent was removed via evaporation in vacuo, and the residue was triturated with water. A solid precipitated and was collected. The filtered product was recrystallized from methanol.

**4.1.3.1. (6-Nitro-1-oxo-1,3-dihydroisoindol-2-yl)acetic acid methyl ester (4a).** Mp = 113–115 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 3.70 (s, 3H), 4.46 (s, 2H), 4.70 (s, 2H), 7.94 (d,  $J$  = 8.3 Hz, 1H), 8.39 (d,  $J$  = 2.1 Hz, 1H), 8.50 (dd,  $J$  = 2.1 Hz, 8.3 Hz, 1H); MS (EI)  $m/z$ : 250.

**4.1.3.2. 3-(6-Nitro-1-oxo-1,3-dihydroisoindol-2-yl)propionic acid methyl ester (4b).** Mp = 150–152 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 2.74 (t,  $J$  = 7.0 Hz, 2H), 3.62 (s, 3H), 3.80 (t,  $J$  = 7.0 Hz, 2H), 4.67

(s, 2H), 7.90 (d,  $J$  = 8.3 Hz, 1H), 8.33 (d,  $J$  = 2.0 Hz, 1H), 8.46 (dd,  $J$  = 8.3, 2.0 Hz, 1H); MS (EI)  $m/z$ : 264.

**4.1.3.3. 3-(6-Nitro-1-oxo-1,3-dihydroisoindol-2-yl)butyric acid methyl ester (4c).** Mp = 125–126 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.32 (d,  $J$  = 6.8 Hz, 3H) 2.76 (ddd,  $J$  = 35.4, 15.2, 7.3 Hz, 2H) 3.57 (s, 3H) 4.59–4.67 (m, 3H) 7.91 (d,  $J$  = 8.3 Hz, 1H) 8.34 (s, 1H) 8.46 (dd,  $J$  = 8.3, 2.2 Hz, 1H); MS (EI)  $m/z$ : 278.

**4.1.3.4. 4-(6-Nitro-1-oxo-1,3-dihydroisoindol-2-yl)butyric acid methyl ester (4d).** Mp = 86–87 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.90 (dt,  $J$  = 13.8, 6.7 Hz, 2H), 2.37 (t,  $J$  = 7.0 Hz, 2H), 3.55 (s, 3H), 3.59 (t,  $J$  = 6.7 Hz, 2H), 4.65 (s, 2H), 7.90 (d,  $J$  = 8.3 Hz, 1H), 8.34 (s, 1H), 8.46 (d,  $J$  = 8.3 Hz, 1H); MS (EI)  $m/z$ : 278.

#### 4.1.4. General procedure for a preparation of compound 5

The nitro compound 4 (0.015 mol) dissolved in methanol (100 ml) was subjected to catalytic hydrogenation at room temperature for 7 h in the presence of 3% palladium on carbon (1 g). The filtered solution was then evaporated in vacuo to give compound 5.

**4.1.4.1. (6-Amino-1-oxo-1,3-dihydroisoindol-2-yl)acetic acid methyl ester (5a).** Mp = 142–143 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 3.68 (s, 3H), 4.32 (s, 2H), 4.34 (s, 2H), 5.38 (s, 2H), 6.83 (dd,  $J$  = 2.1 Hz, 8.1 Hz, 1H), 6.87 (d,  $J$  = 2.1 Hz, 1H), 7.23 (d,  $J$  = 8.1 Hz, 1H); MS (EI)  $m/z$ : 220.

**4.1.4.2. 3-(6-Amino-1-oxo-1,3-dihydroisoindol-2-yl)propionic acid methyl ester (5b).** Mp = 147–148.5 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 2.66 (t,  $J$  = 7.1 Hz, 2H), 3.61 (s, 3H), 3.71 (t,  $J$  = 7.1 Hz, 2H), 4.28 (s, 2H), 5.30 (s, 2H), 6.79 (dd,  $J$  = 2.1 Hz,  $J$  = 8.0 Hz, 1H), 6.82 (d,  $J$  = 2.1 Hz, 1H), 7.19 (d,  $J$  = 8.0 Hz, 1H); MS (EI)  $m/z$ : 234.

**4.1.4.3. 3-(6-Amino-1-oxo-1,3-dihydroisoindol-2-yl)butyric acid methyl ester (5c).** Mp = 86–86.5 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.25 (d,  $J$  = 6.8 Hz, 3H), 2.68 (ddd,  $J$  = 27.0, 15.1, 7.3 Hz, 2H), 3.56 (s, 3H), 4.23 (dd,  $J$  = 24.5, 16.4 Hz, 2H), 4.58 (dt,  $J$  = 13.9, 6.8 Hz, 1H), 5.30 (s, 2H), 6.79 (dd,  $J$  = 8.1, 1.5 Hz, 1H), 6.82 (s, 1H), 7.20 (d,  $J$  = 8.1 Hz, 1H); MS (EI)  $m/z$ : 248.

**4.1.4.4. 4-(6-Amino-1-oxo-1,3-dihydroisoindol-2-yl)butyric acid methyl ester (5d).** Mp = 97–98 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 1.82 (dt,  $J$  = 14.2, 7.2 Hz, 2H), 2.30 (t,  $J$  = 7.2 Hz, 2H), 3.46 (t,  $J$  = 7.2 Hz, 2H), 3.52 (s, 3H), 4.23 (s, 2H), 5.28 (s, 2H), 6.76 (dd,  $J$  = 8.1, 2.1 Hz, 1H), 6.81 (d,  $J$  = 2.1 Hz, 1H), 7.17 (d,  $J$  = 8.1 Hz, 1H); MS (EI)  $m/z$ : 248.

#### 4.1.5. 1-Boc-4-piperidineacetic acid (6)

1-Boc-piperidine-4-carboxylic acid (2.29 g, 10 mmol) was dissolved in anhydrous THF (25 ml) under argon. The solution was cooled to –15 °C, and triethylamine (1.4 ml, 10 mmol), and then isobutyl chloroformate (1.3 ml, 10 mmol), were added. The reactor was protected from light. After 30 min, an ethereal solution of diazomethane was added. The mixture was allowed to warm to room temperature for 2 h without stirring and was left overnight. The reaction solution was diluted with chloroform (final volume –200 ml). The excess of diazomethane was destroyed by addition of few drops of acetic acid, and saturated aqueous solution of  $\text{NaHCO}_3$  (10 ml) was added carefully. The aqueous layer was separated, and the organic layer was washed with saturated aqueous  $\text{NaCl}$  (10 ml). The organic layer was dried over anhydrous  $\text{Na}_2\text{SO}_4$ , filtered, and the solvent was evaporated in vacuo to dryness. The crude product, the diazoketone, was used directly in the next step.

The diazoketone was dissolved in THF (40 ml), and water (10 ml) and silver oxide (0.3 g) were added, and the mixture (protected from light) was stirred for 2–3 h at 45 °C (the reaction was monitored by TLC). Silver oxide was removed by centrifugation and the supernatant was evaporated in vacuo to dryness. The residue was dissolved in chloroform (100 ml) and the solution was washed with 5% aqueous solution of NaOH (30 ml). The aqueous layer was separated and pH of aqueous solution was brought to three, and the product was extracted with chloroform. The organic layer was washed with saturated aqueous NaCl (50 ml). The organic layer was dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>, filtered, and chloroform was evaporated to yield oily residue of **6**. After a period of few days, oil was converted into solid waxy substance.

#### 4.1.6. 2-Acetyl-1,2,3,4-tetrahydroisoquinoline-7-carboxylic (**13**)

Compound **12** (2.5 g, 0.117 mol) was dissolved in anhydrous CH<sub>2</sub>Cl<sub>2</sub> (50 ml). The solution was cooled to –10 °C, and with intensive stirring oxalyl chloride (5 ml, 0.117 mol) was added. AlCl<sub>3</sub> (2.85, 0.643 mol) was added in small portions to the stirred solution with temperature kept below –9 °C. The mixture was stirred for 1 h at –10 °C. Then additional AlCl<sub>3</sub> (3.81, 0.643 mol) was added, and the mixture was stirred for another 2 h at –10 °C. The reaction mixture was allowed to stand overnight at room temperature and then poured onto ice (100 g). The aqueous phase was extracted with CH<sub>2</sub>Cl<sub>2</sub> (100 ml). The organic phase was dried with anhydrous Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent was evaporated in vacuo. The residue was dissolved in acid, with stirring and heating (40 °C). To the warm solution, ice was added. pH of the mixture was brought to three and the product was extracted with chloroform. The chloroform phase was dried with anhydrous Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent was evaporated in vacuo to dryness. The resulting **13** obtained as a solid residue was used in subsequent transformations without further purification. Cream powder, yield 53%. Mp = 183.5–184.5 °C.

#### 4.1.7. General procedure for a preparation of compound **15**

The 10 mmol of acid (**6–8**, **11a,b** or **14**) was dissolved in anhydrous acetonitrile (25 ml). The solution was cooled to –5 °C, and triethylamine (1.4 ml, 10 mmol), and then HBTU (3.79 g, 10 mmol) or HATU (3.8 g, 10 mmol), were added. The mixture was stirred for 1 h at –5 °C and then 10 mmol of amine **5** was added. The reaction mixture was allowed to stand overnight at room temperature. The residual amount of the activated ether (Bt- or At-ether of starting acid) was destroyed by addition of few drops of *N,N*-dimethylpropane-1,3-diamine, and the solvent was evaporated in vacuo to dryness. The residue was dissolved in 100 ml of chloroform. The solution was washed with water (40 mL), aqueous solution of 1 M HCl (40 ml) and 5% aqueous solution of NaHCO<sub>3</sub> (40 ml). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered off, and the solvent was evaporated in vacuo to dryness. The resulting residue was triturated with warm hexane (20 ml), and the precipitate was collected by filtration and dried.

**4.1.7.1. {6-[2-(1-Boc-piperidin-4-yl)acetyl]amino}-1-oxo-1,3-dihydroisoindol-2-yl}acetic acid methyl ester (**15a**). Mp = 193–195 °C; <sup>1</sup>H NMR δ (400 MHz, **d**<sub>6</sub>-DMSO) 1.07 (ddd, *J* = 24.3, 12.1, 3.9 Hz, 2H), 1.38 (s, 9H), 1.65 (d, *J* = 11.0 Hz, 2H), 1.94 (m, *J* = 3.7 Hz, 1H), 2.27 (d, *J* = 6.8 Hz, 2H), 2.64–2.86 (m, 2H), 3.67 (s, 3H), 3.90 (d, *J* = 11.7 Hz, 2H), 4.37 (s, 2H), 4.45 (s, 2H), 7.52 (d, *J* = 8.1 Hz, 1H), 7.68 (dd, *J* = 8.3, 2.0 Hz, 1H), 8.07 (d, *J* = 1.2 Hz, 1H), 10.13 (s, 1H); HRMS (FAB) *m/z* calcd for C<sub>23</sub>H<sub>33</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 446.5280, found: 446.5273.**

**4.1.7.2. {6-[4-(1-Boc-piperidin-4-yl)butyrylamino]-1-oxo-1,3-dihydroisoindol-2-yl}acetic acid methyl ester (**15b**). Mp = 149–150 °C; <sup>1</sup>H NMR δ (400 MHz, **d**<sub>6</sub>-DMSO) 0.95 (ddd, *J* = 24.3, 12.2,**

4.0 Hz, 2H), 1.24 (dd, *J* = 14.9, 6.6 Hz, 2H), 1.39 (s, 9H), 1.41–1.50 (m, 1H), 1.59–1.67 (m, 4H), 2.33 (t, *J* = 7.3 Hz, 2H), 2.60–2.75 (m, 2H), 3.69 (s, 3H), 3.92 (d, *J* = 11.0 Hz, 2H), 4.39 (s, 2H), 4.46 (s, 2H), 7.54 (d, *J* = 8.2 Hz, 1H), 7.72 (dd, *J* = 8.2, 1.8 Hz, 1H), 8.08 (d, *J* = 1.8 Hz, 1H), 10.13 (s, 1H); HRMS (FAB) *m/z* calcd for C<sub>25</sub>H<sub>36</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 474.5822, found: 474.5826.

**4.1.7.3. {6-[4-(1-Boc-piperidin-4-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}acetic acid methyl ester (**15c**). Mp = 189–191 °C; <sup>1</sup>H NMR δ (400 MHz, **d**<sub>6</sub>-DMSO) 1.43 (s, 9H), 1.55 (ddd, *J* = 24.9, 12.4, 3.8 Hz, 2H), 1.79 (d, *J* = 12.7 Hz, 2H), 2.75–2.90 (m, 3H), 4.10 (d, *J* = 10.8 Hz, 2H), 4.29 (s, 2H), 4.50 (s, 2H), 7.44 (d, *J* = 7.6 Hz, 2H), 7.59 (d, *J* = 8.3 Hz, 1H), 7.93 (d, *J* = 7.3 Hz, 2H), 7.97 (d, *J* = 8.3 Hz, 1H), 8.22 (s, 1H), 10.41 (s, 1H); HRMS (FAB) *m/z* calcd for C<sub>28</sub>H<sub>34</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 508.5997, found: 508.6002.**

**4.1.7.4. {6-[4-(4-Boc-piperazin-1-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}acetic acid methyl ester (**15d**). Mp = 219–220.5 °C; <sup>1</sup>H NMR δ (500 MHz, **d**<sub>6</sub>-DMSO) 1.44 (s, 9H), 3.31 (t, *J* = 4.6 Hz, 4H), 3.48 (t, *J* = 4.6 Hz, 4H), 3.70 (s, 3H), 4.41 (s, 2H), 4.50 (s, 2H), 7.05 (d, *J* = 8.7 Hz, 2H), 7.58 (d, *J* = 8.2 Hz, 1H), 7.93 (d, *J* = 8.7 Hz, 2H), 7.98 (dd, *J* = 8.2, 1.0 Hz, 1H), 8.23 (d, *J* = 1.0, 1H), 10.18 (s, 1H); HRMS (FAB) *m/z* calcd for C<sub>27</sub>H<sub>33</sub>N<sub>4</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 509.5873, found: 509.5880.**

**4.1.7.5. {6-[(2-Boc-1,2,3,4-tetrahydroisoquinoline-7-carbonyl)amino]-1-oxo-1,3-dihydroisoindol-2-yl}acetic acid methyl ester (**15e**). Mp = 134–137 °C; <sup>1</sup>H NMR δ (400 MHz, **d**<sub>6</sub>-DMSO) 1.45 (s, 9H), 2.87 (t, *J* = 5.8 Hz, 2H), 3.60 (t, *J* = 5.8 Hz, 2H), 3.70 (s, 3H), 4.42 (s, 2H), 4.51 (s, 2H), 4.61 (s, 2H), 7.35 (d, *J* = 8.1 Hz, 1H), 7.61 (d, *J* = 8.3 Hz, 1H), 7.81 (d, *J* = 8.1 Hz, 1H), 7.83 (s, 1H), 8.00 (dd, *J* = 8.2, 1.7 Hz, 1H), 8.24 (d, *J* = 1.7 Hz, 1H), 10.41 (s, 1H); HRMS (FAB) *m/z* calcd for C<sub>26</sub>H<sub>30</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 480.5455, found: 480.5447.**

**4.1.7.6. 3-{6-[2-(1-Boc-piperidin-4-yl)acetyl]amino}-1-oxo-1,3-dihydroisoindol-2-yl}propionic acid methyl ester (**15f**). Mp = 168–170 °C; <sup>1</sup>H NMR δ (500 MHz, **d**<sub>6</sub>-DMSO) 1.09 (ddd, *J* = 23.9, 11.7, 3.9 Hz, 2H), 1.40 (s, 9H), 1.67 (d, *J* = 11.7 Hz, 2H), 1.94–1.98 (m, 1H), 2.28 (d, *J* = 6.7 Hz, 2H), 2.70 (t, *J* = 6.7 Hz, 2H), 2.74–2.80 (m, 2H), 3.61 (s, 3H), 3.76 (t, *J* = 6.7 Hz, 2H), 3.92 (d, *J* = 9.1 Hz, 2H), 4.43 (s, 5H), 7.50 (d, *J* = 8.3 Hz, 1H), 7.68 (d, *J* = 8.3 Hz, 1H), 8.04 (s, 1H), 10.12 (s, 1H); HRMS (FAB) *m/z* calcd for C<sub>24</sub>H<sub>34</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 460.5551, found: 460.5560.**

**4.1.7.7. 3-{6-[4-(1-Boc-piperidin-4-yl)butyrylamino]-1-oxo-1,3-dihydroisoindol-2-yl}propionic acid methyl ester (**15g**). Mp = 171–172 °C; <sup>1</sup>H NMR δ (400 MHz, **d**<sub>6</sub>-DMSO) 0.94 (ddd, *J* = 24.2, 12.2, 4.2 Hz, 2H), 1.23 (dd, *J* = 15.2, 6.8 Hz, 2H), 1.38 (s, 9H), 1.40–1.45 (m, 1H), 1.58–1.65 (m, 4H), 2.31 (t, *J* = 7.5 Hz, 2H), 2.58–2.65 (m, 2H), 2.68 (t, *J* = 7.0 Hz, 2H), 3.60 (s, 3H), 3.74 (t, *J* = 7.0 Hz, 2H), 3.91 (d, *J* = 12.5 Hz, 2H), 4.41 (s, 2H), 7.49 (d, *J* = 8.2 Hz, 1H), 7.67 (dd, *J* = 8.2, 1.7 Hz, 1H), 8.02 (d, *J* = 1.7 Hz, 1H), 10.08 (s, 1H); HRMS (FAB) *m/z* calcd for C<sub>26</sub>H<sub>38</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 488.6093, found: 488.6087.**

**4.1.7.8. 3-{6-[4-(1-Boc-piperidin-4-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}propionic acid methyl ester (**15h**). Mp = 199.5–200.5 °C; <sup>1</sup>H NMR δ (400 MHz, **d**<sub>6</sub>-DMSO) 1.41 (s, 9H), 1.52 (ddd, *J* = 25.2, 12.6, 4.0 Hz, 2H), 1.77 (d, *J* = 12.6 Hz, 2H), 2.69 (t, *J* = 7.0 Hz, 2H), 2.73–2.90 (m, 3H), 3.59 (s, 3H), 3.75 (t, *J* = 7.0 Hz, 2H), 4.08 (d, *J* = 11.5 Hz, 2H), 4.44 (s, 2H), 7.41 (d, *J* = 8.1 Hz, 2H), 7.54 (d, *J* = 8.3 Hz, 1H), 7.90 (d, *J* = 8.1 Hz, 2H), 7.92 (dd, *J* = 8.3, 1.7 Hz, 1H), 8.16 (d, *J* = 1.7 Hz, 1H), 10.36 (s, 1H); HRMS (FAB) *m/z* calcd for C<sub>29</sub>H<sub>36</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 522.6268, found: 522.6260.**

**4.1.7.9. 3-{6-[4-(4-Boc-piperazin-1-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}propionic acid methyl ester (15i).** Mp = 185–187 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 1.43 (s, 9H), 2.71 (t,  $J$  = 6.9 Hz, 2H), 3.30 (s, 4H), 3.48 (s, 4H), 3.61 (s, 3H), 3.77 (t,  $J$  = 6.9 Hz, 2H), 4.45 (s, 2H), 7.04 (d,  $J$  = 9.0 Hz, 2H), 7.54 (d,  $J$  = 8.7 Hz, 1H), 7.91 (d,  $J$  = 9.0 Hz, 2H), 7.95 (dd,  $J$  = 8.7, 1.7 Hz, 1H), 8.18 (d,  $J$  = 1.7 Hz, 1H), 10.15 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{28}\text{H}_{35}\text{N}_4\text{O}_6$   $[\text{M}+\text{H}]^+$ : 523.6144, found: 523.6137.

**4.1.7.10. 3-{6-[(2-Boc-1,2,3,4-tetrahydroisoquinoline-7-carbonyl)amino]-1-oxo-1,3-dihydroisoindol-2-yl}propionic acid methyl ester (15j).** Mp = 102–103 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 1.45 (s, 9H), 2.72 (t,  $J$  = 7.5 Hz, 2H), 2.87 (t,  $J$  = 5.9 Hz, 2H), 3.60 (t,  $J$  = 5.6 Hz, 2H), 3.62 (s, 3H), 3.78 (t,  $J$  = 7.0 Hz, 2H), 4.47 (s, 2H), 4.61 (s, 2H), 7.34 (d,  $J$  = 8.1 Hz, 1H), 7.57 (d,  $J$  = 8.1 Hz, 1H), 7.80 (d,  $J$  = 8.1 Hz, 1H), 7.82 (s, 1H), 7.95 (dd,  $J$  = 8.1, 1.5 Hz, 1H), 8.19 (d,  $J$  = 1.5 Hz, 1H), 10.37 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{27}\text{H}_{32}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 494.5726, found: 494.5730.

**4.1.7.11. 3-{6-[3-(4-Boc-piperazin-1-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}propionic acid methyl ester (15k).** Mp = 181–183 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.44 (s, 9H), 2.72 (t,  $J$  = 6.5 Hz, 2H), 3.21 (s, 4H), 3.50 (s, 4H), 3.62 (s, 3H), 3.78 (t,  $J$  = 6.1 Hz, 2H), 4.47 (s, 2H), 7.19 (d,  $J$  = 6.2 Hz, 1H), 7.38–7.44 (m, 2H), 7.50 (s, 1H), 7.57 (d,  $J$  = 7.8 Hz, 1H), 7.95 (d,  $J$  = 8.0 Hz, 1H), 8.17 (m, 1H), 10.35 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{28}\text{H}_{35}\text{N}_4\text{O}_6$   $[\text{M}+\text{H}]^+$ : 523.6144, found: 523.6146.

**4.1.7.12. 3-{6-[2-(1-Boc-piperidin-4-yl)acetyl-amino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid methyl ester (15l).** Mp = 130–132 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.09 (ddd,  $J$  = 22.5, 11.7, 3.1 Hz, 2H), 1.28 (d,  $J$  = 5.7 Hz, 3H), 1.40 (s, 9H), 1.67 (d,  $J$  = 11.7 Hz, 2H), 1.91–2.00 (m, 1H), 2.27 (d,  $J$  = 6.2 Hz, 2H), 2.65–2.82 (m, 4H), 3.56 (s, 3H), 3.93 (d,  $J$  = 8.3 Hz, 2H), 4.38 (dd,  $J$  = 22.8, 17.1 Hz, 2H), 4.60 (m, 1H), 7.51 (d,  $J$  = 8.1 Hz, 1H), 7.69 (d,  $J$  = 8.1 Hz, 1H), 8.03 (s, 1H), 10.10 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{25}\text{H}_{36}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 474.5822, found: 474.5823.

**4.1.7.13. 3-{6-[4-(1-Boc-piperidin-4-yl)butyrylamino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid methyl ester (15m).** Mp = 133–134 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 0.95 (ddd,  $J$  = 24.0, 12.2, 3.9 Hz, 2H), 1.22–1.25 (m, 2H), 1.28 (d,  $J$  = 6.6 Hz, 3H), 1.39 (s, 9H), 1.61–1.66 (m, 4H), 2.32 (t,  $J$  = 7.3 Hz, 2H), 2.65–2.77 (m, 4H), 3.56 (s, 3H), 3.92 (d,  $J$  = 9.3 Hz, 2H), 4.38 (dd,  $J$  = 24.0, 17.1 Hz, 2H), 4.61 (td,  $J$  = 13.7, 6.8 Hz, 1H), 7.50 (d,  $J$  = 8.3 Hz, 1H), 7.69 (d,  $J$  = 8.3 Hz, 1H), 8.02 (s, 1H), 10.07 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{27}\text{H}_{40}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 502.6364, found: 502.6367.

**4.1.7.14. 3-{6-[4-(1-Boc-piperidin-4-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid methyl ester (15n).**  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.30 (d,  $J$  = 6.8 Hz, 3H), 1.43 (s, 9H), 1.55 (ddd,  $J$  = 24.5, 11.7, 2.9 Hz, 2H), 1.79 (d,  $J$  = 11.7 Hz, 2H), 2.68–2.85 (m, 5H), 3.57 (s, 3H), 4.11 (d,  $J$  = 7.6 Hz, 2H), 4.42 (dd,  $J$  = 24.0, 17.6 Hz, 2H), 4.63 (dt,  $J$  = 13.9, 6.7 Hz, 1H), 7.43 (d,  $J$  = 8.1 Hz, 2H), 7.57 (d,  $J$  = 8.1 Hz, 1H), 7.92–7.96 (m, 3H), 8.18 (s, 1H), 10.38 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{30}\text{H}_{38}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 536.6539, found: 536.6535

**4.1.7.15. 3-{6-[4-(1-Boc-piperazin-4-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid methyl ester (15o).** Mp = 95–97 °C (decomposes);  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.29 (d,  $J$  = 6.8 Hz, 3H), 1.44 (s, 9H), 2.68–2.78 (m, 2H), 3.30 (t,  $J$  = 4.0 Hz, 4H), 3.48 (t,  $J$  = 4.0 Hz, 4H), 3.57 (s, 3H), 4.41 (dd,  $J$  = 24.2, 17.4 Hz, 2H), 4.63 (td,  $J$  = 13.9, 6.5 Hz, 1H), 7.05 (d,  $J$  = 9.0 Hz, 2H), 7.55 (d,  $J$  = 8.2 Hz, 1H), 7.92 (d,  $J$  = 9.0 Hz, 2H),

7.95 (dd,  $J$  = 8.2, 1.8 Hz, 1H), 8.17 (d,  $J$  = 1.8 Hz, 1H), 10.15 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{29}\text{H}_{37}\text{N}_4\text{O}_6$   $[\text{M}+\text{H}]^+$ : 537.6414, found: 537.66417.

**4.1.7.16. 3-{6-[(2-Boc-1,2,3,4-tetrahydroisoquinoline-7-carbonyl)amino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid methyl ester (15p).**  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO +  $\text{CCl}_4$ ) 1.29 (d,  $J$  = 6.6 Hz, 3H), 1.45 (s, 9H), 2.67–2.77 (m, 2H), 2.86 (t,  $J$  = 5.0 Hz, 2H), 3.57 (s, 3H), 3.58–3.60 (m, 2H), 4.40 (dd,  $J$  = 22.0, 17.4 Hz, 2H), 4.59–4.66 (m, 3H), 7.33 (d,  $J$  = 7.5 Hz, 1H), 7.57 (d,  $J$  = 8.3 Hz, 1H), 7.79–7.82 (m, 2H), 7.95 (d,  $J$  = 7.9 Hz, 1H), 8.18 (s, 1H), 10.36 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{28}\text{H}_{34}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 508.5997, found: 508.5995.

**4.1.7.17. 3-{6-[3-(4-Boc-piperazin-1-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid methyl ester (15q).** Mp = 142–143.5 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 1.30 (d,  $J$  = 6.7 Hz, 3H), 1.44 (s, 9H), 2.68–2.79 (m, 2H), 3.24 (s, 4H), 3.53 (s, 4H), 3.57 (s, 3H), 4.43 (dd,  $J$  = 24.4, 17.1 Hz, 2H), 4.63 (m, 1H), 7.25 (d,  $J$  = 7.8 Hz, 1H), 7.41 (t,  $J$  = 7.8 Hz, 1H), 7.47 (d,  $J$  = 7.5 Hz, 1H), 7.58 (d,  $J$  = 7.0 Hz, 2H), 7.98 (d,  $J$  = 8.0 Hz, 1H), 8.17 (s, 1H), 10.40 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{29}\text{H}_{37}\text{N}_4\text{O}_6$   $[\text{M}+\text{H}]^+$ : 537.6414, found: 537.66416.

**4.1.7.18. 4-{6-[2-(1-Boc-piperidin-4-yl)acetyl-amino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid methyl ester (15r).** Mp = 159–160 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.06–1.12 (m, 2H), 1.40 (s, 9H), 1.67 (d,  $J$  = 11.7 Hz, 2H), 1.87 (t,  $J$  = 6.8, 2H), 1.92–1.99 (m, 1H), 2.28 (d,  $J$  = 6.5, 2H), 2.34 (t,  $J$  = 6.8, 2H), 2.65–2.80 (m, 2H), 3.52–3.56 (m, 5H), 3.92 (d,  $J$  = 8.6 Hz, 2H), 4.40 (s, 2H), 7.50 (d,  $J$  = 7.9 Hz, 1H), 7.68 (d,  $J$  = 7.9 Hz, 1H), 8.03 (s, 1H), 10.09 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{25}\text{H}_{36}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 474.5822, found: 474.5819.

**4.1.7.19. 4-{6-[4-(1-Boc-piperidin-4-yl)butyrylamino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid methyl ester (15s).** Mp = 151–153 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 0.94 (ddd,  $J$  = 24.4, 12.2, 4.0 Hz, 2H), 1.23 (dd,  $J$  = 15.7, 7.3 Hz, 2H), 1.38 (s, 9H), 1.41–1.43 (m, 1H), 1.60–1.66 (m, 4H), 1.86 (dt,  $J$  = 14.2, 7.1 Hz, 2H), 2.32 (m,  $J$  = 7.1 Hz, 4H), 2.59–2.73 (m, 2H), 3.52 (t,  $J$  = 7.1 Hz, 2H), 3.53 (s, 3H), 3.91 (d,  $J$  = 11.7 Hz, 2H), 4.39 (s, 2H), 7.49 (d,  $J$  = 8.1 Hz, 1H), 7.68 (dd,  $J$  = 8.1, 1.6 Hz, 1H), 8.2 (d,  $J$  = 1.6 Hz, 1H), 10.07 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{27}\text{H}_{40}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 502.6364, found: 502.6363.

**4.1.7.20. 4-{6-[4-(1-Boc-piperidin-4-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid methyl ester (15t).** Mp = 166–167 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 1.4 (s, 9H) 1.5 (ddd,  $J$  = 22.3, 12.0, 2.4 Hz, 2H) 1.8 (d,  $J$  = 12.2 Hz, 2H) 1.9 (dt,  $J$  = 13.4, 6.6 Hz, 2H) 2.3 (t,  $J$  = 6.6 Hz, 2H) 2.8 (m, 3H) 3.5 (m, 5H) 4.1 (d,  $J$  = 9.0 Hz, 2H) 4.4 (s, 2H) 7.4 (d,  $J$  = 7.3 Hz, 2H) 7.5 (d,  $J$  = 8.1 Hz, 1H) 7.9 (m,  $J$  = 6.7, 6.7 Hz, 3H) 8.2 (s, 1H) 10.4 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{30}\text{H}_{38}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 536.6539, found: 536.6537.

**4.1.7.21. 4-{6-[(2-Boc-1,2,3,4-tetrahydroisoquinoline-7-carbonyl)amino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid methyl ester (15u).** Mp = 142–145 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 1.44 (s, 9H), 1.88 (dt,  $J$  = 13.8, 6.6 Hz, 4H), 2.35 (t,  $J$  = 7.1 Hz, 2H), 2.86 (t,  $J$  = 4.9, 2H), 3.54 (s, 3H), 3.59 (t,  $J$  = 4.9, 4H), 4.44 (s, 2H), 4.60 (s, 2H), 7.33 (d,  $J$  = 7.6 Hz, 1H), 7.56 (d,  $J$  = 8.1 Hz, 2H), 7.78–7.81 (m, 2H), 7.94 (d,  $J$  = 8.1 Hz, 1H), 8.18 (s, 1H), 10.37 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{28}\text{H}_{34}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 508.5997, found: 508.5995.

#### 4.1.8. General procedure for a preparation of compound 16

The compound **15** (10 mmol) was dissolved in methanol (25 ml), and to this solution was added 1 M NaOH aqueous solu-

tion (5 ml). The reaction mixture was left at room temperature overnight. Then chloroform (50 ml) was added, and pH of the mixture was brought to three with intensive stirring. When the product was not soluble in chloroform layer, an suspension was formed, the product was filtered off. The precipitate was washed (on the filter) with water, chloroform and ether, and dried in air. Thus obtained product **16** was not need to be further purified. When the reaction product was soluble in chloroform, the chloroform phase was collected and washed with water (20 ml). The chloroform phase was dried with anhydrous Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent was evaporated in vacuo to dryness. The resulting residue was triturated with ether (20 ml) and the precipitate was collected by filtration and dried in air.

**4.1.8.1. {6-[2-(1-Boc-piperidin-4-yl)acetylamino]-1-oxo-1,3-dihydroisoindol-2-yl}acetic acid (16a).** Mp = 198–201 °C; <sup>1</sup>H NMR δ (400 MHz, d<sub>6</sub>-DMSO) 1.07 (ddd, *J* = 23.6, 11.9, 3.2 Hz, 2H), 1.38 (s, 9H), 1.65 (d, *J* = 12.2 Hz, 2H), 1.91–1.97 (m, 1H), 2.26 (d, *J* = 6.8 Hz, 2H), 2.65–2.80 (m, 2H), 3.91 (d, *J* = 11.0 Hz, 2H), 4.25 (s, 2H), 4.44 (s, 2H), 7.51 (d, *J* = 8.3 Hz, 1H), 7.68 (d, *J* = 7.8 Hz, 1H), 8.06 (s, 1H), 10.13 (s, 1H), 12.86 (br s, 1H); HRMS (FAB) *m/z* calcd for C<sub>22</sub>H<sub>30</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 432.5009, found: 432.5005.

**4.1.8.2. {6-[4-(1-Boc-piperidin-4-yl)butyrylamino]-1-oxo-1,3-dihydroisoindol-2-yl}acetic acid (16b).** Mp = 165.5–167 °C; <sup>1</sup>H NMR δ (400 MHz, d<sub>6</sub>-DMSO) 0.95 (ddd, *J* = 24.3, 12.3, 4.0 Hz, 2H), 1.22–1.27 (m, 2H), 1.39 (s, 9H), 1.42–1.45 (m, 1H), 1.59–1.68 (m, 4H), 2.33 (t, *J* = 7.5 Hz, 2H), 2.60–2.75 (m, 2H), 3.93 (d, *J* = 11.2 Hz, 2H), 4.27 (s, 2H), 4.46 (s, 2H), 7.53 (d, *J* = 8.1 Hz, 1H), 7.71 (dd, *J* = 8.3, 1.7 Hz, 1H), 8.08 (d, *J* = 1.5 Hz, 1H), 10.12 (s, 1H), 12.96 (br s, 1H); HRMS (FAB) *m/z* calcd for C<sub>24</sub>H<sub>32</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 460.5551, found: 460.5559.

**4.1.8.3. {6-[4-(1-Boc-piperidin-4-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}acetic acid (16c).** Mp = 223–225 °C; <sup>1</sup>H NMR δ (400 MHz, d<sub>6</sub>-DMSO) 1.43 (s, 9H), 1.53 (ddd, *J* = 24.9, 12.4, 3.8 Hz, 2H), 1.79 (d, *J* = 12.7 Hz, 2H), 2.77–2.85 (m, 3H), 4.11 (d, *J* = 10.8 Hz, 2H), 4.30 (s, 2H), 4.50 (s, 2H), 7.44 (d, *J* = 7.6 Hz, 2H), 7.59 (d, *J* = 8.3 Hz, 1H), 7.93 (d, *J* = 7.3 Hz, 2H), 7.97 (d, *J* = 8.3 Hz, 1H), 8.27 (s, 1H), 10.41 (s, 1H); HRMS (FAB) *m/z* calcd for C<sub>27</sub>H<sub>32</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 494.5726, found: 494.5721.

**4.1.8.4. {6-[4-(4-Boc-piperazin-1-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}acetic acid (16d).** Mp = 215–216.5 °C (decomposes); <sup>1</sup>H NMR δ (500 MHz, d<sub>6</sub>-DMSO) 1.44 (s, 9H), 3.31 (s, 4H), 3.48 (s, 4H), 4.29 (s, 2H), 4.49 (s, 2H), 7.05 (d, *J* = 8.3 Hz, 2H), 7.56 (d, *J* = 8.3 Hz, 1H), 7.93 (d, *J* = 8.3 Hz, 2H), 7.98 (d, *J* = 8.3 Hz, 1H), 8.23 (s, 1H), 10.17 (s, 1H), 12.97 (br s, 1H); HRMS (FAB) *m/z* calcd for C<sub>26</sub>H<sub>31</sub>N<sub>4</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 495.5602, found: 495.5610.

**4.1.8.5. {6-[(2-Boc-1,2,3,4-tetrahydroisoquinoline-7-carbonyl)amino]-1-oxo-1,3-dihydroisoindol-2-yl}acetic acid (16e).** Mp = 218–220 °C (decomposes); <sup>1</sup>H NMR δ (400 MHz, d<sub>6</sub>-DMSO) 1.45 (s, 9H), 2.87 (t, *J* = 5.7 Hz, 2H), 3.60 (t, *J* = 5.7 Hz, 2H), 4.29 (s, 2H), 4.50 (s, 2H), 4.61 (s, 2H), 7.35 (d, *J* = 8.1 Hz, 1H), 7.60 (d, *J* = 8.3 Hz, 1H), 7.81 (d, *J* = 8.1 Hz, 1H), 7.83 (s, 1H), 7.98 (dd, *J* = 8.3, 2.0 Hz, 1H), 8.23 (d, *J* = 1.7 Hz, 1H), 10.41 (s, 1H), 12.96 (br s, 1H); HRMS (FAB) *m/z* calcd for C<sub>25</sub>H<sub>28</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 466.5184, found: 466.5187.

**4.1.8.6. 3-{6-[2-(1-Boc-piperidin-4-yl)acetylamino]-1-oxo-1,3-dihydroisoindol-2-yl}propionic acid (16f).** Mp = 193–194 °C; <sup>1</sup>H NMR δ (400 MHz, d<sub>6</sub>-DMSO) 1.08 (ddd, *J* = 24.6, 12.4, 3.3 Hz, 2H), 1.40 (s, 9H), 1.66 (d, *J* = 11.5 Hz, 2H), 1.90–2.01 (m, 1H), 2.28 (d, *J* = 7.1 Hz, 2H), 2.60 (t, *J* = 7.0 Hz, 2H), 2.66–2.82 (m,

2H), 3.72 (t, *J* = 7.1 Hz, 2H), 3.92 (d, *J* = 11.7 Hz, 2H), 4.43 (s, 2H), 7.51 (d, *J* = 8.1 Hz, 1H), 7.67 (dd, *J* = 8.1, 1.2 Hz, 1H), 8.04 (d, *J* = 1.2 Hz, 1H), 10.11 (s, 1H), 12.36 (br s, 1H); HRMS (FAB) *m/z* calcd for C<sub>23</sub>H<sub>32</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 446.5280, found: 546.5274.

**4.1.8.7. 3-{6-[4-(1-Boc-piperidin-4-yl)butyrylamino]-1-oxo-1,3-dihydroisoindol-2-yl}propionic acid (16g).** Mp = 176.5–177.5 °C; <sup>1</sup>H NMR δ (400 MHz, d<sub>6</sub>-DMSO) 0.93 (ddd, *J* = 23.0, 10.5, 3.2 Hz, 2H), 1.22 (dd, *J* = 14.2, 6.9 Hz, 2H), 1.36 (s, 9H), 1.49–1.46 (m, 1H), 1.57–1.64 (m, 4H), 2.30 (t, *J* = 7.3 Hz, 2H), 2.58 (t, *J* = 6.9 Hz, 2H), 2.62–2.76 (m, 2H), 3.69 (t, *J* = 6.9 Hz, 2H), 3.90 (d, *J* = 10.5 Hz, 2H), 4.41 (s, 2H), 7.48 (d, *J* = 8.1 Hz, 1H), 7.66 (dd, *J* = 8.1, 1.5 Hz, 1H), 8.01 (d, *J* = 1.5 Hz, 1H), 10.10 (s, 1H), 12.34 (br s, 1H); HRMS (FAB) *m/z* calcd for C<sub>25</sub>H<sub>36</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 474.5822, found: 508.5828.

**4.1.8.8. 3-{6-[4-(1-Boc-piperidin-4-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}propionic acid (16h).** Mp = 195–196 °C; <sup>1</sup>H NMR δ (400 MHz, d<sub>6</sub>-DMSO) 1.41 (s, 9H), 1.52 (ddd, *J* = 25.1, 12.6, 3.9 Hz, 2H), 1.77 (d, *J* = 12.5 Hz, 2H), 2.60 (t, *J* = 7.1 Hz, 2H), 2.74–2.85 (m, 3H), 3.72 (t, *J* = 7.0 Hz, 2H), 4.08 (d, *J* = 10.3 Hz, 2H), 4.45 (s, 2H), 7.41 (d, *J* = 8.3 Hz, 2H), 7.54 (d, *J* = 8.1 Hz, 1H), 7.90 (d, *J* = 8.3 Hz, 2H), 7.92 (dd, *J* = 8.1, 1.7 Hz, 1H), 8.17 (d, *J* = 1.7 Hz, 1H), 10.37 (s, 1H), 12.37 (br s, 1H); HRMS (FAB) *m/z* calcd for C<sub>28</sub>H<sub>34</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 508.5997, found: 508.6002.

**4.1.8.9. 3-{6-[4-(4-Boc-piperazin-1-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}propionic acid (16i).** Mp >300 °C (decomposes); <sup>1</sup>H NMR δ (400 MHz, d<sub>6</sub>-DMSO) 1.43 (s, 9H), 2.62 (t, *J* = 6.0 Hz, 2H), 3.30 (s, 4H), 3.48 (s, 4H), 3.74 (t, *J* = 6.0 Hz, 2H), 4.46 (s, 2H), 7.04 (d, *J* = 7.9 Hz, 2H), 7.54 (d, *J* = 7.5 Hz, 1H), 7.91–7.95 (m, 2H), 8.18 (s, 1H), 10.15 (s, 1H), 12.35 (br s, 1H); HRMS (FAB) *m/z* calcd for C<sub>27</sub>H<sub>33</sub>N<sub>4</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 509.5873, found: 509.5870.

**4.1.8.10. 3-{6-[(2-Boc-1,2,3,4-tetrahydroisoquinoline-7-carbonyl)amino]-1-oxo-1,3-dihydroisoindol-2-yl}propionic acid (16j).** Mp = 118–120 °C (decomposes); <sup>1</sup>H NMR δ (400 MHz, d<sub>6</sub>-DMSO) 1.45 (s, 9H), 2.63 (t, *J* = 7.1 Hz, 2H), 2.87 (t, *J* = 5.7 Hz, 2H), 3.60 (t, *J* = 5.5 Hz, 2H), 3.75 (t, *J* = 7.0 Hz, 2H), 4.48 (s, 2H), 4.61 (s, 2H), 7.34 (d, *J* = 8.1 Hz, 1H), 7.57 (d, *J* = 8.3 Hz, 1H), 7.80 (d, *J* = 8.1 Hz, 1H), 7.83 (s, 1H), 7.95 (dd, *J* = 8.2, 1.6 Hz, 1H), 8.19 (d, *J* = 1.6 Hz, 1H), 10.38 (s, 1H), 12.35 (br s, 1H); HRMS (FAB) *m/z* calcd for C<sub>26</sub>H<sub>30</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 480.5455, found: 480.5457.

**4.1.8.11. 3-{6-[3-(4-Boc-piperazin-1-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}propionic acid (16k).** Mp = 196–196.5 °C; <sup>1</sup>H NMR δ (500 MHz, d<sub>6</sub>-DMSO) 1.43 (s, 9H), 2.53–2.54 (m, 2H), 3.21 (t, *J* = 6.4 Hz, 4H), 3.49 (t, *J* = 6.4 Hz, 4H), 3.72 (t, *J* = 6.8 Hz, 2H), 4.48 (s, 2H), 7.18 (d, *J* = 7.6 Hz, 1H), 7.38 (t, *J* = 7.8 Hz, 1H), 7.44 (d, *J* = 7.3 Hz, 1H), 7.53–7.56 (m, 2H), 7.98 (d, *J* = 8.1 Hz, 1H), 8.19 (s, 1H), 10.49 (s, 1H); HRMS (FAB) *m/z* calcd for C<sub>27</sub>H<sub>33</sub>N<sub>4</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 509.5873, found: 509.5879.

**4.1.8.12. 3-{6-[2-(1-Boc-piperidin-4-yl)acetylamino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid (16l).** Mp = 197–199 °C; <sup>1</sup>H NMR δ (500 MHz, d<sub>6</sub>-DMSO) 1.09 (ddd, *J* = 23.6, 11.9, 3.1 Hz, 2H), 1.27 (d, *J* = 6.5 Hz, 3H), 1.40 (s, 9H), 1.67 (d, *J* = 12.2 Hz, 2H), 1.91–2.02 (m, 1H), 2.28 (d, *J* = 7.3 Hz, 2H), 2.62 (ddd, *J* = 31.5, 15.2, 7.3 Hz, 2H), 2.69–2.80 (m, 2H), 3.92 (d, *J* = 9.1 Hz, 2H), 4.39 (dd, *J* = 27.5, 17.4 Hz, 2H), 4.60 (td, *J* = 13.1, 6.0 Hz, 1H), 7.51 (d, *J* = 8.3 Hz, 1H), 7.68 (d, *J* = 8.0 Hz, 1H), 8.02 (s, 1H), 10.09 (s, 1H), 12.26 (s, 1H); HRMS (FAB) *m/z* calcd for C<sub>24</sub>H<sub>34</sub>N<sub>3</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 460.5551, found: 450.5555.

**4.1.8.13.** 3-{6-[4-(1-Boc-piperidin-4-yl)butyrylamino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid (**16m**). Mp = 139–140.5 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 0.95 (ddd,  $J = 24.5, 12.2, 3.7$  Hz, 2H), 1.20–1.25 (m, 5H), 1.39 (s, 9H), 1.42–1.44 (m, 1H), 1.59–1.65 (m, 4H), 2.32 (t,  $J = 7.1$  Hz, 2H), 2.62–2.69 (m, 2H), 3.92 (d,  $J = 8.1$  Hz, 2H), 4.36 (s, 2H), 4.59 (m, 1H), 7.46 (d,  $J = 7.8$  Hz, 1H), 7.69 (d,  $J = 7.8$  Hz, 1H), 8.03 (s, 1H), 10.23 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{26}\text{H}_{38}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 488.6093, found: 488.6089.

**4.1.8.14.** 3-{6-[4-(1-Boc-piperidin-4-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid (**16n**). Mp = 175–176 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.25 (d,  $J = 6.4$  Hz, 3H), 1.43 (s, 9H), 1.54 (ddd,  $J = 24.8, 12.2, 3.5$  Hz, 2H), 1.79 (d,  $J = 12.2$  Hz, 2H), 2.74–2.89 (m, 3H), 4.10 (d,  $J = 6.4$  Hz, 2H), 4.41 (s, 2H), 4.62 (dt,  $J = 13.5, 6.5$  Hz, 1H), 7.42 (d,  $J = 7.8$  Hz, 2H), 7.55 (d,  $J = 8.3$  Hz, 1H), 7.93–7.96 (m, 3H), 8.18 (s, 1H), 10.43 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{29}\text{H}_{36}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 522.6268, found: 522.6270.

**4.1.8.15.** 3-{6-[4-(4-Boc-piperazin-1-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid (**16o**). Mp = 155.5–156 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.29 (d,  $J = 6.6$  Hz, 3H), 1.44 (s, 9H), 2.64 (ddd,  $J = 32.3, 15.3, 7.5$  Hz, 2H), 3.30 (s, 4H), 3.48 (s, 4H), 4.41 (dd,  $J = 27.1, 17.1$  Hz, 2H), 4.62 (td,  $J = 13.8, 7.0$  Hz, 1H), 7.04 (d,  $J = 8.8$  Hz, 2H), 7.55 (d,  $J = 8.1$  Hz, 1H), 7.93 (d,  $J = 8.1$  Hz, 2H), 7.95 (d,  $J = 8.8$  Hz, 1H), 8.18 (s, 1H), 10.16 (s, 1H), 12.27 (br s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{28}\text{H}_{35}\text{N}_4\text{O}_6$   $[\text{M}+\text{H}]^+$ : 523.6144, found: 523.6139.

**4.1.8.16.** 3-{6-[(2-Boc-1,2,3,4-tetrahydroisoquinoline-7-carbonyl)-amino]-1-oxo-1,3-dihydroisoindol-2-yl}-butyric acid (**16p**). Mp = 153–154 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.29 (d,  $J = 6.6$  Hz, 3H), 1.45 (s, 9H), 2.64 (ddd,  $J = 32.8, 15.2, 7.5$  Hz, 2H), 2.87 (t,  $J = 5.2$  Hz, 2H), 3.60 (t,  $J = 5.2$  Hz, 2H), 4.43 (dd,  $J = 27.1, 17.4$  Hz, 2H), 4.55–4.64 (m, 3H), 7.34 (d,  $J = 7.8$  Hz, 1H), 7.57 (d,  $J = 8.3$  Hz, 1H), 7.80 (d,  $J = 8.8$  Hz, 1H), 7.83 (s, 1H), 7.95 (dd,  $J = 8.1, 1.5$  Hz, 2H), 8.18 (s, 1H), 10.36 (s, 1H), 12.30 (br s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{27}\text{H}_{32}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 494.5726, found: 493.5723.

**4.1.8.17.** 3-{6-[3-(4-Boc-piperazin-1-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid (**16q**). Mp >300 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.29 (d,  $J = 3.9$  Hz, 3H), 1.44 (s, 9H), 2.65 (ddd,  $J = 33.2, 18.0, 6.4$  Hz, 2H), 3.22 (s, 4H), 3.51 (s, 4H), 4.43 (dd,  $J = 26.2, 16.1$  Hz, 2H), 4.62 (td,  $J = 14.1, 6.2$  Hz, 1H), 7.20 (d,  $J = 4.4$  Hz, 1H), 7.38–7.43 (m, 2H), 7.51 (s, 1H), 7.58 (d,  $J = 6.5$  Hz, 1H), 7.95 (d,  $J = 7.0$  Hz, 1H), 8.16 (s, 1H), 10.35 (s, 1H), 12.30 (br s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{28}\text{H}_{35}\text{N}_4\text{O}_6$   $[\text{M}+\text{H}]^+$ : 523.6144, found: 523.6139.

**4.1.8.18.** 4-{6-[2-(1-Boc-piperidin-4-yl)acetyl-amino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid (**16r**). Mp = 191.5–192 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.06–1.13 (m, 2H), 1.40 (s, 9H), 1.67 (d,  $J = 12.2$  Hz, 2H), 1.84 (t,  $J = 7.2$ , 2H), 1.94–1.99 (m, 1H), 2.25–2.28 (m, 4H), 2.67–2.79 (m, 2H), 3.54 (t,  $J = 7.2$  Hz, 2H), 3.93 (d,  $J = 10.1$  Hz, 2H), 4.41 (s, 2H), 7.50 (d,  $J = 7.0$  Hz, 1H), 7.68 (d,  $J = 7.0$  Hz, 1H), 8.03 (s, 1H), 10.09 (s, 1H), 12.07 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{24}\text{H}_{34}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 460.5551, found: 459.5559.

**4.1.8.19.** 4-{6-[4-(1-Boc-piperidin-4-yl)butyrylamino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid (**16s**). Mp = 190–191 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 0.94 (ddd,  $J = 24.4, 12.3, 3.9$  Hz, 2H), 1.20–1.26 (m, 2H), 1.38 (s, 9H), 1.42–1.44 (m, 1H), 1.60–1.65 (m, 4H), 1.82 (dt,  $J = 14.2, 7.1$  Hz, 2H), 2.24 (t,  $J = 7.2$  Hz, 2H), 2.31 (t,  $J = 7.3$  Hz, 2H), 2.63–2.72 (m, 2H), 3.51 (t,  $J = 7.0$  Hz, 2H), 3.91 (d,  $J = 10.3$  Hz, 2H), 4.40 (s, 2H), 7.49 (d,  $J = 8.3$  Hz, 1H), 7.68 (dd,  $J = 8.1, 1.5$  Hz, 1H), 8.02 (s, 1H), 10.08 (s, 1H), 12.09 (br s, 1H);

HRMS (FAB)  $m/z$  calcd for  $\text{C}_{26}\text{H}_{37}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 488.6093, found: 488.6090.

**4.1.8.20.** 4-{6-[4-(1-Boc-piperidin-4-yl)benzoylamino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid (**16t**). Mp = 197–197.5 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 1.42 (s, 9H), 1.53 (ddd,  $J = 25.1, 12.8, 3.9$  Hz, 2H), 1.78 (d,  $J = 13.7$  Hz, 2H), 1.84 (dt,  $J = 14.2, 6.8$  Hz, 2H), 2.25 (t,  $J = 7.2$  Hz, 2H), 2.75–2.85 (m, 3H), 3.54 (t,  $J = 6.8$  Hz, 2H), 4.09 (d,  $J = 9.8$  Hz, 2H), 4.44 (s, 2H), 7.42 (d,  $J = 8.1$  Hz, 2H), 7.55 (d,  $J = 8.3$  Hz, 1H), 7.91–7.94 (m, 3H), 8.17 (s, 1H), 10.38 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{29}\text{H}_{36}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 522.6268, found: 522.6261.

**4.1.8.21.** 4-{6-[(2-Boc-1,2,3,4-tetrahydroisoquinoline-7-carbonyl)-amino]-1-oxo-1,3-dihydroisoindol-2-yl}butyric acid (**16u**). Mp = 90–93 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 1.44 (s, 9H), 1.84 (dt,  $J = 13.9, 7.1$  Hz, 2H), 2.26 (t,  $J = 7.0$  Hz, 2H), 2.86 (t,  $J = 5.1$  Hz, 2H), 3.54 (t,  $J = 6.7$  Hz, 2H), 3.59 (t,  $J = 4.6$  Hz, 2H), 4.45 (s, 2H), 4.60 (s, 2H), 7.33 (d,  $J = 7.6$  Hz, 1H), 7.56 (d,  $J = 8.1$  Hz, 2H), 7.78–7.82 (m, 2H), 7.94 (d,  $J = 8.1$  Hz, 1H), 8.18 (s, 1H), 10.37 (s, 1H), 12.14 (br s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{27}\text{H}_{32}\text{N}_3\text{O}_6$   $[\text{M}+\text{H}]^+$ : 494.5726, found: 494.5721.

#### 4.1.9. General procedures for a preparation of compound 17

**Method A.** This method was used when the compound **16** was not soluble in anhydrous  $\text{CH}_2\text{Cl}_2$ . The compound **16** (10 mmol) was suspended in anhydrous  $\text{CH}_2\text{Cl}_2$ , and trifluoroacetic acid (1 ml) was added. The precipitate was dissolved in a short time, and the reaction solution was kept for 2 h, protected from moisture. After the solvent was evaporated in vacuo to dryness, the residue was dried in vacuo (2 mmHg) for 2 h at 40 °C.

**Method B.** This method was used for the compound **16** soluble in anhydrous  $\text{CH}_2\text{Cl}_2$ . The compound **16** (10 mmol) was dissolved in anhydrous  $\text{CH}_2\text{Cl}_2$ , and the stream of dry HCl was passed through the solution for 1 h. The solvent was evaporated, and the solid residue was dried in vacuo (2 mm Hg) for 2 h at 40 °C.

**4.1.9.1. [1-Oxo-6-(2-piperidinium-4-yl-acetyl-amino)-1,3-dihydroisoindol-2-yl]acetic acid trifluoroacetate (17a).** Mp = 236–239 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.41 (ddd,  $J = 22.6, 10.6, 2.6$  Hz, 2H), 1.85 (d,  $J = 13.8$  Hz, 2H), 2.05–2.13 (m, 1H), 2.33 (d,  $J = 6.7$  Hz, 2H), 2.91 (dd,  $J = 22.8, 11.6$  Hz, 2H), 3.27 (d,  $J = 11.9$  Hz, 2H), 4.28 (s, 2H), 4.46 (s, 2H), 7.54 (d,  $J = 8.2$  Hz, 1H), 7.69 (d,  $J = 8.2$  Hz, 1H), 8.10 (s, 1H), 8.37 (d,  $J = 8.7$  Hz, 1H), 8.65 (d,  $J = 8.7$  Hz, 1H), 10.23 (s, 1H);  $^{13}\text{C NMR } \delta$  (125 MHz,  $d_6$ -DMSO) 28.64 (2C), 31.17 (1C), 43.05 (1C), 43.58 (2C), 43.95 (1C), 50.41 (1C), 113.61 (1C), 116.75 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 123.10 (1C), 124.17 (s, 1C), 132.64 (1C), 137.06 (1C), 139.42 (1C), 158.81 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 168.15 (1C), 170.37 (1C), 170.98 (1C); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{17}\text{H}_{22}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 332.3826, found: 332.3825.

**4.1.9.2. [1-Oxo-6-(4-piperidinium-4-yl-butyl-amino)-1,3-dihydroisoindol-2-yl]acetic acid trifluoroacetate (17b).** Mp = 114–115 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 0.75–0.85 (m, 4H), 1.02–1.11 (m, 1H), 1.16 (dt,  $J = 14.9, 7.5$  Hz, 2H), 1.35 (d,  $J = 13.4$  Hz, 2H), 1.88 (t,  $J = 7.2$  Hz, 2H), 2.37 (dd,  $J = 32.5, 11.9$  Hz, 1H), 2.80 (d,  $J = 12.2$  Hz, 2H), 3.81 (s, 2H), 3.99 (s, 2H), 7.06 (d,  $J = 8.3$  Hz, 1H), 7.24 (dd,  $J = 8.2, 1.1$  Hz, 1H), 7.64 (s, 1H), 7.86 (d,  $J = 8.1$  Hz, 1H), 8.17 (d,  $J = 8.1$  Hz, 1H), 9.71 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{19}\text{H}_{26}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 360.4368, found: 360.4374.

**4.1.9.3. [1-Oxo-6-(4-piperidinium-4-yl-benzoylamino)-1,3-dihydroisoindol-2-yl]acetic acid trifluoroacetate (17c).** Mp >300 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 1.84 (ddd,  $J = 26.7, 12.5, 2.9$  Hz, 2H), 1.99 (d,  $J = 13.2$  Hz, 2H), 2.93–2.98 (m, 1H), 3.04 (dd,  $J = 23.5, 12.2$  Hz, 2H), 3.41 (d,  $J = 11.7$  Hz, 2H), 4.30 (s, 2H), 4.50



(s, 2H), 7.41 (d,  $J = 8.3$  Hz, 2H), 7.60 (d,  $J = 8.3$  Hz, 1H), 7.95–7.99 (m, 3H), 8.23 (d,  $J = 1.5$  Hz, 1H), 8.42 (dd,  $J = 21.3, 10.8$  Hz, 1H), 8.69 (d,  $J = 11.7$  Hz, 1H), 10.44 (s, 1H);  $^{13}\text{C}$  NMR  $\delta$  (125 MHz,  $d_6$ -DMSO) 29.67 (2C), 39.25 (1C), 43.98 (1C), 44.05 (2C), 50.46 (1C), 114.75 (1C), 116.89 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 124.08 (1C), 124.22 (1C), 127.06 (2C), 128.58 (2C), 132.60 (1C), 133.56 (1C), 137.50 (1C), 139.58 (1C), 148.86 (1C), 158.77 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 165.95 (1C), 168.18 (1C), 170.99 (1C); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{22}\text{H}_{24}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 394.4543, found: 394.4545.

**4.1.9.4. [1-Oxo-6-(4-piperazinium-1-yl-benzoylamino)-1,3-dihydroisoindol-2-yl]acetic acid trifluoroacetate (17d).** Mp = 241–242 °C;  $^1\text{H}$  NMR  $\delta$  (500 MHz,  $d_6$ -DMSO) 3.27 (s, 4H) 3.54 (s, 4H) 4.30 (s, 2H) 4.49 (s, 2H) 7.11 (d,  $J = 8.3$  Hz, 2H) 7.58 (d,  $J = 8.1$  Hz, 1H) 7.96–7.99 (m, 3H) 8.23 (s, 1H) 8.96 (s, 2H) 10.23 (s, 1H);  $^{13}\text{C}$  NMR  $\delta$  (125 MHz,  $d_6$ -DMSO) 42.98 (2C), 43.97 (1C), 44.90 (2C), 50.45 (1C), 114.70 (1C), 114.73 (2C), 116.67 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 123.97 (1C), 124.19 (1C), 125.11 (1C), 129.70 (2C), 132.55 (1C), 137.15 (1C), 139.84 (1C), 152.63 (1C), 158.75 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 165.47 (1C), 168.23 (1C), 171.01 (1C); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{21}\text{H}_{23}\text{N}_4\text{O}_4$   $[\text{M}+\text{H}]^+$ : 395.4419, found: 366.4422.

**4.1.9.5. {1-Oxo-6-[(1,2,3,4-tetrahydroisoquinolinium-7-carbonyl)amino]-1,3-dihydroisoindol-2-yl}acetic acid chloride (17e).** Mp = 297.5–298 °C (decomposes);  $^1\text{H}$  NMR  $\delta$  (400 MHz,  $d_6$ -DMSO) 3.11 (t,  $J = 6.0$  Hz, 2H), 3.40 (dd,  $J = 10.5, 6.3$  Hz, 2H), 4.30 (s, 2H), 4.35 (t,  $J = 4.0$  Hz, 2H), 4.50 (s, 2H), 7.40 (d,  $J = 8.3$  Hz, 1H), 7.60 (d,  $J = 8.3$  Hz, 1H), 7.90–7.92 (m, 2H), 8.01 (dd,  $J = 8.3, 1.7$  Hz, 1H), 8.26 (d,  $J = 1.7$  Hz, 1H), 9.76 (s, 2H), 10.59 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{20}\text{H}_{20}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 366.4001, found: 366.3997.

**4.1.9.6. 3-[1-Oxo-6-(2-piperidinium-4-yl-acetyl-amino)-1,3-dihydroisoindol-2-yl]propionic acid trifluoroacetate (17f).** Mp = 112–114 °C;  $^1\text{H}$  NMR  $\delta$  (500 MHz,  $d_6$ -DMSO) 1.40 (dd,  $J = 24.9, 13.2$  Hz, 2H) 1.85 (d,  $J = 15.1$  Hz, 2H) 2.03–2.14 (m, 1H), 2.33 (d,  $J = 7.0$  Hz, 2H), 2.61 (t,  $J = 6.8$  Hz, 2H), 2.91 (dd,  $J = 22.8, 11.5$  Hz, 2H), 3.27 (d,  $J = 11.9$  Hz, 2H), 3.73 (t,  $J = 6.8$  Hz, 2H), 4.44 (s, 2H), 7.51 (d,  $J = 8.0$  Hz, 1H), 7.66 (d,  $J = 8.0$  Hz, 1H), 8.05 (s, 1H), 8.30 (d,  $J = 10.1$  Hz, 1H), 8.59 (d,  $J = 7.0$  Hz, 1H), 10.19 (s, 1H);  $^{13}\text{C}$  NMR  $\delta$  (125 MHz,  $d_6$ -DMSO) 28.65 (2C), 31.16 (1C), 33.36 (1C), 38.60 (1C), 43.04 (1C), 43.59 (2C), 49.97 (1C), 113.42 (1C), 116.58 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 122.76 (1C), 124.02 (1C), 133.25 (1C), 136.85 (1C), 139.34 (1C), 158.74 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 167.64 (1C), 170.31 (1C), 173.26 (1C); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{18}\text{H}_{24}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 346.4097, found: 374.4690.

**4.1.9.7. 3-[1-Oxo-6-(4-piperidinium-4-yl-butrylamino)-1,3-dihydroisoindol-2-yl]propionic acid trifluoroacetate (17g).** Mp = 196–197 °C;  $^1\text{H}$  NMR  $\delta$  (400 MHz,  $d_6$ -DMSO) 1.20–1.30 (m, 4H), 1.49–1.55 (m, 1H), 1.58–1.65 (m, 2H), 1.81 (d,  $J = 12.9$  Hz, 2H), 2.34 (t,  $J = 6.6$  Hz, 2H), 2.60 (t,  $J = 6.9$  Hz, 2H), 2.83 (dd,  $J = 20.8, 10.0$  Hz, 2H), 3.26 (d,  $J = 11.2$  Hz, 2H), 3.72 (t,  $J = 6.9$  Hz, 2H), 4.43 (s, 2H), 7.50 (d,  $J = 7.9$  Hz, 1H), 7.67 (d,  $J = 7.5$  Hz, 1H), 8.05 (s, 1H), 8.32 (s, 1H), 8.64 (s, 1H), 10.14 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{20}\text{H}_{28}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 374.4639, found: 374.4643.

**4.1.9.8. 3-[1-Oxo-6-(4-piperidinium-4-yl-benzoylamino)-1,3-dihydroisoindol-2-yl]propionic acid trifluoroacetate (17h).** Mp = 243–244 °C (decomposes);  $^1\text{H}$  NMR  $\delta$  (400 MHz,  $d_6$ -DMSO) 1.85 (dd,  $J = 23.7, 11.5$  Hz, 2H), 1.99 (d,  $J = 13.2$  Hz, 2H), 2.63 (t,  $J = 7.0$  Hz, 2H), 2.93–3.09 (m, 3H), 3.42 (d,  $J = 11.7$  Hz, 2H), 3.75 (t,  $J = 7.0$  Hz, 2H), 4.48 (s, 2H), 7.41 (d,  $J = 8.1$  Hz, 2H), 7.57 (d,  $J = 8.3$  Hz, 1H), 7.94 (dd,  $J = 8.3, 1.5$  Hz, 1H), 7.97 (d,  $J = 8.1$  Hz, 2H), 8.20 (d,  $J = 1.5$  Hz, 1H), 8.57 (s, 1H), 8.77 (t,  $J = 12.7, 2H$ ), 10.43 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{23}\text{H}_{26}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 408.4814, found: 408.4816.

**4.1.9.9. 3-[1-Oxo-6-(4-piperazinium-1-yl-benzoylamino)-1,3-dihydroisoindol-2-yl]propionic acid trifluoroacetate (17i).** Mp = 212–214 °C;  $^1\text{H}$  NMR  $\delta$  (400 MHz,  $d_6$ -DMSO +  $\text{CCl}_4$ ) 2.62 (t,  $J = 6.6$  Hz, 2H), 3.27 (s, 4H), 3.53 (s, 4H), 3.74 (t,  $J = 6.6$  Hz, 2H), 4.46 (s, 2H), 7.10 (d,  $J = 8.7$  Hz, 2H), 7.55 (d,  $J = 8.3$  Hz, 1H), 7.93–7.99 (m, 3H), 8.19 (s, 1H), 9.02 (s, 2H), 10.21 (s, 1H);  $^{13}\text{C}$  NMR  $\delta$  (125 MHz,  $d_6$ -DMSO) 33.39 (1C), 38.61 (1C), 42.98 (2C), 44.89 (2C), 50.01 (1C), 114.50 (1C), 114.72 (2C), 116.68 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 123.83 (1C), 123.86 (1C), 129.69 (2C), 133.15 (1C), 136.94 (1C), 139.76 (1C), 152.61 (1C), 158.57 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 165.44 (1C), 167.74 (1C), 173.29 (1C); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{22}\text{H}_{25}\text{N}_4\text{O}_4$   $[\text{M}+\text{H}]^+$ : 409.4689, found: 409.4696.

**4.1.9.10. 3-[1-Oxo-6-[(1,2,3,4-tetrahydroisoquinolinium-7-carbonyl)amino]-1,3-dihydroisoindol-2-yl]-propionic acid chloride (17j).** Mp = 276–277.5 °C;  $^1\text{H}$  NMR  $\delta$  (400 MHz,  $d_6$ -DMSO) 2.63 (t,  $J = 7.0$  Hz, 2H), 3.10 (t,  $J = 5.9$  Hz, 2H), 3.40 (dd,  $J = 6.5, 4.4$  Hz, 2H), 3.74 (t,  $J = 7.0$  Hz, 2H), 4.35 (s, 2H), 4.48 (s, 2H), 7.40 (d,  $J = 8.6$  Hz, 1H), 7.56 (d,  $J = 8.3$  Hz, 1H), 7.89–7.92 (m, 2H), 7.97 (dd,  $J = 8.3, 1.7$  Hz, 1H), 8.21 (d,  $J = 1.2$  Hz, 1H), 9.67 (s, 2H), 10.54 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{21}\text{H}_{22}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 380.4272, found: 380.4281.

**4.1.9.11. 3-[1-Oxo-6-(3-piperazinium-1-yl-benzoylamino)-1,3-dihydroisoindol-2-yl]propionic acid trifluoroacetate (17k).** Mp = 102–103 °C;  $^1\text{H}$  NMR  $\delta$  (500 MHz,  $d_6$ -DMSO) 2.63 (t,  $J = 7.1$  Hz, 2H), 3.29 (s, 4H), 3.46 (t,  $J = 4.8$  Hz, 4H), 3.75 (t,  $J = 6.8$  Hz, 2H), 4.48 (s, 2H), 7.24 (d,  $J = 8.3$  Hz, 1H), 7.43 (t,  $J = 7.7$  Hz, 1H), 7.48 (d,  $J = 7.6$  Hz, 1H), 7.52 (s, 1H), 7.58 (d,  $J = 8.1$  Hz, 1H), 7.94 (d,  $J = 7.8$  Hz, 1H), 8.17 (s, 1H), 8.88 (s, 2H), 10.39 (s, 1H);  $^{13}\text{C}$  NMR  $\delta$  (125 MHz,  $d_6$ -DMSO) 33.39 (1C), 38.62 (1C), 43.17 (2C), 45.93 (2C), 50.04 (1C), 114.69 (1C), 115.37 (1C), 116.09 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 119.66 (1C), 119.78 (1C), 123.93 (1C), 124.01 (1C), 129.72 (1C), 133.21 (1C), 136.18 (1C), 137.34 (1C), 139.44 (1C), 150.46 (1C), 158.76 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 166.30 (1C), 167.67 (1C), 173.28 (1C); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{22}\text{H}_{25}\text{N}_4\text{O}_4$   $[\text{M}+\text{H}]^+$ : 409.4689, found: 409.4687.

**4.1.9.12. 3-[1-Oxo-6-(2-piperidinium-4-yl-acetyl-amino)-1,3-dihydroisoindol-2-yl]butyric acid trifluoroacetate (17l).** Glassy substance;  $^1\text{H}$  NMR  $\delta$  (500 MHz,  $d_6$ -DMSO) 1.27 (d,  $J = 6.7$  Hz, 3H), 1.37–1.44 (m, 2H), 1.85 (d,  $J = 13.5$  Hz, 2H), 2.05–2.12 (m, 1H), 2.33 (d,  $J = 6.7$  Hz, 2H), 2.62 (ddd,  $J = 30.8, 15.3, 7.4$  Hz, 2H), 2.92 (dd,  $J = 22.8, 11.9$  Hz, 2H), 3.27 (d,  $J = 11.9$  Hz, 2H), 4.39 (dd,  $J = 27.8, 17.1$  Hz, 2H), 4.59 (dt,  $J = 14.4, 7.5$  Hz, 1H), 7.52 (d,  $J = 8.0$  Hz, 1H), 7.67 (d,  $J = 8.0$  Hz, 1H), 8.04 (s, 1H), 8.35 (d,  $J = 7.3$  Hz, 1H), 8.62 (d,  $J = 7.0$  Hz, 1H), 10.20 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{19}\text{H}_{26}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 360.4368, found: 360.4660.

**4.1.9.13. 3-[1-Oxo-6-(4-piperidinium-4-yl-butrylamino)-1,3-dihydroisoindol-2-yl]butyric acid chloride (17m).** Mp = 231–234 °C;  $^1\text{H}$  NMR  $\delta$  (500 MHz,  $d_6$ -DMSO) 1.21–1.29 (m, 5H), 1.32 (dd,  $J = 23.1, 11.4$  Hz, 2H), 1.49–1.56 (m, 1H), 1.61 (t,  $J = 6.6$  Hz, 2H), 1.79 (d,  $J = 13.5$  Hz, 2H), 2.36 (t,  $J = 5.8$  Hz, 2H), 2.61 (ddd,  $J = 31.1, 15.1, 7.8$  Hz, 2H), 2.80 (dd,  $J = 21.2, 10.5$  Hz, 2H), 3.21 (d,  $J = 11.2$  Hz, 2H), 4.37 (dd,  $J = 26.7, 17.7$  Hz, 2H), 4.59 (dd,  $J = 12.8, 6.4$  Hz, 1H), 7.49 (d,  $J = 7.8$  Hz, 1H), 7.74 (d,  $J = 7.0$  Hz, 1H), 8.07 (s, 1H), 8.82 (d,  $J = 7.6$  Hz, 1H), 9.09 (d,  $J = 7.6$  Hz, 1H), 10.36 (s, 1H);  $^{13}\text{C}$  NMR  $\delta$  (125 MHz,  $d_6$ -DMSO) 18.51 (1C), 22.02 (1C), 28.30 (2C), 32.86 (1C), 35.06 (1C), 36.33 (1C), 43.15 (2C), 44.43 (1C), 45.58 (1C), 112.87 (1C), 122.25 (1C), 123.48 (1C), 133.02 (1C), 136.05 (1C), 139.27 (1C), 166.82 (1C), 171.47 (1C), 172.25 (1C); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{21}\text{H}_{30}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 388.4910, found: 388.4872.

**4.1.9.14. 3-[1-Oxo-6-(4-piperidinium-4-yl-benzoylamino)-1,3-dihydroisoindol-2-yl]butyric acid chloride (17n).** Mp = 200–202 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.28 (d,  $J = 6.7$  Hz, 3H), 1.94–2.02 (m, 4H), 2.64 (ddd,  $J = 31.4, 15.2, 7.4$  Hz, 2H), 2.94–3.04 (m, 3H), 3.36 (d,  $J = 10.6$  Hz, 2H), 4.43 (dd,  $J = 27.8, 17.4$  Hz, 2H), 4.61 (td,  $J = 14.0, 7.0$  Hz, 1H), 7.40 (d,  $J = 7.8$  Hz, 2H), 7.56 (d,  $J = 8.0$  Hz, 1H), 7.97–8.01 (m, 3H), 8.20 (s, 1H), 9.23–9.27 (m, 1H), 9.32–9.34 (m,  $J = 9.6$  Hz, 1H), 10.50 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{24}\text{H}_{28}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 422.5085, found: 422.5091.

**4.1.9.15. 3-[1-Oxo-6-(4-piperazinium-4-yl-benzoylamino)-1,3-dihydroisoindol-2-yl]butyric acid chloride (17o).** Mp = 258–258.5 °C (decomposes);  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.28 (d,  $J = 6.8$  Hz, 3H), 2.64 (ddd,  $J = 31.2, 15.3, 7.1$  Hz, 2H), 3.21 (s, 4H), 3.57 (t,  $J = 4.3, 4\text{H}$ ), 4.41 (dd,  $J = 27.9, 17.6$  Hz, 2H), 4.61 (td,  $J = 13.6, 6.5$  Hz, 1H), 7.10 (d,  $J = 8.3$  Hz, 2H), 7.55 (d,  $J = 8.3$  Hz, 2H), 7.97 (m,  $J = 8.1$  Hz, 3H), 8.19 (s, 1H), 9.51 (s, 2H), 10.26 (s, 1H);  $^{13}\text{C NMR } \delta$  (125 MHz,  $d_6$ -DMSO) 18.99 (1C), 42.74 (2C), 44.76 (2C), 44.88 (1C), 46.06 (1C), 114.53 (2C), 114.69 (1C), 123.83 (1C), 123.85 (1C), 125.03 (1C), 129.73 (2C), 133.42 (1C), 136.84 (1C), 139.80 (1C), 152.63 (1C), 165.45 (1C), 167.25 (1C), 172.71 (1C); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{23}\text{H}_{27}\text{N}_4\text{O}_4$   $[\text{M}+\text{H}]^+$ : 423.4960, found: 423.4953.

**4.1.9.16. 3-{1-Oxo-6-[(1,2,3,4-tetrahydroisoquinolinium-7-carbonyl)amino]-1,3-dihydroisoindol-2-yl}butyric acid chloride (17p).** Hygroscopic substance;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.29 (d,  $J = 6.6$  Hz, 3H), 2.64 (ddd,  $J = 30.8, 15.2, 7.1$  Hz, 2H), 3.09 (t,  $J = 6.0$  Hz, 2H), 3.45 (d,  $J = 5.1$  Hz, 2H), 4.39–4.48 (m, 4H), 4.62 (td,  $J = 14.2, 7.3$  Hz, 1H), 7.42 (d,  $J = 8.1$  Hz, 1H), 7.59 (d,  $J = 8.6$  Hz, 1H), 7.88–7.95 (m, 3H), 8.18 (s, 1H), 9.17 (s, 2H), 10.46 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{22}\text{H}_{24}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 394.4543, found: 394.4546.

**4.1.9.17. 3-[1-Oxo-6-(3-piperazinium-4-yl-benzoylamino)-1,3-dihydroisoindol-2-yl]butyric acid chloride (17q).** Mp = 232–233 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.28 (d,  $J = 5.4$  Hz, 3H), 2.64 (ddd,  $J = 30.6, 14.8, 7.8$  Hz, 2H), 3.24 (s, 4H), 3.52 (s, 4H), 4.42 (dd,  $J = 26.5, 17.4$  Hz, 2H), 4.59–4.65 (m, 1H), 7.23 (d,  $J = 7.5$  Hz, 1H), 7.41 (t,  $J = 6.7$  Hz, 1H), 7.50 (d,  $J = 6.5$  Hz, 1H), 7.56 (d,  $J = 7.8$  Hz, 1H), 7.60 (s, 1H), 7.99 (d,  $J = 7.3$  Hz, 1H), 8.20 (s, 1H) 9.53 (s, 2H), 10.52 (s, 1H);  $^{13}\text{C NMR } \delta$  (125 MHz,  $d_6$ -DMSO) 18.99 (1C), 38.72 (1C), 42.92 (2C), 44.90 (1C), 45.81 (2C), 46.10 (1C), 114.80 (1C), 115.49 (1C), 119.75 (1C), 123.91 (1C), 124.12 (1C), 129.69 (1C), 133.43 (1C), 136.08 (1C), 137.24 (1C), 139.49 (1C), 150.42 (1C), 166.28 (1C), 167.20 (1C), 172.70 (1C); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{23}\text{H}_{27}\text{N}_4\text{O}_4$   $[\text{M}+\text{H}]^+$ : 423.4960, found: 423.4967.

**4.1.9.18. 4-[1-Oxo-6-(2-piperidinium-4-yl-acetylamino)-1,3-dihydroisoindol-2-yl]butyric acid trifluoroacetate (17r).** Mp = 184–185 °C;  $^1\text{H NMR } \delta$  (500 MHz,  $d_6$ -DMSO) 1.37–1.45 (m, 2H), 1.80–1.87 (m, 4H), 2.05–2.12 (m, 1H), 2.25 (t,  $J = 6.9$  Hz, 2H), 2.33 (d,  $J = 6.7$  Hz, 2H), 2.91 (dd,  $J = 23.0, 11.2$  Hz, 2H), 3.28 (d,  $J = 11.2$  Hz, 2H), 3.53 (t,  $J = 6.9$  Hz, 2H), 4.41 (s, 2H), 7.51 (d,  $J = 7.8$  Hz, 1H), 7.66 (d,  $J = 7.8$  Hz, 1H), 8.05 (s, 1H), 8.33–8.38 (m, 1H), 8.61–8.66 (m, 1H), 10.20 (s, 1H);  $^{13}\text{C NMR } \delta$  (125 MHz,  $d_6$ -DMSO) 23.72 (1C), 28.65 (2C), 31.17 (1C), 31.47 (1C), 41.66 (1C), 43.05 (1C), 43.58 (2C), 49.58 (1C), 113.49 (1C), 116.17 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 122.68 (1C), 124.02 (1C), 133.34 (1C), 136.83 (1C), 139.33 (1C), 158.56 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 167.76 (1C), 170.31 (1C), 174.42 (1C); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{19}\text{H}_{26}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 360.4368, found: 360.4372.

**4.1.9.19. 4-[1-Oxo-6-(4-piperidinium-4-yl-butrylamino)-1,3-dihydroisoindol-2-yl]butyric acid trifluoroacetate (17s).** Mp = 200–202 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 1.20–1.28 (m, 4H), 1.48–1.55 (m, 1H), 1.6 (dt,  $J = 15.0, 7.3$  Hz, 2H), 1.78–1.86 (m, 4H), 2.24 (t,  $J = 7.2$  Hz, 2H), 2.33 (t,  $J = 7.1$  Hz, 2H), 2.83 (dd,  $J = 21.7, 10.9$  Hz, 2H), 3.25 (d,  $J = 12.2$  Hz, 2H), 3.51 (t,  $J = 6.8$  Hz, 2H), 4.40 (s, 2H), 7.49 (d,  $J = 8.1$  Hz, 1H), 7.66 (d,  $J = 8.1$  Hz, 1H), 8.04 (s, 1H), 8.27 (s, 1H), 8.68 (s, 1H), 10.13 (s, 1H);  $^{13}\text{C NMR } \delta$  (125 MHz,  $d_6$ -DMSO) 22.41 (1C), 23.74 (1C), 28.87 (2C), 31.50 (1C), 33.30 (1C), 35.49 (1C), 36.81 (1C), 41.69 (1C), 43.80 (2C), 49.59 (1C), 113.38 (1C), 116.44 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 122.61 (1C), 123.96 (1C), 133.34 (1C), 136.61 (1C), 139.57 (1C), 157.75 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 167.81 (1C), 171.78 (1C), 174.39 (1C); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{21}\text{H}_{30}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 388.4910, found: 388.4904.

**4.1.9.20. 4-[1-Oxo-6-(4-piperidinium-4-yl-benzoylamino)-1,3-dihydroisoindol-2-yl]butyric acid trifluoroacetate (17t).** Mp = 280–281 °C;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 1.79–1.90 (m, 4H), 2.00 (d,  $J = 12.7$  Hz, 2H), 2.26 (t,  $J = 6.6$  Hz, 2H), 2.93–3.08 (m, 3H), 3.40–3.43 (m, 2H), 3.54 (t,  $J = 6.5$  Hz, 2H), 4.45 (s, 2H), 7.40 (d,  $J = 7.6$  Hz, 2H), 7.56 (d,  $J = 7.6$  Hz, 2H), 7.92–7.97 (m, 3H), 8.18 (s, 1H), 8.52 (s, 1H), 8.77 (s, 1H), 10.43 (s, 1H);  $^{13}\text{C NMR } \delta$  (125 MHz,  $d_6$ -DMSO) 23.76 (1C), 29.67 (2C), 31.53 (1C), 39.25 (1C), 41.72 (1C), 44.05 (2C), 49.66 (1C), 114.65 (1C), 116.72 (1C,  $\text{CF}_3\text{CO}_2^-$ ), 123.83 (1C), 123.93 (1C), 127.05 (2C), 128.57 (2C), 133.31 (1C), 133.58 (1C), 137.27 (1C), 139.48 (1C), 148.85 (1C), 158.77 (1C,  $\text{CF}_3\text{CCO}_2^-$ ), 165.91 (1C), 167.81 (1C), 174.41 (1C); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{24}\text{H}_{28}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 422.5085, found: 422.5088.

**4.1.9.21. 4-{1-Oxo-6-[(1,2,3,4-tetrahydroisoquinolinium-7-carbonyl)amino]-1,3-dihydroisoindol-2-yl}butyric acid chloride (17u).** Hygroscopic substance;  $^1\text{H NMR } \delta$  (400 MHz,  $d_6$ -DMSO) 1.86 (dt,  $J = 13.7, 6.7$  Hz, 2H), 2.27 (t,  $J = 6.7$  Hz, 2H), 3.09 (t,  $J = 5.2$  Hz, 2H), 3.43–3.49 (m, 2H), 3.55 (t,  $J = 6.7$  Hz, 2H), 4.40 (s, 2H), 4.46 (s, 2H), 7.42 (d,  $J = 7.8$  Hz, 1H), 7.58 (d,  $J = 8.3$  Hz, 1H), 7.88 (s, 1H), 7.90–7.96 (m, 2H), 8.19 (s, 1H), 9.21 (s, 2H), 10.46 (s, 1H); HRMS (FAB)  $m/z$  calcd for  $\text{C}_{22}\text{H}_{24}\text{N}_3\text{O}_4$   $[\text{M}+\text{H}]^+$ : 394.4543, found: 394.4540.

## 4.2. X-ray diffraction

The X-ray data for **5b**, **14**, **17c**–TfaOH·5H<sub>2</sub>O, **17h**–TfaOH·2H<sub>2</sub>O, and **17t**·H<sub>2</sub>O were collected utilizing MoK $\alpha$  radiation at Nonius kappa CCD diffractometer at 100 K, and at Xcalibur Oxford Diffraction CCD diffractometer at room temperature for **17c**, **17h**, **17i**·H<sub>2</sub>O, and **17j**·H<sub>2</sub>O. Final unit cell dimensions were obtained and refined on an entire data set. All calculations to solve the structures and to refine the models were carried out with the programs SHELXS97 and SHELXL97.<sup>22</sup> In all structures nonhydrogen nondisordered atoms have been refined with anisotropic displacement parameters. In **17c** the Tfa anion is disordered over two positions with the occupancies 0.876(3) and 0.124(4). The F atoms in the minor component were refined in isotropic approximation. In **17i**·H<sub>2</sub>O the fluorine atoms in the Tfa anion are disordered over three positions with the occupancies 0.756(12), 0.156(10) and 0.088(7). The minor components were refined in isotropic approximation. The X-ray data for **17j**·H<sub>2</sub>O obtained from the poor-diffracting crystal revealed the disordering of chloride anion and water molecule. For both of these disordered species five close proximal positions with different occupancies have been found and refined in isotropic approximation with the combined occupancies of unity, both for Cl anion and H<sub>2</sub>O. In all structures the C-bound H atoms were placed in cal-

culated positions and were treated in a riding model approximation with  $U_{\text{iso}}(\text{H}) = 1.2 U_{\text{eq}}(\text{C})$ , the O- and N-bound H-atoms were found from difference Fourier maps and refined with isotropic displacement parameters  $U_{\text{iso}}(\text{H}) = 1.5 U_{\text{eq}}(\text{O})$ ,  $U_{\text{iso}}(\text{H}) = 1.2 U_{\text{eq}}(\text{N})$ . The Figures were produced using Mercury.<sup>23</sup> CCDC 912329–912337 contain the crystallographic data for studied compounds. These data can be obtained free of charge from the Cambridge Crystallographic Data Centre via [www.ccdc.cam.ac.uk/data\\_request/cif](http://www.ccdc.cam.ac.uk/data_request/cif).

#### 4.2.1. Crystal data and structure refinement parameters for 5b

Empirical formula:  $\text{C}_{12}\text{H}_{14}\text{N}_2\text{O}_3$  formula weight: 234.25. Crystal system: monoclinic. Space group:  $P2_1/n$ . Unit cell dimensions:  $a = 5.9940(3)$ ,  $b = 8.1340(5)$ ,  $c = 22.2920(13)$ ,  $\beta = 96.223(2)^\circ$ ,  $V = 1080.45(11) \text{ \AA}^3$ . Index ranges:  $-7 \leq h \leq 7$ ,  $-10 \leq k \leq 9$ ,  $-23 \leq l \leq 28$ . Range for data collection from  $1.84^\circ$  to  $26.99^\circ$ ,  $Z = 4$ .  $D_{\text{calc}} = 1.440 \text{ mg/m}^3$ .  $F(000) = 496$ . Reflections collected 6247. Independent reflections 2360 [ $R(\text{int}) = 0.0429$ ]. Data/restraints/parameters 2360:0:163. Refinement method: Full matrix least squares on  $F^2$ . Goodness-of-fit on  $F^2$ : 1.006. Final  $R$  indices [ $I > 2(I)$ ]:  $R_1 = 0.0493$ ,  $wR_2 = 0.1350$ .  $R$  indices (all data):  $R_1 = 0.0666$ ,  $wR_2 = 0.1464$ . Largest diff. peak and hole: 0.285 and  $-0.286 \text{ e \AA}^{-3}$ .

#### 4.2.2. Crystal data and structure refinement parameters for 14

Empirical formula:  $\text{C}_{15}\text{H}_{19}\text{NO}_4$  formula weight: 277.31. Crystal system: monoclinic. Space group:  $P2_1/c$ . Unit cell dimensions:  $a = 11.9373(9)$ ,  $b = 10.3546(8)$ ,  $c = 11.5307(9)$ ,  $\beta = 102.543(4)^\circ$ ,  $V = 1391.25(19) \text{ \AA}^3$ . Index ranges:  $-15 \leq h \leq 15$ ,  $-12 \leq k \leq 13$ ,  $-14 \leq l \leq 14$ . Range for data collection from  $2.63^\circ$  to  $27.00^\circ$ ,  $Z = 4$ .  $D_{\text{calc}} = 1.324 \text{ mg/m}^3$ .  $F(000) = 592$ . Reflections collected 5573. Independent reflections 3015 [ $R(\text{int}) = 0.0541$ ]. Data/restraints/parameters 3015:0:188. Refinement method: Full matrix least squares on  $F^2$ . Goodness-of-fit on  $F^2$ : 1.006. Final  $R$  indices [ $I > 2(I)$ ]:  $R_1 = 0.0633$ ,  $wR_2 = 0.1400$ .  $R$  indices (all data):  $R_1 = 0.0910$ ,  $wR_2 = 0.1546$ . Largest diff. peak and hole: 0.248 and  $-0.309 \text{ e \AA}^{-3}$ .

#### 4.2.3. Crystal data and structure refinement parameters for [1-oxo-6-(4-piperidinium-4-yl-benzoylamino)-1,3-dihydroisoindol-2-yl]acetate pentahydrate (17c-TfaOH·5H<sub>2</sub>O)

Empirical formula:  $\text{C}_{22}\text{H}_{33}\text{N}_3\text{O}_9$  formula weight: 483.51. Crystal system: monoclinic. Space group:  $P2_1/n$ . Unit cell dimensions:  $a = 12.8160(3)$ ,  $b = 11.1350(3)$ ,  $c = 17.1290(5) \text{ \AA}$ ,  $\beta = 108.7240(8)^\circ$ ,  $V = 2315.04(11) \text{ \AA}^3$ . Index ranges:  $-16 \leq h \leq 16$ ,  $-13 \leq k \leq 13$ ,  $-21 \leq l \leq 21$ . Range for data collection from  $2.40^\circ$  to  $26.50^\circ$ ,  $Z = 4$ .  $D_{\text{calc}} = 1.387 \text{ mg/m}^3$ .  $F(000) = 1032$ . Reflections collected 18213. Independent reflections 4793 [ $R(\text{int}) = 0.0500$ ]. Data/restraints/parameters 4793:0:346. Refinement method: Full matrix least squares on  $F^2$ . Goodness-of-fit on  $F^2$ : 1.003. Final  $R$  indices [ $I > 2(I)$ ]:  $R_1 = 0.0474$ ,  $wR_2 = 0.1117$ .  $R$  indices (all data):  $R_1 = 0.0596$ ,  $wR_2 = 0.1185$ . Largest diff. peak and hole: 0.364 and  $-0.366 \text{ e \AA}^{-3}$ .

#### 4.2.4. Crystal data and structure refinement parameters for 17c

Empirical formula:  $\text{C}_{24}\text{H}_{24}\text{F}_3\text{N}_3\text{O}_6$  formula weight: 507.46. Crystal system: triclinic. Space group:  $P-1$ . Unit cell dimensions:  $a = 7.2240(8)$ ,  $b = 9.8920(9)$ ,  $c = 17.329(2)$ ,  $\alpha = 97.142(7)^\circ$ ,  $\beta = 93.852(4)^\circ$ ,  $\gamma = 97.901(7)^\circ$ ,  $V = 1212.6(2) \text{ \AA}^3$ . Index ranges:  $-8 \leq h \leq 8$ ,  $-11 \leq k \leq 8$ ,  $-20 \leq l \leq 20$ . Range for data collection from  $2.10^\circ$  to  $25.00^\circ$ ,  $Z = 2$ .  $D_{\text{calc}} = 1.390 \text{ mg/m}^3$ .  $F(000) = 528$ . Reflections collected 6908. Independent reflections 4135 [ $R(\text{int}) = 0.0374$ ]. Data/restraints/parameters 4135:16:360. Refinement method: Full matrix least squares on  $F^2$ . Goodness-of-fit on  $F^2$ : 1.001. Final  $R$  indices [ $I > 2(I)$ ]:  $R_1 = 0.0781$ ,  $wR_2 = 0.1818$ .  $R$  indices (all data):  $R_1 = 0.0972$ ,  $wR_2 = 0.1918$ . Largest diff. peak and hole: 0.639 and  $-0.459 \text{ e \AA}^{-3}$ .

#### 4.2.5. Crystal data and structure refinement parameters for 17h

Empirical formula:  $\text{C}_{25}\text{H}_{26}\text{F}_3\text{N}_3\text{O}_6$  formula weight: 521.49. Crystal system: monoclinic. Space group:  $P2_1/c$ . Unit cell dimensions:  $a = 13.675(4)$ ,  $b = 6.202(8)$ ,  $c = 30.779(17) \text{ \AA}$ ,  $\beta = 115.42(3)^\circ$ ,  $V = 2358(3) \text{ \AA}^3$ . Index ranges:  $-14 \leq h \leq 14$ ,  $-6 \leq k \leq 6$ ,  $-32 \leq l \leq 31$ . Range for data collection from  $1.65^\circ$  to  $22.06^\circ$ ,  $Z = 4$ .  $D_{\text{calc}} = 1.469 \text{ mg/m}^3$ .  $F(000) = 1088$ . Reflections collected 4516. Independent reflections 2242 [ $R(\text{int}) = 0.1700$ ]. Data/restraints/parameters 2242:228:316. Refinement method: Full matrix least squares on  $F^2$ . Goodness-of-fit on  $F^2$ : 1.133. Final  $R$  indices [ $I > 2(I)$ ]:  $R_1 = 0.2046$ ,  $wR_2 = 0.3531$ .  $R$  indices (all data):  $R_1 = 0.2967$ ,  $wR_2 = 0.3908$ . Largest diff. peak and hole: 0.516 and  $-0.414 \text{ e \AA}^{-3}$ .

#### 4.2.6. Crystal data and structure refinement parameters for 3-[1-oxo-6-(4-piperidinium-4-yl-benzoylamino)-1,3-dihydroisoindol-2-yl]propionate dihydrate (17h-TfaOH·2H<sub>2</sub>O)

Empirical formula:  $\text{C}_{23}\text{H}_{29}\text{N}_3\text{O}_6$  formula weight: 443.49. Crystal system: triclinic. Space group:  $P-1$ . Unit cell dimensions:  $a = 7.8610(4)$ ,  $b = 10.0294(4)$ ,  $c = 15.0560(5) \text{ \AA}$ ,  $\alpha = 96.388(2)^\circ$ ,  $\beta = 99.306(2)^\circ$ ,  $\gamma = 111.4031(18)^\circ$ ,  $V = 1071.73(8) \text{ \AA}^3$ . Index ranges:  $-9 \leq h \leq 9$ ,  $-12 \leq k \leq 12$ ,  $-18 \leq l \leq 18$ . Range for data collection from  $2.22^\circ$  to  $25.49^\circ$ ,  $Z = 2$ .  $D_{\text{calc}} = 1.374 \text{ mg/m}^3$ .  $F(000) = 472$ . Reflections collected 7408. Independent reflections 3961 [ $R(\text{int}) = 0.0525$ ]. Data/restraints/parameters 3961:0:310. Refinement method: Full matrix least squares on  $F^2$ . Goodness-of-fit on  $F^2$ : 1.007. Final  $R$  indices [ $I > 2(I)$ ]:  $R_1 = 0.0696$ ,  $wR_2 = 0.1778$ .  $R$  indices (all data):  $R_1 = 0.0923$ ,  $wR_2 = 0.1913$ . Largest diff. peak and hole: 0.400 and  $-0.284 \text{ e \AA}^{-3}$ .

#### 4.2.7. Crystal data and structure refinement parameters for 3-[1-oxo-6-(4-piperazinium-1-yl-benzoylamino)-1,3-dihydroisoindol-2-yl]propionic acid trifluoroacetate hydrate (17i·H<sub>2</sub>O)

Empirical formula:  $\text{C}_{24}\text{H}_{27}\text{F}_3\text{N}_4\text{O}_7$  formula weight: 540.50. Crystal system: triclinic. Space group:  $P-1$ . Unit cell dimensions:  $a = 8.9199(5)$ ,  $b = 11.8606(9)$ ,  $c = 12.0078(7) \text{ \AA}$ ,  $\alpha = 92.329(5)^\circ$ ,  $\beta = 93.603(5)^\circ$ ,  $\gamma = 106.407(6)^\circ$ ,  $V = 1213.97(13) \text{ \AA}^3$ . Index ranges:  $-6 \leq h \leq 10$ ,  $-14 \leq k \leq 13$ ,  $-14 \leq l \leq 14$ . Range for data collection from  $3.02^\circ$  to  $25.05^\circ$ ,  $Z = 2$ .  $D_{\text{calc}} = 1.479 \text{ mg/m}^3$ .  $F(000) = 564$ . Reflections collected 7655. Independent reflections 4295 [ $R(\text{int}) = 0.0251$ ]. Data/restraints/parameters 4295:109:383. Refinement method: Full matrix least squares on  $F^2$ . Goodness-of-fit on  $F^2$ : 1.004. Final  $R$  indices [ $I > 2(I)$ ]:  $R_1 = 0.0507$ ,  $wR_2 = 0.1147$ .  $R$  indices (all data):  $R_1 = 0.0923$ ,  $wR_2 = 0.1241$ . Largest diff. peak and hole: 0.495 and  $-0.278 \text{ e \AA}^{-3}$ .

#### 4.2.8. Crystal data and structure refinement parameters for 3-{1-oxo-6-[(1,2,3,4-tetrahydroisoquinolinium-7-carbonyl)amino]-1,3-dihydroisoindol-2-yl]propionic acid chloride hydrate (17j·H<sub>2</sub>O)

Empirical formula:  $\text{C}_{21}\text{H}_{22}\text{ClN}_3\text{O}_5$  formula weight: 431.87. Crystal system: monoclinic. Space group:  $C2/c$ . Unit cell dimensions:  $a = 46.068(7)$ ,  $b = 5.0936(7)$ ,  $c = 18.848(3) \text{ \AA}$ ,  $\beta = 109.607(13)^\circ$ ,  $V = 4166.3(11) \text{ \AA}^3$ . Index ranges:  $-54 \leq h \leq 54$ ,  $-6 \leq k \leq 5$ ,  $-22 \leq l \leq 18$ . Range for data collection from  $2.98^\circ$  to  $25.05^\circ$ ,  $Z = 8$ .  $D_{\text{calc}} = 1.377 \text{ mg/m}^3$ .  $F(000) = 1808$ . Reflections collected 6442. Independent reflections 3648 [ $R(\text{int}) = 0.0647$ ]. Data/restraints/parameters 3648:2:290. Refinement method: Full matrix least squares on  $F^2$ . Goodness-of-fit on  $F^2$ : 1.004. Final  $R$  indices [ $I > 2(I)$ ]:  $R_1 = 0.0500$ ,  $wR_2 = 0.0917$ .  $R$  indices (all data):  $R_1 = 0.2115$ ,  $wR_2 = 0.1289$ . Largest diff. peak and hole: 0.209 and  $-0.279 \text{ e \AA}^{-3}$ .

#### 4.2.9. Crystal data and structure refinement parameters for 4-[1-oxo-6-(4-piperidinium-4-yl-benzoylamino)-1,3-dihydroisoindol-2-yl]butyric acid trifluoroacetate hydrate (17t·H<sub>2</sub>O)

Empirical formula: C<sub>26</sub>H<sub>30</sub>F<sub>3</sub>N<sub>3</sub>O<sub>7</sub> formula weight: 553.53. Crystal system: triclinic. Space group: *P*-1. Unit cell dimensions: *a* = 7.1440(6), *b* = 9.3970(6), *c* = 20.6680(16) Å,  $\alpha$  = 84.549(5),  $\beta$  = 88.466(5),  $\gamma$  = 69.296(5)°, *V* = 1291.98(17) Å<sup>3</sup>. Index ranges:  $-7 \leq h \leq 7$ ,  $-10 \leq k \leq 10$ ,  $-22 \leq l \leq 22$ . Range for data collection from 1.98° to 22.50°, *Z* = 2. *D*<sub>calc</sub> = 1.423 mg/m<sup>3</sup>. *F*(000) = 580. Reflections collected 5446. Independent reflections 3288 [*R*(int) = 0.0838]. Data/restraints/parameters 3288:1:362. Refinement method: Full matrix least squares on *F*<sup>2</sup>. Goodness-of-fit on *F*<sup>2</sup>: 1.011. Final *R* indices [*I* > 2(*I*): *R*<sub>1</sub> = 0.0887, *wR*<sub>2</sub> = 0.1985. *R* indices (all data): *R*<sub>1</sub> = 0.1432, *wR*<sub>2</sub> = 0.2232. Largest diff. peak and hole: 0.345 and -0.342 e Å<sup>-3</sup>.

#### Acknowledgments

This article is dedicated to the memory of Dr. Yurii A. Simonov, who took active part in this research.

The authors are grateful to Dr. V.V. Polovinko (Enamine LTD, Ukraine) for the <sup>1</sup>H NMR spectra. We also thank Dr. A.V. Mazepa and Dr. I.M. Rakipov (A.V. Bogatsky Physico-Chemical Institute of the National Academy of Sciences of Ukraine) for the Mass spectra. T.M. Khristova thanks the French Embassy in Ukraine for a PhD fellowship.

#### Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.bmc.2013.05.019>. These data include MOL files and InChIKeys of the most important compounds described in this article.

#### References and notes

- Tello-Montoliu, A.; Jover, E.; Rivera, J.; Valdés, M.; Angiolillo, D. J.; Marín, F. *Curr. Med. Chem.* **2012**, *19*, 406.
- Wang, X.; Dorsam, R. T.; Lauver, A.; Wang, H.; Barbera, F. A.; Gibbs, S.; Varon, D.; Savion, N.; Friedman, S. M.; Feuerstein, G. Z. *J. Pharmacol. Exp. Ther.* **2002**, *303*, 1114.
- Clappers, N.; Brouwer, M. A.; Verheugt, F. W. *Heart* **2007**, *93*, 258.
- Freson, K.; Thys, C.; Wittevrongel, C.; Van Geet, C. *Mini Rev. Med. Chem.* **2006**, *6*, 719.
- Krakat, J.; Mousa, S.; Root, R.; Mousa, S. S. *Drugs Future* **2009**, *34*, 27.
- Berger, P. B. *J. Am. Coll. Cardiol.* **2010**, *56*, 476.
- Zhu, J.; Choi, W.-S.; McCoy, J. G.; Negri, A.; Zhu, J.; Naini, S.; Li, J.; Shen, M.; Huang, W.; Bougie, D.; Rasmussen, M.; Aster, R.; Thomas, C. J.; Filizola, M.; Springer, T. A.; Coller, B. S. *Sci. Transl. Med.* **2012**, *4*, 125ra32.
- Andronati, S. A.; Karaseva, T. L.; Krysko, A. A. *Curr. Med. Chem.* **2004**, *11*, 1183.
- Perdih, A.; Dolenc, M. S. *Curr. Med. Chem.* **2010**, *17*, 2371.
- Krysko, A. A.; Samoylenko, G. V.; Polishchuk, P. G.; Andronati, S. A.; Kabanova, T. A.; Khristova, T. M.; Kuz'min, V. E.; Kabanov, V. M.; Krysko, O. L.; Varnek, A. A.; Grygorash, R. Y. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 5971.
- Bera, M.; Musie, G. T.; Powell, D. R. *Inorg. Chem. Commun.* **2008**, *11*, 293.
- Wilczyńska, D.; Kosson, P.; Kwasiborska, M.; Ejchart, A.; Olma, A. *J. Pept. Sci.* **2009**, *15*, 777.
- Klein, S. I.; Molino, B. F.; Czekaj, M.; Gardner, C. J.; Chu, V.; Brown, K.; Sabatino, R. D.; Bostwick, J. S.; Kasiewski, C.; Bentley, R.; Windisch, V.; Perrone, M.; Dunwiddie, C. T.; Leadley, R. J. *J. Med. Chem.* **1998**, *41*, 2492.
- Frank, H.; Guenter, L.; Volkhard, A.; Helmut, P.; Thomas, M.; Johannes, W.; Brian, G. Eur. Patent 0,604,800, 1993.
- Duggan, M. E.; Egbertson, M. S.; Hartman, G. D.; Young, S. D.; Ihle, N. C. U.S. Patent 5,854,245, 1998.
- Malovichko, O. L.; Krysko, A. A.; Kabanova, T. A.; Andronati, S. A.; Grishkovets, V. I.; Kachala, V. V.; Panov, D. A. *Med. Chem.* **2009**, *5*, 158.
- Born, G. V. R. *Nature* **1962**, *194*, 927.
- Hantgan, R. *Biochim. Biophys. Acta.* **1987**, *927*, 55.
- Xia, Z.; Wong, T.; Liu, Q.; Kasirer-Friede, A.; Brown, E.; Frojmnvic, M. M. *Br. J. Haematol.* **1996**, *93*, 204.
- FlexX, 2.1.2; BioSolveIT GmbH: Sankt Augustin, Germany.
- (a) <http://dx.doi.org/10.2210/pdb2vdm/pdb>; (b) Xiao, T.; Coller, B. S.; Wang, J.-H.; Springer, T. A. *Nature* **2004**, *432*, 59.
- Sheldrick, G. M. *Acta Cryst.* **2008**, *A64*, 112.
- Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. *J. Appl. Cryst.* **2008**, *41*, 466.

## **APPENDIX 3.**

### **RÉSUMÉ DE LA THÈSE DE DOCTORAT**

## INTRODUCTION

Les maladies thrombotiques sont une cause de forte mortalité dans le monde entier. Le processus thrombogène est complexe et comprend de multiples étapes. De nombreux systèmes de récepteurs sont impliqués dans la thrombose pathogène. Les récepteurs à la surface des plaquettes participent activement à ce processus, en particulier le récepteur du fibrinogène (intégrine  $\alpha_{IIb}\beta_3$ ) et le récepteur de la thromboxane A2. Le premier est responsable de l'interaction des plaquettes avec le fibrinogène pour former des caillots, et le second est responsable de l'activation des plaquettes par l'un des agonistes excrétés par les plaquettes adjacentes – la thromboxane A2.

Les antagonistes classiques du récepteur  $\alpha_{IIb}\beta_3$ , les peptidomimétiques RGD, ont prouvé leur capacité à réduire le risque d'infarctus du myocarde péri-interventionnel et de revascularisation en urgence du vaisseau cible pendant la cathétérisation et ont revendiqué une place en thérapie pour ces indications. De nos jours trois antagonistes du récepteur  $\alpha_{IIb}\beta_3$  sont des médicaments commerciaux : l'abciximab, l'éptifibatide et le tirofiban. Ils sont assez onéreux et présentent d'importants effets secondaires, tels que la thrombocytopénie et l'hémorragie, ce qui incite les chercheurs à développer de nouveaux antagonistes efficaces du récepteur  $\alpha_{IIb}\beta_3$ .

La thrombocytopénie est associée avec les changements conformationnels des récepteurs  $\alpha_{IIb}\beta_3$ , qui sont induits par la liaison de nouveaux antagonistes. Récemment, un nouvel antagoniste non classique RUC-2 du récepteur  $\alpha_{IIb}\beta_3$  a été développé. Contrairement aux médicaments susmentionnés, qui se lient à la forme ouverte du récepteur  $\alpha_{IIb}\beta_3$ , le nouveau ligand se lie à la forme fermée. Cela n'induit pas de changements conformationnels pour la protéine, et par conséquent réduit le risque d'effets indésirables. Ainsi, le développement de ligands ayant un mécanisme de liaison similaire à RUC-2 représente un moyen prometteur de créer de nouveaux agents antithrombotiques.

Les antagonistes du récepteur de la thromboxane A2 (TP) pourraient être particulièrement utiles pour le traitement de l'ischémie myocardique aiguë et l'insuffisance cardiaque, et présentent des effets cardioprotecteurs. Cependant, jusqu'à présent, aucun antagoniste du récepteur TP n'est disponible sur le marché. Tous les agents étudiés sont moins efficaces que l'aspirine. Cette dernière n'est pas un antagoniste du récepteur TP mais cible la synthèse du thromboxane A2 des plaquettes par l'acétylation irréversible des cyclooxygénases 1 et 2. L'aspirine présente quelques effets secondaires : toxicité gastro-intestinale, résistance,

événements hémorragiques, essentiellement saignement gastro-intestinal mineur ou généralisé, et accidents vasculaires cérébraux. Au niveau pharmacologique, les antagonistes du récepteur de la thromboxane A2 pourraient être plus avantageux que de l'aspirine à faible dose. En effet, les antagonistes du récepteur de la thromboxane A2 inhibent les effets délétères d'autres ligands endogènes du récepteur de la thromboxane A2, tels que les endoperoxydes, les prostanoïdes et les isoprostanes, dont la formation ne serait pas affectée par l'aspirine, d'autres inhibiteurs COX ou inhibiteurs de la thromboxane synthase.

Le but de cette étude est la création assistée par ordinateur de nouveaux antagonistes classiques et non classiques du récepteur  $\alpha_{IIb}\beta_3$ , et de nouveaux antagonistes du récepteur TP. Pour cela, nous avons utilisé un ensemble d'approches informatisées modernes – QSAR, modélisation par pharmacophore, amarrage (*docking*) moléculaire, et analyse de similarité de champ moléculaire et forme moléculaire intégrés dans un flux de travail de criblage virtuel.

Notre étude a abouti sur la suggestion de nouveaux ligands potentiels capables de se lier à la forme ouverte ou fermée de l'intégrine  $\alpha_{IIb}\beta_3$  et du récepteur de la thromboxane A2. Nos antagonistes de l'intégrine  $\alpha_{IIb}\beta_3$  ont été synthétisés et testés expérimentalement à l'Institut de Physico-Chimie A.V. Bogatsky de l'Académie Nationale des Sciences d'Ukraine (PCI). Les expériences effectuées montrent que certains des composés synthétisés sont plus efficaces que le tirofiban – un médicament commercialisé.

Le manuscrit est divisé en quatre parties. La première présente un examen de la littérature existante décrivant le mécanisme de la thrombose, les agents connus de la thérapie antithrombotique et les études de modélisation déjà effectuées. La seconde partie décrit les méthodes de criblage virtuel utilisées dans cette étude. Les deux dernières parties exposent le développement assisté par ordinateur de nouveaux antagonistes du récepteur  $\alpha_{IIb}\beta_3$  et des antagonistes du récepteur de la thromboxane A2.

## **PARTIE 1 : ÉTUDE DE LA LITTÉRATURE EXISTANTE**

La première section décrit le mécanisme de la thrombose. Une attention particulière est donnée à l'étape d'activation des plaquettes impliquant les récepteurs de la thromboxane A<sub>2</sub>, et à la formation du thrombus via l'interaction de l'intégrine  $\alpha_{\text{IIb}}\beta_3$  avec le fibrinogène.

Dans la seconde section, nous analysons les études de modélisation décrites dans la littérature. La plupart des modèles de pharmacophore et QSAR développés ont été créés à partir de petits ensembles congénériques de composés, et n'ont pas été validés correctement. De plus, dans la plupart des cas ils n'étaient pas bien détaillés, ce qui rend leur application au criblage virtuel impossible.

## **PARTIE 2 : MÉTHODES DE CRIBLAGE VIRTUEL**

Dans cette partie nous considérons différentes approches informatisées pour le criblage virtuel ainsi que les logiciels utilisés dans ce travail : QSAR (ISIDA, SiRMS), similarité de forme moléculaire (ROCS) et de champ moléculaire (FieldAlign), pharmacophores (LigandScout), docking (PLANTS, FlexX et MOE), et profilage pharmacologique (PASS). Nous donnons aussi des informations sur les méthodes d'apprentissage automatique (Random Forest) et les descripteurs moléculaires (SiRMS, ISIDA/SMF et les triplets pharmacophoriques flous dépendants du pH (ISIDA/FPT)) utilisés dans la modélisation QSAR.

## **PARTIE 3 : CONCEPTION D'ANTAGONISTES DU RÉCEPTEUR $\alpha_{\text{IIb}}\beta_3$**

Cette partie décrit la conception assistée par ordinateur de nouveaux antagonistes du récepteur du fibrinogène. Deux scénarios différents sont considérés : lier soit la forme ouverte soit la forme fermée de l'intégrine  $\alpha_{\text{IIb}}\beta_3$  (sections 3.1 and 3.2 respectivement). À la première



étape de chaque étude, les données expérimentales à disposition ont été utilisées pour développer des modèles de pharmacophore et QSAR, et des modèles de similarité de forme moléculaire et de champ moléculaire. Des antagonistes potentiels de l'intégrine ont été sélectionnés dans une base de données contenant des structures de composés réels ou générés par ordinateur, en utilisant un flux de travail de criblage virtuel qui combinait différentes méthodes de calcul.

### 3.1. Conception de ligands de la forme ouverte de l'intégrine $\alpha_{\text{IIb}}\beta_3$ .

**Données à disposition.** Deux jeux de données de peptidomimétiques RGD contenant les données sur leur affinité pour  $\alpha_{\text{IIb}}\beta_3$  (45 composés) ou leur activité antiagrégante (53 composés) ont été fournis par PCI. Puisque les deux jeux de données contenaient un nombre de composés relativement peu important et étaient significativement déséquilibrés, ils ont été augmentés de composés provenant de la base de données ChEMBL. Les jeux de données résultants, comprenant 338 composés associés à des valeurs d'affinité pour  $\alpha_{\text{IIb}}\beta_3$  et 453 composés associés à une activité antiagrégante, ont été utilisés par la suite pour la modélisation QSAR et pour la modélisation par pharmacophore.

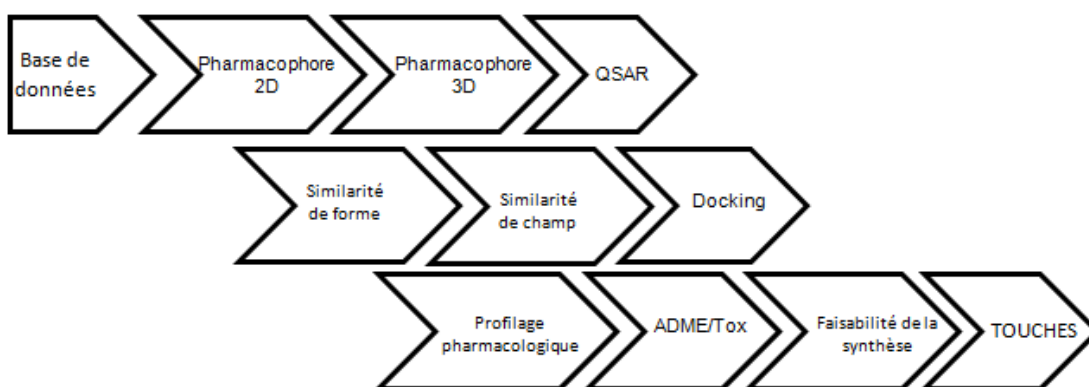
Trois complexes formés de la forme ouverte de  $\alpha_{\text{IIb}}\beta_3$  et de trois différents ligands ont été sélectionnés dans la base de données PDB. Leurs structures ont été utilisées pour le docking ligand-à-protéine. Trois programmes de docking ont été testés : PLANTS, FlexX, et MOE. Le programme MOE s'est révélé le meilleur d'après les expériences de re-docking et de cross-docking.

**Étape de modélisation.** Les modèles QSAR 2D ont été construits pour l'affinité pour  $\alpha_{\text{IIb}}\beta_3$  ainsi que pour l'activité antiagrégante en utilisant la méthode Random Forest avec trois types de descripteurs (SiRMS, ISIDA/SMF et ISIDA/FPT). Malgré le bruit significatif présent dans les données, les modèles obtenus montrent une performance prédictive raisonnable en validation croisée à 5 échantillons :  $R^2_{\text{cv}} = 0.76$ ,  $\text{RMSE}_{\text{cv}} = 0.72$  (affinité pour  $\alpha_{\text{IIb}}\beta_3$ ) et  $R^2_{\text{cv}} = 0.54$ ,  $\text{RMSE}_{\text{cv}} = 0.74$  (activité antiagrégante).

Les modèles de pharmacophore 3D (basés soit sur la structure du récepteur soit sur le ligand) ont été développés avec le logiciel LigandScout. La validation sur le jeu de test externe a montré que les modèles basés sur la structure du récepteur n'étaient pas assez sélectifs. Les modèles basés sur la structure des ligands étaient bien plus performants (rappel 0,27 et précision 0,92), et ont donc été utilisés par la suite pour le criblage virtuel. Les deux modèles sélectionnés contiennent des fonctions anioniques et cationiques séparées par 15-16 Å, ce qui peut être couvert par au moins 12 liaisons. Cette information a été utilisée pour construire un pharmacophore 2D utilisé pour le criblage rapide de grandes bases de données.

Les modèles pour l'analyse de similarité de forme moléculaire et de champ moléculaire ont été préparés à partir de structures 3D du tirofiban et de L-739758 issues de leurs complexes avec le récepteur  $\alpha_{IIb}\beta_3$ . Leur application au jeu de test externe a abouti sur des paramètres statistiques raisonnables: ROC AUC = 0.64 et 0.68 pour la similarité de forme et de champ respectivement.

Tous les modèles développés dans cet ouvrage ont été intégrés au flux de travail pour le criblage virtuel destiné à sélectionner les ligands les plus efficaces de la forme ouverte du récepteur  $\alpha_{IIb}\beta_3$  (Figure 1). Les modèles décrits précédemment dans la littérature ont également été utilisés pour établir le profil pharmacologique et quelques propriétés ADME/Tox de certaines molécules sélectionnées.

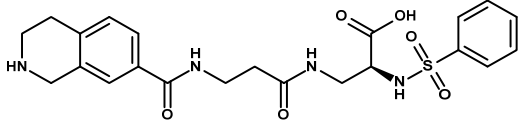
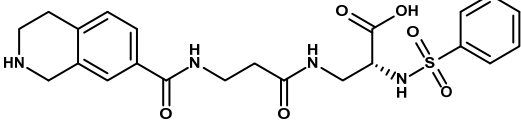
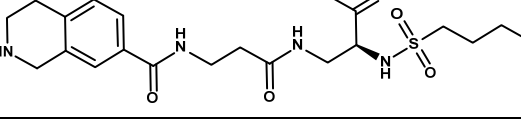
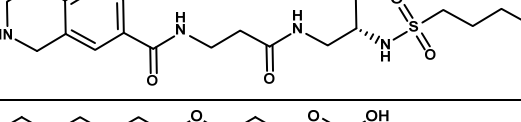
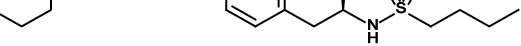


**Figure 1.** Flux de travail pour le criblage virtuel combinant les méthodes basées sur la structure du récepteur et celles basées sur la structure des ligands, utilisé pour découvrir des molécules se liant à la forme ouverte du récepteur  $\alpha_{IIb}\beta_3$ .

Criblage virtuel et conception assistée par ordinateur. De façon surprenante, le criblage avec l'ensemble des modèles décrits dans la Figure 1 de la base de données BioInfoDB contenant environ 3 millions de composés chimiques disponibles sur le marché n'a retourné

aucune touche ou *hit*. En conséquence, nous avons utilisé des mimétiques appropriés Asp et Arg avec des fragments "*linkers*" pour générer une librairie combinatoire de 6930 composés. Son criblage a donné deux touches, chacune desquelles étant représentée par des énantiomères S et R. Les quatre composés ont été synthétisés et leur activité biologique mesurée expérimentalement dans l'Institut de Physico-Chimie A. V. Bogatsky d'Odessa. L'expérience a confirmé le pouvoir antithrombotique de ces composés. Ainsi, I-S et I-R sont même plus puissants que le médicament de référence, le tirofiban (Table 1). Dans les deux cas les énantiomères S, avec un plus haut score en docking, sont plus puissants que les énantiomères R correspondants.

**Table 1.** Affinité pour  $\alpha_{IIb}\beta_3$  et activité antiagrégante du tirofiban (médicament de référence) et des composés mis au point dans ce travail.

		Affinité (pIC <sub>50</sub> )		Antiagrégation (pIC <sub>50</sub> )	
		pred	exp	pred	exp
I-S		8.32	9.66	7.37	8.21
I-R		8.32	9.02	7.37	7.60
II-S		8.24	7.21	7.27	6.49
II-R		8.24	7.10	7.27	6.17
Tirofiban			8.62		7.48

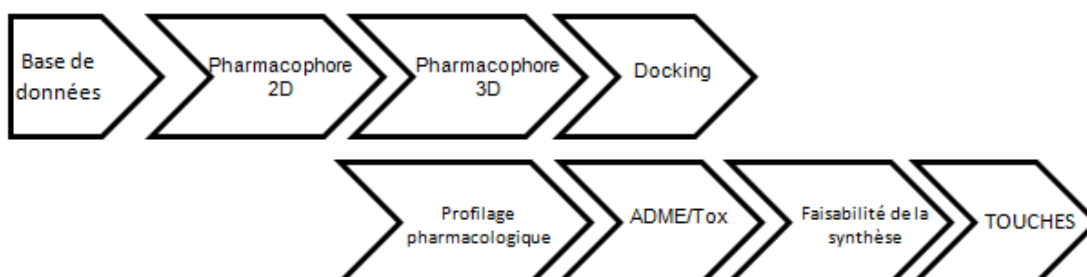
### 3.2. Conception de ligands de la forme fermée de l'intégrine $\alpha_{IIb}\beta_3$ .

**Données disponibles.** Seules quelques données sont disponibles pour les ligands récemment découverts de la forme fermée de  $\alpha_{IIb}\beta_3$ . Le complexe de la forme fermée de  $\alpha_{IIb}\beta_3$  avec le ligand

RUC-2 a été tiré de la base de données PDB. Il a été utilisé pour la modélisation de pharmacophore basée sur la structure de la protéine et le docking protéine-ligand. En se basant sur les expériences de re-docking, le programme FlexX a été choisi, s'étant révélé un meilleur outil de docking que le programme MOE.

**Étape de modélisation.** Les modèles de pharmacophore basés sur la structure 3D de la protéine ont été obtenus avec le programme LigandScout. Ces modèles présentent deux centres cationiques communs séparés de 16 Å. Cette information a été utilisée pour préparer des pharmacophores 2D contenant deux atomes chargés positivement séparés de 12 liaisons. L'application d'un modèle de pharmacophore 2D a accéléré de manière significative la première étape du criblage virtuel.

Puisqu'aucune donnée sur la liaison à la forme fermée du récepteur  $\alpha_{11b}\beta_3$  d'analogues de RUC-2 ne sont disponibles, nous avons dû faire l'impasse sur la plupart des approches basées sur la structure des ligands dans le flux de travail pour le criblage virtuel (Figure 2).

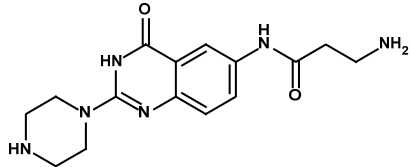
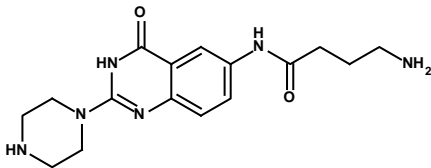
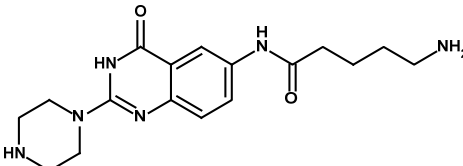
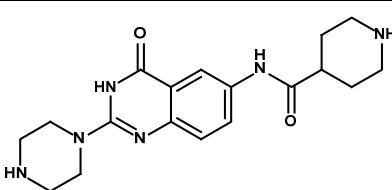


**Figure 2.** Flux de travail pour le criblage virtuel utilisé pour trouver des molécules se liant à la forme fermée du récepteur  $\alpha_{11b}\beta_3$ .

**Criblage virtuel et conception assistée par ordinateur.** Le criblage de trois bases de données (les bases de données Enamine Advanced et HTS, la base de données Enamine REAL, et la base de données ZINC) couvrant près de 35 millions de composés a permis de trouver deux touches qui sont passées par tous les filtres de criblage. L'une d'entre elles était un composé connu, le nafamostat, avec des propriétés antiagrégantes déjà décrites, tandis que la seconde n'était pas disponible sur le marché.

Ainsi, une petite librairie combinatoire d'analogues du ligand RUC-2 a été générée en utilisant les informations récupérées des modèles de pharmacophore et du docking moléculaire. Les quatre composés ont une forte affinité pour  $\alpha_{11b}\beta_3$  ( $< 10$  nM, cf. Table 2).

**Table 2.** Affinité pour le récepteur  $\alpha_{11b}\beta_3$  des analogues de RUC-2 conçus *in silico*.

		Affinité pour $\alpha_{11b}\beta_3$ (pIC <sub>50</sub> )
1		8.30
2		8.66
3		8.85
4		8.42

#### **PARTIE 4 : CONCEPTION D'ANTAGONISTES DU RÉCEPTEUR DE LA THROMBOXANE A2**

**Données disponibles.** Les deux jeux de données d'antagonistes de la thromboxane A2 contenant leur affinité pour le récepteur de la thromboxane A2 (TP) (174 composés) ou leur activité antiagrégante (93 composés) ont été extraits de la base de données ChEMBL. Aucune structure cristallographique expérimentale du récepteur de la thromboxane A2 ou de ses complexes n'est disponible. Par conséquent, seules les méthodes basées sur la structure des ligands ont été

utilisées : la modélisation par pharmacophore et QSAR, et l'analyse de forme et champ moléculaires.

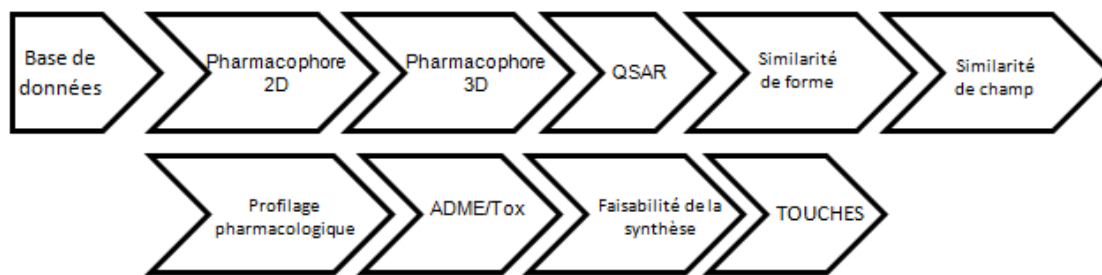
**Étape de modélisation.** Des modèles QSAR ont été construits pour l'affinité pour TP (modèle de régression) et l'activité antiagrégante (modèle de classification binaire) en utilisant la méthode Random Forest avec trois types de descripteurs (SiRMS, ISIDA/SMF et ISIDA/FPT). Les modèles obtenus donnent une performance prédictive raisonnable en validation croisée à 5 échantillons :  $R^2_{cv} = 0.67$ ,  $RMSE_{cv} = 0.61$  (affinité pour TP) et BA (*balanced accuracy*) = 0.81 avec une spécificité de 0.83 et une sensibilité de 0.79 (activité antiagrégante).

Des modèles de pharmacophore basés sur la structure 3D des ligands ont été développés avec le programme LigandScout. Le meilleur modèle donne une précision élevée de 0.96 et un rappel modéré de 0.30 sur tous les composés du jeu de test externe. Il contient un centre anionique, deux centres accepteurs de liaison hydrogène et deux centres hydrophobes. Comme dans les sections précédentes, un pharmacophore 2D a été développé à partir de ce pharmacophore 3D.

Les conformères des composés actifs du jeu de données contenant les valeurs d'affinités qui correspondent au modèle de pharmacophore 3D ont été pré-alignés dans MOE puis utilisés dans le développement d'un modèle de forme moléculaire avec le programme ROCS (OpenEye). Toutes les combinaisons de deux ou trois composés ont été considérées pour la génération des modèles suivie par l'application de chaque modèle au jeu de test externe. Le meilleur modèle basé sur les trois composés (ROC AUC = 0.80) a été sélectionné pour le criblage virtuel.

L'un des composés actifs à conformation rigide du jeu de données d'affinité a été choisi comme référence pour l'analyse de similarité de champ moléculaire. Le modèle de champ moléculaire obtenu, offrant de bonnes performances sur le jeu de test (ROC AUC = 0.79), a été utilisé pour le criblage virtuel.

Les modèles développés dans cet ouvrage ainsi que des modèles décrits précédemment dans la littérature ont été intégrés dans un flux de travail pour le criblage virtuel, illustré par la Figure 3.



**Figure 3.** Flux de travail pour le criblage virtuel utilisé pour découvrir de nouveaux antagonistes du récepteur de la thromboxane A2.

**Criblage virtuel et conception assistée par ordinateur.** Le criblage de la base de données BioInfoDB pour trouver de nouveaux ligands pour TP n'a pas permis de trouver de touche. La conception de nouveaux composés a été entreprise en utilisant deux approches différentes. La première consistait à modifier manuellement quelques-uns des fragments structuraux d'antagonistes de TP présentant une forte activité, ce qui permit de créer 52 nouveaux composés avec une grande similarité de champ moléculaire avec les composés parents. La seconde approche concerne le modèle de pharmacophore 3D obtenu dans l'étape de modélisation. Ce modèle, impliquant 5 centres pharmacophoriques a été divisé en 2 modèles "partiaux", chacun constitué de trois centres dont un qui leur est commun. La base de données de fragments Preswick contenant 2800 fragments issus de médicaments connus, a été criblée avec les deux modèles partiels de pharmacophore. Puis, à l'aide du programme MOE, les fragments trouvés ont été liés entre eux de manière combinatoire de façon à former un ensemble de 171 composés. Le criblage de l'ensemble des 223 composés mis au point a permis d'obtenir 5 composés recommandés pour la synthèse et l'évaluation de leurs propriétés biologiques.

## CONCLUSIONS

Durant ces travaux, des antagonistes présentant une forte activité pour les formes ouverte et fermée des récepteurs  $\alpha_{11b}\beta_3$  ont été découverts en criblage virtuel, en utilisant la modélisation par pharmacophore et la modélisation QSAR, le docking moléculaire et l'évaluation d'effets pharmacologiques secondaires et de certaines propriétés ADME/Tox. Tous les composés suggérés ont été synthétisés et testés expérimentalement dans l'Institut de Physico-Chimie A. V. Bogatsky d'Odessa. L'expérience confirme les effets thrombotiques puissants des composés conçus *in silico*. Ainsi, l'un des antagonistes découverts de la forme ouverte du récepteur  $\alpha_{11b}\beta_3$  fait preuve d'une plus forte activité que le composé de référence, le tirofiban. Tous les

antagonistes mis au point pour la forme fermée du récepteur  $\alpha_{IIb}\beta_3$  ont des valeurs d'affinité comparables à celle du tirofiban, ce qui en fait des cibles d'un grand intérêt pour d'autres études biologiques et optimisations structurales.

Cinq nouveaux antagonistes de la thromboxane A2 ont été conçus en utilisant des modèles de pharmacophore et QSAR, l'analyse de similarité de forme moléculaire et champ moléculaire, ainsi que l'évaluation d'effets pharmacologiques possibles et de certaines propriétés ADME/Tox. Les molécules suggérées *in silico* sont considérées pour la synthèse et l'évaluation expérimentale de leurs propriétés biologiques.



## **PUBLICATIONS**

1. Krysko A.A., Samoylenko G.V., Polishchuk P.G., Andronati S.A., Kabanova T.A., Khristova T.M., Kuz'min V.E., Kabanov V.M., Krysko O.L., Varnek A.A. and Grygorash R.Y. "RGD mimetics containing phthalimidine fragment, novel ligands of fibrinogen receptor." *Bioorganic & Medicinal Chemistry Letters*. - 2011. - Vol.21. - P. 5971–5974
2. A.A. Krysko, G.V. Samoylenko, P.G. Polishchuk, M.S. Fonari, V.Ch. Kravtsov, S.A. Andronati, T.A. Kabanova, J. Lipkowski, T.M. Khristova, V.E. Kuz'min, V.M. Kabanov, O.L. Krysko, A.A. Varnek "Synthesis, biological evaluation, X-ray molecular structure and molecular docking studies of RGD mimetics containing phthalimidine fragment as ligands of integrin  $\alpha_{IIb}\beta_3$ ." *Bioorganic & Medicinal Chemistry* (*accepted in May 2013*)
3. T.M. Khristova, P.G. Polishchuk, A.A. Krysko, T.A. Kabanova, V.M. Kabanov, G.V. Samoylenko, V.E. Kuz'min, S.A. Andronati, T. Langer, R. Issenhardt, A. Varnek, "Development of novel antagonists of fibrinogen receptor using multistage screening approach." (*in preparation*)
4. T.M. Khristova, P.G. Polishchuk, A.A. Krysko, T.A. Kabanova, V.M. Kabanov, V.E. Kuz'min, S.A. Andronati, A. Varnek "Development of RUC-2 analogs - antagonists of  $\alpha_{IIb}\beta_3$  integrin" (*in preparation*)

## **POSTER PRESENTATIONS**

1. Khristova T.M., Polishchuk P.G., Kuzmin V.E. and Varnek A.A. "Computer-aided design of new inhibitors of platelet aggregation: RGD-peptidomimetics." 5<sup>eme</sup> Journées Nationales de Chimoinformatique. - Cabourg, 2011 - P. 36
2. Khristova T.M., Polishchuk P.G., Krysko A.A., Kuzmin V.E., Andronati S.A. and Varnek A.A. "Computer-aided design of new inhibitors of platelets aggregation: RGD-peptidomimetics." 3rd Strasbourg Summer School in Chemoinformatics. - Strasbourg, 25-29 June, 2012 - P. P47
3. Khristova T.M., Polishchuk P.G. "Development of new antagonists of thromboxane A2 receptor." XIV Conference of young scientists and student-chemists of South region of Ukraine with international participation. – Odessa, Ukraine, 25-27 April, 2012 – P.70

## **ORAL COMMUNICATIONS**

1. Khristova T.M., Volyanskaya A.V., Krysko O.L. Polishchuk P.G. "Virtual screening of "nonclassical" antagonists of fibrinogen receptor." XV Conference of young scientists and student-chemists of South region of Ukraine with international participation. – Odessa, Ukraine, 17-19 April, 2013