



BABEȘ-BOLYAI UNIVERSITY

FACULTY OF CHEMISTRY AND CHEMICAL ENGINEERING
DEPARTMENT OF ORGANIC CHEMISTRY

TOPOLOGY OF BIOACTIVITY - RESUME

Scientific advisor

Prof. Dr. Csaba Paisz

Prof. Dr. Mircea V. Diudea

Phd student:

Lungu Nicolae Claudiu

Cluj-Napoca

2019

BABEȘ-BOLYAI UNIVERSITY

FACULTY OF CHEMISTRY AND CHEMICAL ENGINEERING

DEPARTMENT OF ORGANIC CHEMISTRY

TOPOLOGY OF BIOACTIVITY

Scientific advisor

Prof Dr. Csaba Paisz

Prof. Dr. Mircea V. Diudea

Phd student: Lungu Nicolae Claudiu

Committee

President: Prof. Dr. Ioan Grosu Committee

Prof. Dr. Ionel Mangalagiu

Prof. Dr. Mihai V. Putz

Prof. Dr. Simion Simon

Foreword

In such an attempt, there are many people to whom I owe gratitude and thanks for their support and support. Without all these people around me, I could not imagine taking this thesis to the end.

I want to thank my mentor, Prof. Dr. Mircea V. Diudea, for the detailed comments and constructive criticism. His guidance helped me throughout this period of research and writing of this thesis.

I would also like to thank the doctoral committee for encouragement, comments, information, and helpful questions.

Contents

Keywords.....	9
THEORETICAL PART.....	10
GENERAL INTRODUCTION.....	10
Chapter I.1.....	12
Molecular recognition.....	12
Chapter I.2.....	12
Analysis of protein-ligand interactions.....	12
I.2.1 EC50 measurement.....	12
I.2.2 Competitive linkage and IC50.....	13
Chapter I.3.....	14
„QSAR " (activity structure relations).....	14
I.3.1 Fragment-based 2D-QSAR.....	14
I.3.2 3D-QSAR.....	14
I.3.3 Multi-dimensional QSAR methods.....	14
I.3.4 Characteristics of receiver selection.....	14
Chapter I.4.....	15
Protein docking and folding.....	15
I.4.1. Rigid docking and receptor docking.....	15
I.4.2. Flexible ligand docking and rigid receiver.....	15
I.4.3 Flexible ligand docking and flexible receptor docking.....	16
I.4.4 Folding of proteins.....	16
I.4.5 Tertiary structure of proteins.....	16
I.4.6 Quaternary structure of proteins.....	16
I.4.7 Modeling of protein folding.....	16

Chapter I.5.....	17
Molecular descriptors.....	17
I.5.1. Basic definitions in graphs	17
I.5.2. Square topological matrices and indices.....	17
I.5.2.1. The adjacency matrix.....	17
I.5.2.2. The distance matrix	17
I.5.2.3. The die matrix.....	18
I.5.2.4. Cluj matrix.....	18
I.5.2. 5. Task matrices - "CH"	19
I.5.2.6. Layer arrays	20
Personal contributions Chapter II	21
II.1Molecular adaptability of the ligand	21
PEI (polyethyleneimines) linear and branched and the space of properties	21
II.1.1Introduction	21
II.1.2 Computational methods.....	21
II.1.3 Results	22
I.1.4 Conclusions	23
Chapter 2.....	24
Molecular adaptability of the receptor	24
Molecular adaptability of flavin adenine dinucleotide (FAD) among adjacent amino acids and its catalytic role in GOx (gluco oxidase).....	24
II.2.1 Introduction	24
II.2.3 Computational methods.....	25
II.2.3 Results	25
II.2.4 Conclusions	26
Chapter 3.....	27

Predicting power	27
The potency and prediction of the binding site of teixobactin and other lipids II ligands through the statistical baseline scores of the conformational spatial maps.....	27
II.3.1 Introduction	27
II.3.2 Computational methods.....	28
II.3.3 Results	28
Chapter 4.....	29
Prediction and representation of the biological surface	29
Modeling of ligands in rigid docking of MraY inhibitors. The polynomial discriminant and the Laplacian operator as predictors of bioactivity.....	29
II.4.1 Introduction	29
II.4.2 Computational methods.....	30
II.4.3 Results	30
II.4.4. Conclusions	32
Chapter 5.....	32
Protein folding and interactions	32
Alpha 1 folding of antitrypsin and interaction with nanomaterials.....	32
II.5.1 Introduction	32
II.5.2 Computational methods.....	32
II.5.3 Results	33
II.5.4 Conclusions	35
Chapter 6.....	36
II.6.1 Pyridine imidazole derivatives: QSAR models.....	36
II.6.1.1 Introduction	36
II.6.1.2 Experimental methods	36
II.6.1.3 Computational methods.....	37

II.6.1.4 Results	38
II.6.1.4. Experimental results	38
II.6.1.4. QSAR models.....	40
II.6.1.5 Conclusions	40
II.6.2. Oxabutan benzamides as inhibitors of UDP-3-O-acyl-N-acetylglucosamine deacetylase: a QSAR study.....	41
II.6.2.1 Introduction	41
II.6.2.2 Computational methods.....	41
II.6.2.3 Results	42
II.6.2.4 Conclusions	43
II.6.3 Famotidine as a urokinase inhibitor: a QSAR study	43
II.6.3.1 Introduction	43
II.6.3.2 Computational methods.....	44
II.6.3.3 Results	44
II.6.3.4 Conclusions	46
II.6.4 Cyclin-dependent kinase two inhibitors: a QSAR artificial neural regression study	46
II.6.4.1 Introduction	46
II.6.4.2 Computational methods.....	47
II.6.4.3 Results	47
II.6.4.4 Conclusions	48
II.6.5 CGMP 3', 5'-cyclic 10A phosphodiesterase inhibitors: a QSAR study.....	48
II.6.5.1 Introduction	48
II.6.5.2 Computational methods.....	48
II.6.5.3 Results	49
II.6.5.4 Conclusions	50
II.6.6 Exploring bioactivity - the toxicity space.....	50

A theoretical toxicity study on "friendly" biologic macromolecules	50
II.6.6.1 Introduction	50
II.6.6.2 Computational methods.....	50
II.6.6.3 Results	51
II.6.6.4 Conclusions	53
II.6.7 Exploration of the bioactivity space - antimicrobial activity	53
Penicillin-binding protein in the mechanism of resistance to <i>Staphylococcus aureus</i>	53
II.6.7.1 Introduction	53
II.6.7.2 Experimental methods.....	54
II.6.7.3 Computational methods.....	54
II.6.7.4 Results	55
II.6.7.5 Conclusions	56
General conclusions	56
List of publications	59
References.....	61

Keywords

Bioactivity

Chemical descriptors

Chemical space

Cofactor

Conformational energy

Docking

Ligand

Lipid II

Low molecular dynamics

Modeling of counterparts

Molecular dynamics

Molecular recognition

MraY

Multiple linear regression

Nanostructure

Neural networks

Pearson correlation

Pharmacophore

Receiver

Relationship structure activity

Structure-property relationship

The topological space

Toxicity

Variety (topological)

THEORETICAL PART

GENERAL INTRODUCTION

Bioactivity or biological activity underlines the property of a stimulus (radiation, molecule, biomolecules, drug) to trigger a biological response in living matter. In pharmacology, this activity is determined by the pharmacophore of ligand molecules, whose constitution and form represent the nature of the interactions between the ligand molecule and the target (respectively the receptor).

The topological space is defined as a set of points/atoms with the corresponding vicinity for each point, whose axiomatic existence is accepted (Stadler and Stadler 2002). All biological variables, which reflect either the discrete presence, the continuous presence or the interaction of biological entities - can be represented and measured in Euclidean space (Fagerland et al., 2011) Topological spaces can be described up to homeomorphism by their topological properties. These are invariable under homeomorphism. A topological space, which is locally Euclidean, is called a variety. As a Euclidean space, a variety can have n-dimensions (Körner 2015). Spaces of varieties can be weighted to "transport" certain properties, such as bioactivity. Thus, a variable describing the bioactivity can be represented using variants (topological spaces).

Pharmacological bioactivity is the result of the biodynamics and biokinetics of a particular drug; these activities can be considered to be the medicine space. The medicine space is a special region, a component of the chemical space. Chemical space is a conceptualization in chemistry, which refers to space, including all possible molecules and chemical compounds defined by a set of building rules and boundary conditions. Theoretical chemical space in cheminformatics refers to the space of potentially pharmacologically active molecules, with an order of the size of 10^{60} molecules.

The Abstract Chemical Service has in its database since July 2009, 49,037,297 organic and inorganic molecules, a number that increases on November 11, 2013, to over 75 million structures cited in journals. The exploration of the chemical space can be performed computationally - in "silico" or experimentally - by chemical reactions. This exploration of the chemical space is realized by using in the database "in silico" of the virtual molecules, visualized by designing the space of the multidimensional property in the lower dimensions, characterized by measurable/quantifiable physical-chemical properties (Vallianatou et al., 2015).

The displacement through the chemical space is achieved by generating stoichiometric combinations of electrons and atomic nuclei, in order to approach all possible topological isomers necessary for the application of the construction rules. In chemistry, such techniques include structures generating motors, fragmentation motors and more discrete techniques, such as conformational search and analysis techniques such as molecular dynamics, docking, etc. (Carlson and McCammon 1999).

In the real world, the Physico-chemical properties are the projection of the chemical properties into the known chemical space. Being only projections, these characteristics are often not unique (respectively degenerate), and different molecules can exhibit similar properties (Raymond and Awale 2012). The design of materials and the discovery of molecules with pharmacological properties involve the exploration of the chemical space.

Chapter I.1

Molecular recognition

Molecular recognition refers to a specific interaction between two molecules, which is a noncovalent one (Bolel et al., 2012). The notion of molecular recognition was introduced by Jean Marie Lehn in 1970. The recognition process is multilayered. The two partners, in the case of biology, are represented by a host or target and the guest or active molecule represented by the drug. The process involves mutual recognition, in particular, based on the complementarity (in local form and task) required to create an obligation based on functional groups.

Recognition is possible in any direction with six degrees of freedom. Natural functional groups, hydrogen donors or hydrogen acceptors are relevant for molecules to "recognize" and interact with each other. The different deletions of the functional atoms around the central atom will result in a different noncovalent recognition in space, with different geometric positioning. In short, bioactivity is the result of both binding recognition and molecular recognition. Molecular recognition determines a subsequent interaction between host (receptor) and guest (ligand).

Chapter I.2

Analysis of protein-ligand interactions

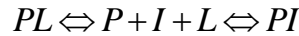
Interactions are described as unique and complicated cases such as multiple binding sites, for example, allosteric (Liu and Nussinov 2016). A large number of macromolecular structures deposited in the protein database (PDB) that describes "complexes between proteins and endogenous compounds or synthetic drugs has made it possible to analyze the interactions that occur between proteins and their ligands. The residues observed with the lowest frequency in the protein sequences, Trp, His, Met, Tyr, and Phe, were found to be the most abundant in the binding pockets.

I.2.1 EC50 measurement

The total concentration of the ligand to which half of the protein binds is called EC50 (50% effect concentration) of the reaction. EC50 is widely used to report test results because estimating the midpoint of inspection titration is simple. Halfway through the titration point, when half of the protein is bound to the ligand, we have the following relation $[P] = [PL]$ (Polla 2010).

I.2.2 Competitive linkage and IC50

In many situations, it is important to evaluate the competition between two ligands for binding to a protein. If one of these is a "reference" ligand with a known binding site and if titration in another "test" ligand diminishes the "reference" binding, it means that the test ligand either binds to the same site as the "reference" or at an allosteric site that influences the binding site of the reference.



At equilibrium, the following can be written:

$$[P] = \frac{K_d[PL]}{[L]} = \frac{K_i[PI]}{[I]}$$

then

$$\frac{[PL]}{[PI]} = \frac{K_i[L]}{K_d[I]}$$

The equation shows that the effect of the inhibitor is proportional; the higher the affinity or concentration of the free inhibitor [I], the lower the ligand ratio.

2.3 Balance

Consider a situation where a quantity of ligand is added to the protein. Initially, each protein molecule will be unbound ($[P] = [P]_T$), and encounters between ligand and protein frequently bind:

$$\frac{d[PL]}{dt} = k_{on}[P][L] - k_{off}[PL] = k_{on}([P]_T - [PL])([L]_T - [PL]) - k_{off}[PL] = k_{on}[P]_T[L]_T - (k_{on}([P]_T + [L]_T) + k_{off})[PL] + k_{on}[PL]^2$$

$$f_b(t) \approx 1 - e^{-(k_{on}([P]_T + [L]_T) + k_{off})t}$$

The time constant for approaching the equilibrium, $t_{relax} = 1 / (k_{on} ([P]_T + [L]_T) + k_{off})$, proves to be a good approximation of the real value in a wide variety of circumstances. The protein-bound fraction will usually reach more than 98% of its equilibrium value in t_{relax} and in practice, this time depends mostly on k_{on} and the total concentration of protein and ligand, because k_{off} is relatively small. Experimenting with design for balance measurements should take into account the time it takes to get there. "For a typical protein-ligand binding assay with 1 pM protein, 2 pM ligand and 1 pM K_d , the equilibrium time can vary dramatically with the kinetic properties of the protein.

Chapter I.3

„QSAR " (activity structure relations)

The quantitative relationship between structure and activity can finally be described as a theory that links "the molecular structure, through molecular descriptors, with the relevant biological activities to create regression or classification models" (Nantasenamat et al., 2009). Therefore, its importance is incalculable, because it is capable of knowing or predicting this connection "from the molecular structure to a certain biological activity that has enormous value in numerous fields of scientific research.

I.3.1 Fragment-based 2D-QSAR

"2D-QSAR methods allow the construction of models when 3D crystal objects or receiver structures are not available. Improved fragment-based QSAR methods have been introduced. The Hologram-QSAR (H-QSAR) was the first QSAR method based on molecular fragments introduced by Tripos (Leal et al., 2015). "

I.3.2 3D-QSAR

3D-QSAR methods are complex. There are two types of 3D-QSAR methods: alignment dependence and alignment independence. Both methods require bioactive conformations of ligands as templates for QSAR model development. CoMFA is one of the well-known 3D-QSAR methods developed by Cramer et al. In this method, "the 3D structure of the ligands is superimposed, then the steric and electrostatic fields are generated around each ligand.

I.3.3 Multi-dimensional QSAR methods

Multidimensional QSAR methods are developed to address the disadvantages of 3D-QSAR methods by incorporating additional descriptors (dimensions). Hopfinger et al. "Developed a 4D-QSAR model using Boltzmann conformational sampling. Receiver geometry is essential for the development of a 4D-QSAR model." This model was developed using XMAP software.

I.3.4 Characteristics of receiver selection

A common problem in any QSAR study is to describe the molecular properties. The nature of the descriptors used and the degree to which they encrypt the structural features related to the dependent variable is a vital part of a QSAR study.

Chapter I.4

Protein docking and folding

"The completion of the human genome project has led to an increasing number of new therapeutic 'targets' for drug discovery. At the same time, high-performance protein purification techniques, crystallography, and nuclear magnetic resonance spectroscopy techniques have been developed which have contributed many structural details of the proteins "(Meng et al., 2011). These advances allow computational strategies to now penetrate all aspects of drug discovery, such as virtual screening (VS) techniques for identifying "hits" - and molecular optimization methods.

I.4.1. Rigid docking and receptor docking

When the ligand and the receptor are both treated as rigid bodies, the search space is minimal, since there are only three degrees of freedom of translation and three degrees of freedom of rotation. In this case, the flexibility of the ligand could be addressed by using a precalculated set of ligand configurations or by allowing a degree of atom-atom overlap between the protein and the ligand. DOCK was the first automatic docking procedure for a molecule at a receptor site and was developed continuously. It characterizes the ligand and the receptor assets of spheres that could be superimposed through a click detection procedure.

FLOG generates the conformations of the ligand-based on the distance geometry and uses a click determination algorithm to calculate the sets of distances. Up to 25 specific ligand conformations could be used to cover some flexibility. FLOG allows users to define critical points that must be associated with a ligand atom (Elkoley and Doersksen, 2013). This approach is useful if a significant interaction is already known before docking. The conformations are marked with a function taking into account van der Waals interactions, electrostatic, hydrogen bonds and hydrophobic interactions.

I.4.2. Flexible ligand docking and rigid receiver

For systems whose behavior follows the induced matching paradigm, it is of vital importance to consider the flexibility of the ligand and the receptor, because in this case both the ligand and the receptor change their conformations to form a perfect minimum energy complex. However, the cost is very high when the receiver is also flexible. Thus, the typical approach, a trade-off between accuracy and computation time, treats the ligand as flexible, while the receptor is held rigid during docking. Almost all docking programs have adopted this methodology, such as AutoDock, FlexX.

I.4.3 Flexible ligand docking and flexible receptor docking

Intrinsic protein mobility is closely related to ligand behavior and has been reviewed by Teague. Incorporating receiver flexibility is a significant challenge in the shower field. Ideally, the use of molecular dynamics (MD) simulations could model all degrees of freedom in the ligand-receptor complex. However, the MD has the problem of inadequate sampling as mentioned earlier. In addition to the induced true fit, several theoretical models have been proposed: conformer selection and conformational induction to illustrate the flexible ligand-protein binding process. According to Teague's definition, conformer selection refers to a process in which a ligand selectively binds to a favorable confirmation from many protein conformations.

I.4.4 Folding of proteins

It is the physical process by which the protein chain acquires its native three-dimensional structure, a conformation that is usually biologically functional in an expedient and reproducible manner. It is the physical process by which a polypeptide folds into the characteristic and functional three-dimensional structure of the random "coil." Each protein exists as a random polypeptide or "coil" when it is translated from an mRNA sequence into a linear amino acid chain.

I.4.5 Tertiary structure of proteins

Once the tertiary structure of the protein is formed and stabilized by hydrophobic interactions, there may also be a covalent bond in the form of disulfide bridges formed between two cysteine residues. The tertiary structure of a protein involves a unique polypeptide chain. However, the new interactions of the folded polypeptide chains give rise to the formation of the quaternary structure.

I.4.6 Quaternary structure of proteins

The tertiary structure may give rise to the formation of the quaternary structure in some proteins, which usually involves the "assembly" or "co-assembly" of the already folded subunits. In other words, multiple chains of polypeptides could interact to form a fully functional quaternary protein.

I.4.7 Modeling of protein folding

The "de novo" or "ab initio" techniques for predicting the structure of the computerized protein are related, but strictly distinct from experimental protein folding studies. Molecular dynamics (MD) is an essential tool for studying folding and protein dynamics in silico. Foldable equilibrium simulations were performed using the implicit solvent model and "umbrella" sampling. Due to the computational costs, the simulations of MD folding from the beginning, with clear water, are limited to minimal peptides and

proteins. The MD simulations of the more abundant proteins remain limited to the dynamics of the experimental structure or the unfolding at high temperatures. Long-lasting folding processes (over 1 millisecond), such as folding of small (approximately 50 residues) or more abundant proteins, can be accessed using coarse-grained models.

Chapter I.5

Molecular descriptors

Chemical graph theory is an interdisciplinary science, between chemistry and mathematics. We try to solve some difficult problems of chemistry, such as the enumeration of isomers, molecular symmetry and / or elucidation of complex chemical structures.

I.5.1. Basic definitions in graphs

A graph is a pair of two sets: the set of vertices/atoms V and the set of edges/links E . A margin is a pair of vertices (i, j) belonging to E (G) (Harary, 1969), G is a (non-directed) molecular graph. The number of points in G is given by the cardinality of the set of points. The term graph was introduced by Sylvester (1874).

I.5.2. Square topological matrices and indices

A molecular graph can be a sequence of numbers, a polynomial, a single number, or a matrix. These representations have the purpose of being unique, for a certain structure.

I.5.2.1. The adjacency matrix

The adjacency matrix $A(G)$ Sylvester (1874) is a symmetrical square table defined as:

If instead of 1, we placed topologically, the bond order ($> C-C <$, $bo = 1$; $> C = C <$, $bo = 2$; aromatic bond, $bo = 1.5$), we have the connectivity matrix $C(G)$. The adjacency matrix allows the recovery of the corresponding graph.

I.5.2.2. The distance matrix

The distance matrix $DI(G)$ is a square, symmetrical mass, with dimensions $n \times n$, defined as (Harary, 1969):

The non-diagonal entries of this matrix are only the topological distance between i and j . The sum of half of all inputs in $DI(G)$ gives the Wiener topological index - well-known "W".

I.5.2.3. The die matrix

In graphs containing cycles, if the shortest paths are replaced by the longest path between two nodes i and j , the Detour DE (G) matrix can be defined (Harary, 1969, Diudea et al., 2002).

I.5.2.4. Cluj matrix

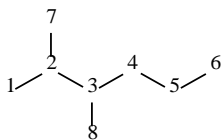
The Cluj CJ matrix (G), proposed by Diudea (Diudea, 1997, Diudea, 2010, Diudea et al. 1997, Janežič et al. 2007), is defined on Cluj fragments that collect the nodes v closer to i than to j , the ends of a path (i, j) , in other words, vertex neighborhoods (see Gutman, 1994) of v vs. any node j , joined by path p , with the distances measured in the subgraph $G \setminus p$:

The entries in the CJ matrices are calculated as the maximum cardinality of all the fragments that can appear in the graphs containing the rings.

Cluj matrices are defined both by distance (UCJDI) and bypass concepts (UCJDE); they are non-symmetric matrices, except for symmetric graphs. These matrices can be symmetrized by multiplying Hadamard with their transposition. The symmetric matrices based on the edges can be calculated from the one based on the path, by multiplying with the adjacent matrix of the graph.

In the trees, the entries of the Cluj matrix represent the number of paths passing through i to j . An example of this is given in the figure below. Notice that $RS(UCJp) = RS(We)$; $CS(UCJp) = CS(DI)$.

	1	2	3	4	5	6	7	8	S(W)
1	0	1	1	1	1	1	1	1	7
2	7	0	3	3	3	3	7	3	29
3	5	5	0	5	5	5	5	7	37
4	3	3	3	0	6	6	3	3	27
5	2	2	2	2	0	7	2	2	19
6	1	1	1	1	1	0	1	1	7
7	1	1	1	1	1	1	0	1	7
8	1	1	1	1	1	1	1	0	
CS(DI)	20	14	12	14	18	24	20	18	



Indexes from topological matrices (see above) are calculated as half of the sum of matrix entries.

$$I(G) = (1/2) \sum_i \sum_j [M]_{ij} = \sum_{i>j} [M]_{ij}$$

In the trees, the sum of half of the Cluj symmetric matrices gives Wiener W and hyper-Wiener WW indices, from SCJe and SCJp, respectively.

I.5.2. 5. Task matrices - "CH"

The CH (G) table of the weighted matrix is calculated based on the Sanderson group electronegativities group (Diudea and Ursu, 2003), as follows:

5.2.7 Layers and shell arrays

5.2.7.1 Layer arrays

A layer of peaks v located at distance k, in partition G (i) is defined as:

$$G(v)_k = \{v \mid d_{i,v} = k\} \quad G(i) = \{G(v)_k ; k \in [0, 1, \dots, ecc_i]\}$$

Where is the eccentricity of i. The entries in the matrix of the layer (of the property of the peaks) LM (Diudea, 1994, Diudea and Ursu, 2003, Diudea, 2010) are defined as

$$[LM]_{i,k} = \sum_{v \mid d_{i,v} = k} p_v$$

The layer matrix is a collection of entries below:

$$LM(G) = \{ [LM]_{i,k} ; i \in V(G); k \in [0, 1, \dots, d(G)] \}$$

D (G) being the diameter of the graph (i.e. the most significant distance in G). The column k = 0 is just the column of the peak properties. When pi = 1 (counting property), the LM matrix is called LC (counting layer matrix). Any atomic property/peak can be used as pi. Any square matrix M can be the matrix - info, that is, the matrix that provides local / peak properties, as implemented in the TopoCluj software (Ursu and Diudea, 2005).

i \ k	0	1	2	3	4	RS
1	1	1	2	2	1	7
2	1	3	2	1	0	7

3	1	3	3	0	0	7
4	1	2	2	2	0	7
5	1	1	1	2	2	7
6	1	1	2	2	1	7
7	1	1	2	3	0	7
CS	7	12	14	12		
kCS	12	28	36	16	92a	

LC layer matrices (G1); a = 2X the Wiener index

I.5.2.6. Layer arrays

The operator of the shell matrix, Sh, (Diudea and Ursu, 2003) transforms a square matrix into a layer/shell matrix:

$$[\text{ShM}]_{i,k} = \sum_{v|d_{i,v}=k} [M]_{i,v}$$

The peak centralization classes are identical to those found by permuting the adjuvant matrix entries, for example, by Mathematica and GAP (Groups, Algorithms, and Programming). The index is, in particular, a descriptor of the molecular form.

Personal contributions C hapter II

II.1Molecular adaptability of the ligand

PEI (polyethyleneimines) linear and branched and the space of properties

II.1.1Introduction

A space of chemical properties defines the adaptability of a molecule to changing conditions and its interaction with other molecular systems determining a pharmacological response. In a congeneric molecular series (compounds with the same derivatization algorithm and therefore the same crude formula), the chemical properties vary monotonically, that is, the congeneric compounds have the same range of chemical properties. The space of chemical properties is a crucial component in molecular design, where some building blocks are functional, that is, derivatized and eventually self-assembled into more complex systems, such as enzyme-ligand systems, of which (physical-chemical properties/bioactivity).) can be predicted by QSPR / QSAR (quantitative-property / activity relationship) studies.

The molecular adaptability of a particular compound represents the remodeling process whereby the compound adopts a new conformation appropriate to its interaction with another molecule, proteins, medium conditions. One way to quantitatively evaluate this intrinsic property of a compound is by calculating the molecular adaptability index. The index is the ratio between the variance of a topological descriptor sensitive to changing the topological coordinates and the attenuation of the energy of the compound.

II.1.2 Computational methods

IEPs are explored computationally through their space of chemical properties. For this class of compounds, two-dimensional graphs and radar graphs are suitable for designing the relevant subspace for the receiver and, respectively, for diversity analysis (Franco et al., 2008; Shelat et al., 2007, Perlman et al., 1999). Using statistical methods to explore the space of the chemical properties of the PEI, their Physico-chemical properties can be predicted even consecutively for derivatization and functionalization.

In other words, the physicochemical behavior of a class of compounds, when moving in its chemical space, can be predicted by computational methods (Mandal et al., 2009). To accurately predict the behavior of derivative compounds, the properties of the essential compounds must be calculated (Guru et al., 2015). In order to study PEI, their space of chemical properties, that is to say the variations of properties following the functionality and derivatization, were developed to simulate the interaction of PEI with molecular systems (i.e. nano-devices) for a set of PEI molecules.

II.1.3 Results

In order to give a quantitative measure of the similarity/dissimilarity presented a mathematical approach was used: a polynomial trend line of order four was calculated for the groups belonging to BPEI 04 (C14N8 and C18N10), thus obtaining fourth-degree polynomials (a total of 15 equations were calculated) as the equations below:

$$\text{C14N8 BPEI 04: } y = 0.0001x^4 - 2 * 10^{-5}x^3 - 30.06x^2 + 0.661x + 97.401 \quad (1)$$

$$\text{C18N8 BPEI 04: } y = 0.000x^4 - 1 * 10^{-5}x^3 - 14.96x^2 + 0.622x + 48.499 \quad (2)$$

Then the roots of the equations were calculated. They consist of real and imaginary numbers (see table 1, for the presented equation). Root calculation was performed using the Wolfram alpha online computational engine.

Table 1. Roots for degree 4 equations: (3), (3)

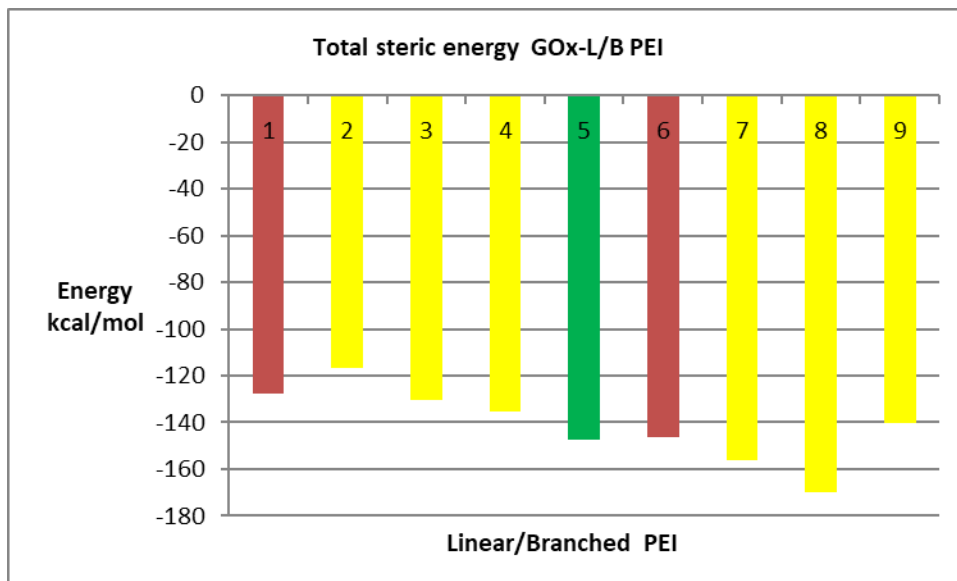
Equation	Root 1	Root 2	Root 3	Root 4
3. C14	-545.181	-57.254	57.275	545.361
3. C18	20.376	50.003	-11.684-18.421i	-11.684+18.421i

To represent in detail the space of chemical properties, ten molecular descriptors were calculated for PEI and the results were combined in a two-dimensional form, and a surface similar to the radial graph was plotted.

Each property was represented by the radial graph. The surfaces LPEI01 and BPEI 02 are not visible being covered by the surfaces BPEI 03 and BPEI 04 of the free PEI series C14N8.

There are apparent differences in the chemical properties of PEI before and after docking. Comparison of two molecules is equivalent to forming a complex cluster (corresponding to their virtual interactions). According to the network theory, once the complexity of a group between two PEI structures is formed, it is directly proportional to their dissimilarity (that is, the more "different" the molecules, the more complex the cluster results).

The comparison between the docking energy (total steric energy) and the after-docking energy of LPEI and BPEI with an approximate number of carbon atoms is shown in the figure below



The steric energy of the GOx-L / B PEI complex: (from left to right) 1 and 6 L PEI C14 and C18 (in red); 2 to 4 and 7 to 9 represent the corresponding B PEI branch isomers (in yellow). The LPEI C16 isomer is represented in green.

I.1.4 Conclusions

LPEIs change their geometry more quickly compared to BPEIs, which means that LPEIs are more adaptable to a particular binding site. LPEIs are adaptable to the location and space of the chemical properties - stable. From the perspective of the variation range, the PEI set C18N10 is more favorable compared to the C14N8 PEI. If you want to build a nanodevice using PEI and to have GOx as a component, then LPEI is preferred, at least in first-generation devices (in which LPEI is docked to GOx). If further functionalization is required, then C18N10 PEI should be chosen to the detriment of C14N8. When talking about the thermoresponsive properties, which are correlated with the central moment of inertia, the LPEI should be chosen preferentially at the BPEI. The central moment is related to the temperature and thus to the thermal stability in the GOx-PEI complex [31]. The optimal size of PEI

should be around C18N10, which is midway along the PEI chain, as suggested by the QSAR model developed in this study, to obtain evidence of an absolute "saturation" of GOx through PEI and vice versa, when the size of the PEI molecule increases. The branched PEIs have a relatively small energetic influence of GOx binding compared to linear PEIs (Lungu et al., 2016).

In conclusion, a non-branched linear compound has better molecular adaptability than a branched compound of the same molecular formula. This is mainly due to steric factors that appear as a result of branching (eg derivatization). For molecular recognition, a linear compound more easily folds its structure, for example, in a "pocket" or a functional group. In molecular design, for a compound to have endless degrees of freedom for a given chemical space, it means a loss in its specificity. The bioactivity space has a rigid domain.

A linear compound can move through the entire bioactivity space, with the consequence of losing the specificity for a specific biological target. In order to limit this unbounded molecular adaptability, a compound (ligand) must be branched so that only a specific region of the chemical space, corresponding to the bioactivity space, can be explored by the compound. Thus, the compound can have only distinct conformations that reach distinct areas of the bioactivity space. These "limitations" increase the specificity of the compound for a particular target (receptor). Finally, from the ligand's point of view (guests), each compound needs a specific adaptability to the receptor, so that molecular recognition takes place.

Chapter 2

Molecular adaptability of the receptor

Molecular adaptability of flavin adenine dinucleotide (FAD) among adjacent amino acids and its catalytic role in GOx (gluco oxidase)

II.2.1 Introduction

The formation of a complex between a receptor and a ligand is the critical event in any biological pathway. In such an interaction, the ligand, and the receptor (for example, the enzyme, the protein) mutually model each other to obtain an optimal conformation, capable of performing biological functions. Except inert functional proteins, such as those found in connective tissue, functionally active proteins, in order to fulfill their function, must exist in a complex with a small molecule called a ligand, whether the ligand is exogenous, endogenous or a cofactor. Cofactors are a different category of endogenous ligands, due to their permanent non-covalent bond with the enzyme (receptor). Typically, a ligand forms non-

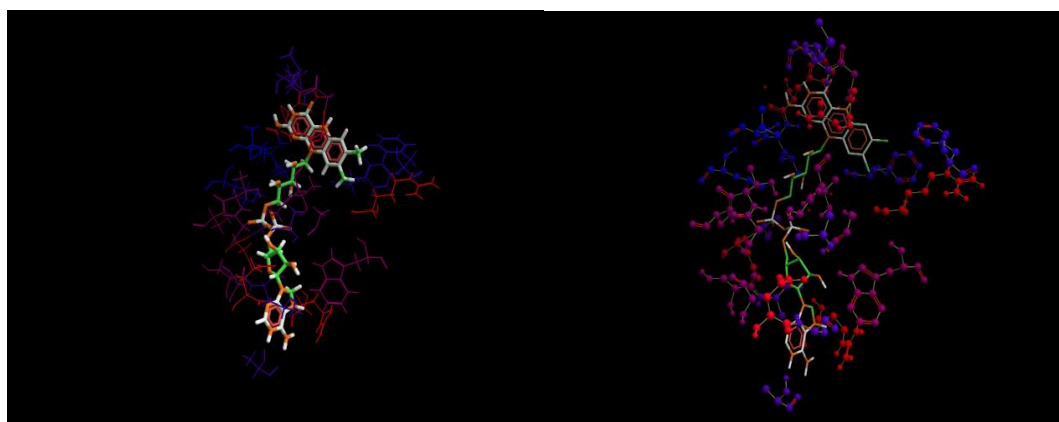
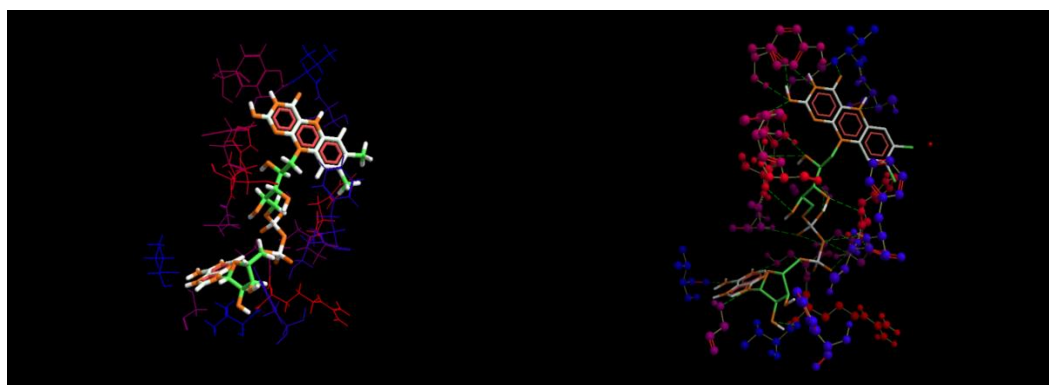
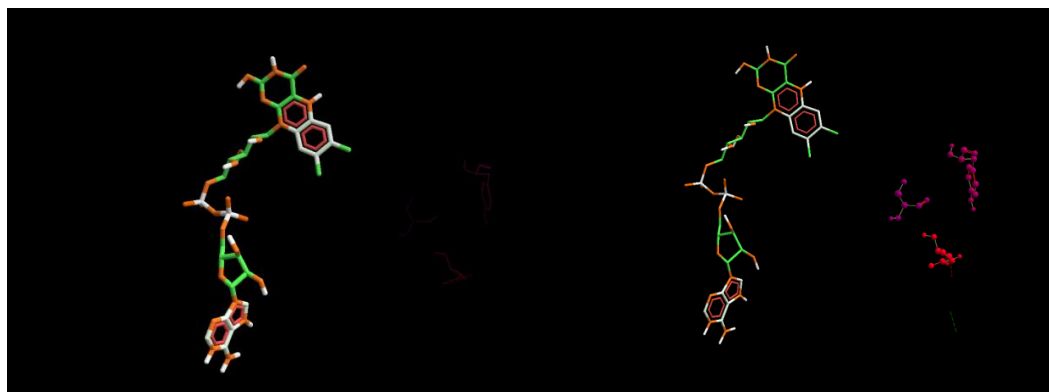
covalent bonds with the receptor for a relatively short period, after which the complex is either metabolized or the ligand is decoupled from the receptor, making the receptor ready for another binding event. Enzymes, which also play the role of a receptor for their substrate (ligands), make permanent structural changes on the substrate (ligand). $K_{ik_{off}}$, the nature, and dynamics of these processes are modulated by cofactors.

II.2.3 Computational methods

In the study of enzymatic systems, computational methods were used - homology, molecular dynamics. All molecular systems designed "in silico" have been minimized, protonated at physiological pH, corrected interatomic loads and distances.

II.2.3 Results

In the homology model, the ADF takes the lowest energy, the most favorable, due to the flexibility of the model (in the model, disulfide bonds between the chains were avoided). The GOx energy potential, as represented by pH and temperature ranges, suggests a temperature variation, independent of temperature. This means that a potential energy decrease of about 0.08305 kcal/mol (see potential = -16691,6689 kcal/mol, the average potential energy to be changed -16691,7519kcal / mol) will keep GOx activity at pH 7 and 50 ° C. The total energy of FAD, mainly steric energy, increases with temperature. The two formation heat variants may, by chance, represent transition states (Solomons and Fryhle, Wiley, 2004) of the FAD. However, this computational result may be a malfunction of the force field used. Factor B (or temperature factor) suggests that GOx has the most excellent flexibility in the euthermic state. In the hyperthermia state, strong atomic mobility is expected, depending on the decrease of the B values, probably producing a decrease of the enzyme activity. In other words, due to the "immobility" of the proteins, the contacts between the FAD and the amino acid network tend to increase, increasing the total and steric energy of the FAD-Aa system. The energy of the amino acids does not correspond to their optimal energy so that the ADF will have an unfavorable confirmation for the catalysis of beta-D-glucose oxidation. Mutation in pocket-bound amino acids can alter FAD behavior to external variations. ADF, after internal amino acid mutation, reacts more to internal changes, without "seeing" the entire network. The figure below illustrates how FAD is in contact with amino acids. The diagram shows FAD surrounded by amino acids, in the presence of water molecules.



FAD - "meshing" amino acid: 1GPE, 2WDW, 4YNT, by hydrophobicity (left column) and by electrostatic interactions (right column).

II.2.4 Conclusions

FAD plays a significant role in modulating flavoproteins (ie, enzymes having FAD as a cofactor), determining the type of substrate and the type of reaction. The ADF can take different conformations; in other words, it has a structure "adaptable" to the amino acid "meshing", thus offering its own "response", through different conformational and energetic geometries. As expected, the external conditions can

produce changes in the FAD geometry by interacting with the inside (at the binding site) and outside (all other amino acids) amino acid networks. Optimization of the amino acid network can mask or make "visible" (for FAD) a part of the internal or external network. The increase of these two amino acid networks allows changes in the adaptability of FAD and consequently in the behavior of enzymes, in terms of stability, substrate specificity and performance.

Chapter 3

Predicting power

The potency and prediction of the binding site of teixobactin and other Lipid II ligands through the statistical baseline scores of the conformational spatial maps.

II.3.1 Introduction

As discussed previously, a receptor binding pocket can be occupied by a ligand in various plausible conformations, of which only a few are energetically linked to biological activity in the field of physiological efficiency. The location of the receiver and the efficiency of this interaction can be predicted. A statistical prediction methodology is proposed here. To validate this methodology, a number of Lipid II inhibitors have been studied. This class of compounds was chosen because of recent interest in Lipid II inhibitors. Of these compounds, teixobactin molecules, which inhibit bacterial wall synthesis by interacting with Lipid II, are the most studied. A new methodology, based on the statistical evaluation of the conformational maps, on the molecular dynamics grouping and the principal component analysis (PCA), is used to achieve this objective. The methodology used is exemplified in one case.

Lipid II, peptidoglycan, is a precursor of bacterial cell synthesis and has both hydrophilic and lipophilic properties. The molecule translocates a bacterial membrane to deliver and incorporate "building blocks" of disaccharide pentapeptide into the peptidoglycan wall. Lipid II is a valid antibiotic target (Wright, 2015), although it has a relatively modest expression in the bacterial cell, less than 2000 molecules per bacterium (Heijenoort et al., 1992). The molecule must be inclined outside the cytoplasmic membrane to incorporate a disaccharide peptide unit into the peptidoglycan, thus being relatively accessible to an outer molecule (i.e., antibiotic). Lipid II is synthesized as a functional and essential molecule even in cell-free organisms (Henrichfreise et al., 2009) (such as Chlamydia, Mycoplasma, Wolbachia, etc.). Antimicrobial

compounds, which act with Lipid II and are selected in this study, include Dalbavacin, Oritravacin, Telvanicin, Teicoplanin, Teixobactin (Wright, 2015), Telvanicin and Vancomycin. Other compounds are known in this regard: Ramoplanin (Heijenoort, 2007), some peptides such as Nisin (De Gruijff et al., 2008), Copsin or alpha defensin human HNP-3 (Ganz et al., 1985).

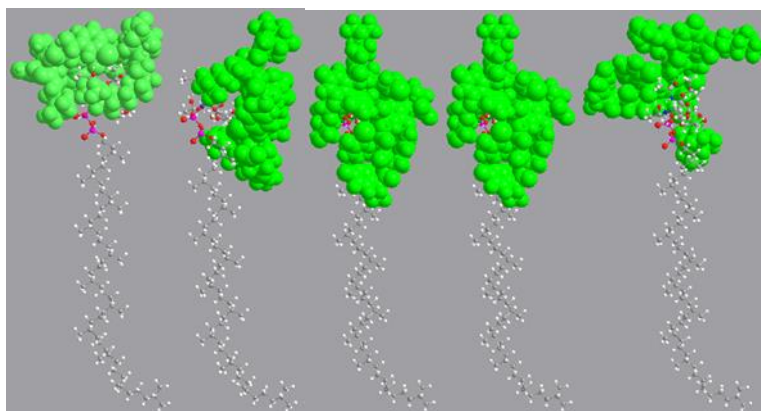
II.3.2 Computational methods

In order to study computationally the complex between Lipid II and ligands, a docking study was first performed. The docking port has been taken from the literature (Hart et al., 2016). The C6 atom with Cartesian coordinates $x = -13.353$, $y = -13.886$ and $z = 1.712$ (D-ala-D-ala motif) was chosen as the center of the docking cavity. The radius was set to 15 Å. After docking, 5 conformations with ligand and 5 consecutive complexes for each molecule, noted from 00 to 04, were taken into account.

As was done in the case of Lipid II, a statistical analysis of the pixel population was performed. A scoring method has been introduced to evaluate the potency of the ligand with respect to the antibacterial effect. The method is composed of 3 rules/evaluation criteria: (i) Evaluation of the receptor-ligand complex according to the "chances" of the receptor to bind the ligand. This criterion refers to the lipid II conformation study. (ii) Evaluation of the ligand-receptor complex in terms of ligand binding efficiency. This rule explores cluster density based on molecular dynamics study. (iii) Evaluation of the complex that characterizes the ligand by PCA. The molecules are arranged using these criteria. The resulting data were compared with those of the specialized literature.

II.3.3 Results

The conformations obtained for Dalbavacin are shown in the figure below. In each image, the ligand is represented as a space-filling colored in green. Lipid II is shown in gray. Five energy feasible conformations were selected for the binding site. For each of these conformations, Lipid II "rotates" inside the ligand in each case.



Dalbavacin complex (conformations, from left to right) 00,01,02,03,04.

The low standard deviation regarding the conformational analysis of Lipid II is correlated with an energetically favorable "population" of the conformations of Lipid II. Our data show that Lipid II has a population of more favorable energy conformations in docking with Teixoplamine (in conformation (structure) 01, with the highest efficiency score (15 points)) compared to Vancomycin (of which the most favorable structure 02 has the lowest score in the series of ligands (3 points).

conclusions

The results obtained by this computational method are in agreement with the experimental data published in the literature, for example, teixobactin is reported to be active in vitro against all gram-positive bacteria tested and efficiently in vivo in mice infected with *S. aureus* with methicillin resistance. Needed to achieve a 50% survival rate against MRSA was only 10% of the dose of Vancomycin (Ling et al., 2015), PD50, a typical antibiotic used for MRSA. Thus, the ratio of Teixobactin: Vancomycin, 1: 5 observed experimentally is found in this study (Teixobactin is five times stronger than Vancomycin 15: 3).

Chapter 4

Prediction and representation of the biological surface

Modeling of ligands in rigid docking of *MraY* inhibitors. The polynomial discriminant and the Laplacian operator as predictors of bioactivity

II.4.1 Introduction

The interaction of a small molecule (ligand) with a protein structure, ie receptor, is the basis of the mechanism of action of the drug for the vast majority of bioactive substances. The ligand and the receptor adapt their geometry and energy, within this interaction, for the benefit of the receptor-ligand complex. In an induction docking mode, the structure of the ligand is most susceptible to changes in topology and energy compared to the receptor. These changes can be described by multiple hypersurfaces, in terms of discriminating polyfunctional operator and Laplacian operator. Such topological surfaces were represented for each *MraY* inhibitor (phospho-MurNAc-pentapeptide translocase), studied before and after docking with *MraY*. The binding energies of all ligands were calculated by this procedure. For each ligand, the Laplacian and polynomial discriminants were correlated with the minimum inhibitory concentrations (MICs) extracted from the literature. There was a strong correlation between all these.

II.4.2 Computational methods

The structures of 5 members of the class of inhibitors representative of *MraY* were studied: Caprazamicin A, Liposidomicin B, Muraimicin Cl, Mureidomicin A, Tunicamicin I.

The topological coordinates for all ligands and their pharmacological targets *MraY* were plotted using scatter plots. The tendency for each coordinate was expressed as a logarithmic equation of trend. The logarithmic equation gathering was represented as a unified log (s) function. The procedure was applied to all the studied molecules. For example, for Caprazamicin A, the logarithmic surface generation equation has the form:

$$Y = \int (0.9741 \ln(x) - 4.0486 \ln(x) - 0.1524 \ln(x) - 0.6333 \ln(x) + 14.913) dx.$$

Using the above equation, 3D graphics were generated for all five structures.

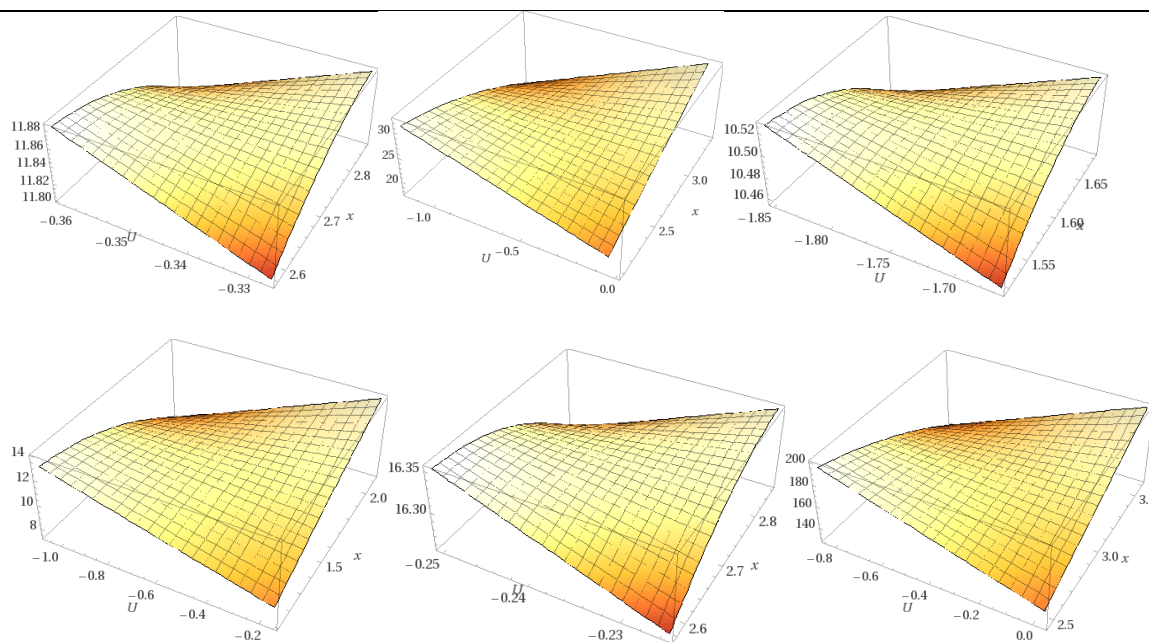
II.4.3 Results

The docking procedure was successfully completed. The receptor (*MraY*) was considered fixed, and the ligands were considered mobile. For each ligand, the best conformation positions were retained. All ligands are attached to the designated binding site. In Figure 2c, the best positions for the studied compounds are colored green, in the *MraY* binding pocket. It is noted that no compound is linked to another. In order to test the docking ability to predict the correct (i.e. bioactive) conformation, re-docking of Muraimicin D on *MraY* proved to be successful.

Logarithmic equations for free and bound ligands.

Molecule	X	Z	Y
Caprazamicin A	$Y = 0.9741 \ln(x) - 4.0486$	$Y = 0.1524 \ln(x) - 0.6333$	$Y = -3.588 \ln(x) + 14.913$
Liposidomicin B	$Y = 0.2486 \ln(x) - 4.9995$	$Y = -2.49 \ln(x) + 8.0632$	$y = -2.371 \ln(x) + 6.8077$
Muraimicin Cl	$Y = 0.3637 \ln(x) + 8.2632$	$Y = -2.35 \ln(x) + 9.3795$	$Y = -0.105 \ln(x) + 0.4055$
Mureidomicin A	$Y = 2.3747 \ln(x) - 7.7925$	$Y = 1.7282 \ln(x) - 5.4429$	$Y = -0.13 \ln(x) + 0.4852$
Tunicamicin I	$Y = -1.382 \ln(x) - 1.382$	$Y = 1.219 \ln(x) + 3.022$	$Y = -1.325 \ln(x) + 1.95$
<i>MraY</i>	$Y = -0.991 \ln(x) + 34.353$	$Y = 3.639 \ln(x) - 12.969$	$Y = -6.486 \ln(x) + 56.661$
Molecules after docking	X	Z	Y

Caprazamicin A	$Y=1.1474\ln(x)+2.1222$	$Y=2.2825\ln(x)+7.4555$	$y=1.4935\ln(x)+4.3625$
Liposidomicin B	$Y=0.6125\ln(x)+8.986$	$Y=1.8278\ln(x)+9.8712$	$Y=0.1306\ln(x)+9.6344$
Muraimicin CI	$Y=1.9153\ln(x)+2.4093$	$Y=-0.492\ln(x)+22.764$	$Y=0.1996\ln(x)+7.3225$
Mureidomicin A	$Y=0.9779\ln(x)+6.4303$	$Y=-0.93\ln(x)+19.552$	$Y=0.6179\ln(x)+7.1176$
Tunicamicin I	$Y=0.6101\ln(x)+10.26$	$Y=1.5282\ln(x)+14.268$	$Y=0.851\ln(x)+6.0062$
MraY	$Y=-0.991\ln(x)+34.353$	$Y=3.639\ln(x)-12.969$	$Y=-6.486\ln(x)+56.661$



Topological collections, after docking, represented in order, from left to right: Caprazamicin A, Liposidomycin B, Muraimicin CI, Mureidomycin A, Tunicamicin I and MraY.

The topological spaces of the ligands, in the free ligand (eg, energy minimized) and stationary (in a complex with MraY), are thus Hausdorff Euclidean spaces, described as topological varieties. The Cartesian coordinates and subsequent topological spaces are modeled by the intramolecular and intermolecular forces. These forces, depending on the type of interaction, are: (i) electrostatic forces, due to the resident charges in the matter; (ii) electrodynamic forces; (iii) the most widespread van der Waals interactions; (iv) the steric forces, caused by entropy, manifested, for example, in the solvation processes (especially in the movement of water molecules from the binding site); (v) hydrogen bonds; (vi) hydrophobic interactions, etc.

II.4.4. Conclusions

The topological surface of the ligand, after docking, is modeled by the receptor. The topological surface of the ligand is a two-dimensional manifestation, modeled by the intra- and intermolecular forces involved in docking. The Laplacian operator and the polynomial discriminant of the second degree Cartesian equation (x, y, z -trends) can be used to distinctly characterize the topological surfaces of each ligand. The Cartesian coordinate equation is strongly correlated with binding affinity (calculated in silico) and MIC. These operators, calculated for a two-degree topological distributor of the docked ligand, can be used to predict bioactivity in a pathway involving the ligand-receptor complex (Lungu et al., 2017).

Chapter 5

Protein folding and interactions

Alpha 1 folding of antitrypsin and interaction with nanomaterials

II.5.1 Introduction

Proteins can only perform their action when their tertiary and quaternary structure is intact. In this state, certain regions of the protein located at the binding site may orient and allow the process of molecular recognition and complete formation of the ligand-receptor complex. In non-physiological conditions, such as high or low body temperature, increased or low osmotic pressure, radiation, chemical reagents, proteins are denatured (Kang et al., 2000). The unfolding is a reversible process when the active regions of the protein (i.e. the receptor) unfold and allow the ligand to detach. In the "revealed" state, protein energy and entropy are high. In the folded native state, energy and entropy are low.

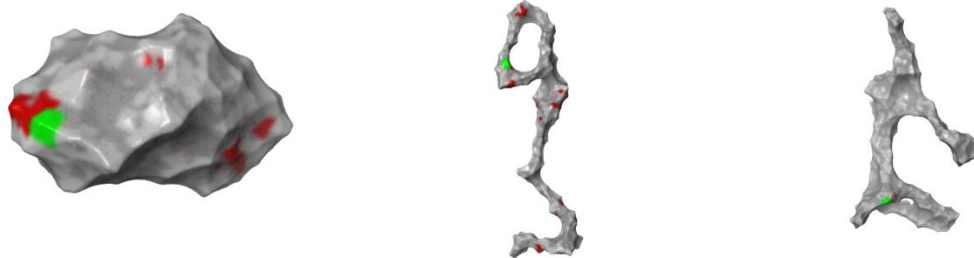
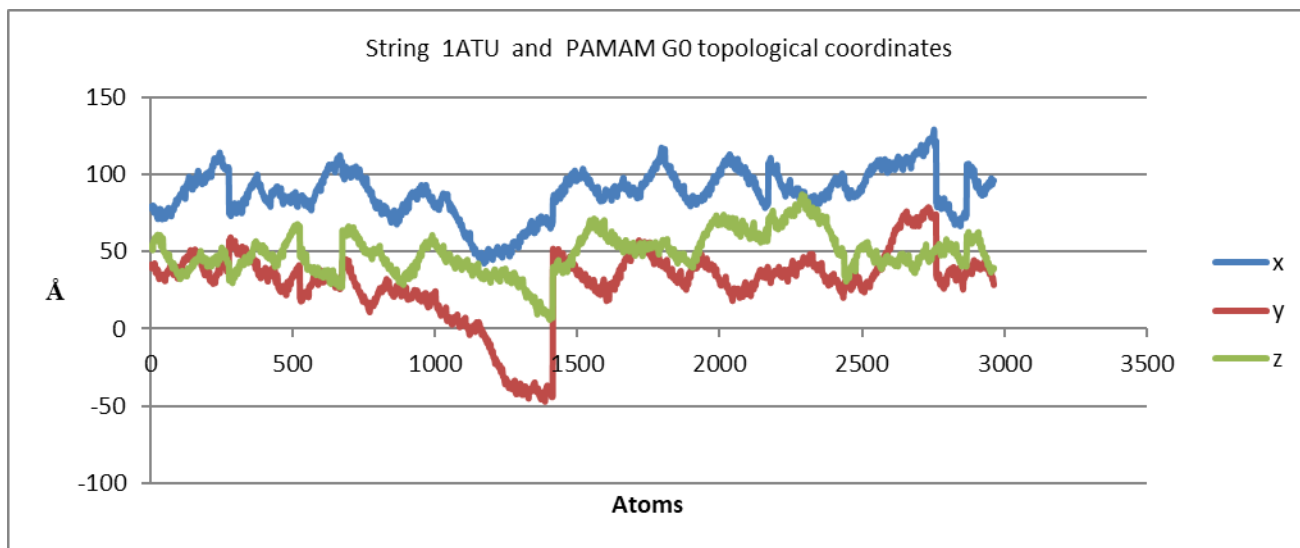
II.5.2 Computational methods

The 1ATU protein with a molecular weight of 41.97 kDa was chosen to be studied. The interaction with proteins was studied compared to a PAMAM polymer monolayer. The PAMAMs studied were C22H48N10O4 (G0), C62H128N26O12 (G1), C110H192N26O44 (G1.5), C142H288N58O28 (G2), C302H512N122O60 (G3), and a series of multiple C4H4 nucleotides: C20H36N2O4, C22H40N2O4, C24H45N3O6, C28H52N2O4. A total of 10 PAMAM protein systems were constructed. The structures were optimized for the reasonable intravascular condition: 0.15M NaCl, 310.15K temperature, a dielectric constant of 78 was chosen in terms of published data.

The aggregation surface of the protein was studied. Surface aggregation was studied on all the proteins, the resulting strings plus the "native" ones, with the 1ATU structure and on the "free" "PAMAM result," the "protein" string. Thus the properties of biological interaction and their topology (in terms of biochemical reactions, aggregation) can be described and compared. The overall aggregation process was also characterized by calculating the Zeta potential for a temperature range from 273K to 320K and a pH range from 7.4 to 7.8. The force field used to calculate the aggregation surfaces was OPLS3.

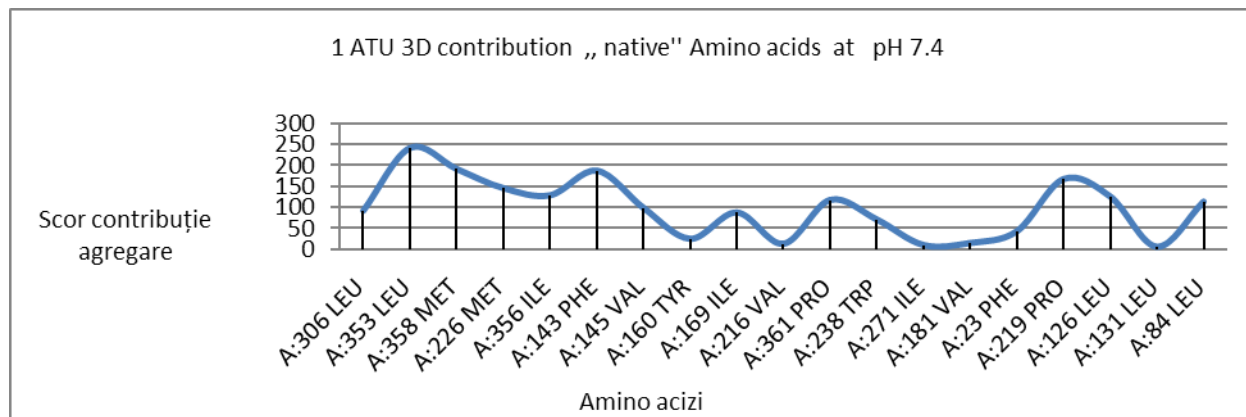
II.5.3 Results

The topological Cartesian coordinates of the protein chain presented distinct patterns for each corresponding protein-polymer system chain. Two major types of models have been distinguished: a model corresponding to the free protein with intact tertiary structure and a second model closed at the dendrimer of generation 1. Without the intersection of the Cartesian coordinates at the end of the terminal, the last 500 topological points correlate with a lower rate. Of interaction with the polymer layer (they have the same model as the free protein chain and the free protein with tertiary structure).



1ATU 3D structure "native," "free," and G1.5 respectively. The aggregation surface was calculated in red; all Aa are represented which can potentially contribute to an aggregation process. In green, Aa is represented with the highest chance of producing an aggregation.

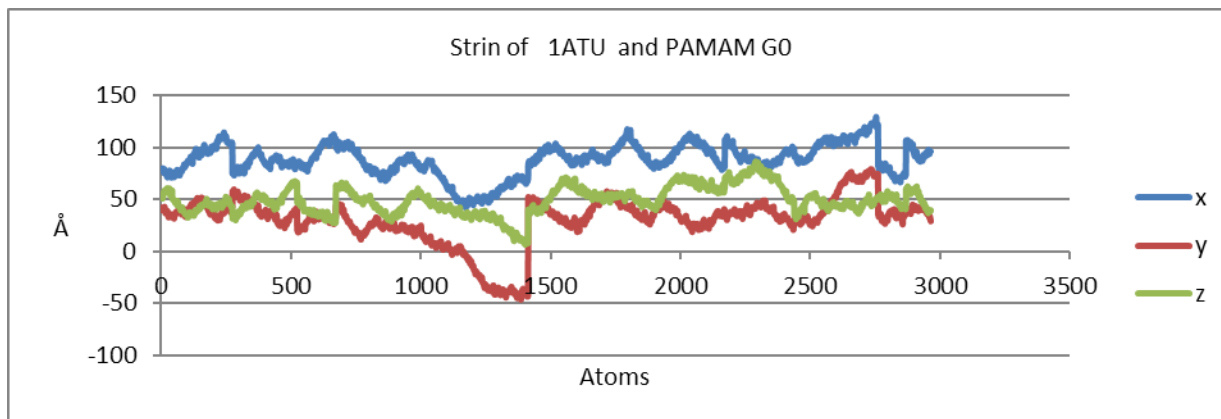
Protein aggregation score was calculated (Saunders et al., 2016). The lowest score has the string generated by the G1.5 layer. The figure shows the native protein score, the free string score, and the G1.5 score comparatively.



The protein or nucleic acid that folds or unfolds is a critical process in biology. Denaturation, the process in which the quaternary structure is lost, has relevance in disclosing the secondary and primary protein structure, present in the native state of the protein. Denatured proteins illustrate a wide range of representative properties from loss of solubility to aggregation. When proteins are "fixed" on covalent or noncovalent surfaces, a phenomenon similar to denaturation occurs. There is no doubt that one of the most frequently observed consequences of the attachment of proteins to the surface of carbon nanomaterials and nanopolymers is the appearance of conformational changes in their structures, which can lead to exposure of new epitopes, changes in protein functionality and avidity, therefore, to induce a strong immune response and / or unintended toxicity. The phenomenon of protein interaction with the surface of nanomaterials is a very complicated process, not only because of the dynamism of the observed changes, an effect similar to the Vroman effect, which can never reach a steady-state in vivo, but especially because of the complexity of the parameters of the tested system, which determines adsorption processes and thus the layer of biomolecular composition is formed.

In short, the protein "corona" is presented as a differentiated multi-layer structure in the primary, small part of the protein-nanomaterial interface. The interaction is represented by a "hard" part of this structure, formed from the surface of the proteins of recognition of the nanomaterials with high affinity to their adsorption and the slow exchange; and a secondary dynamic part of the "soft" type of change, short in

protein form and reversibly linked to the "hard" part. A significant impact on the process of creating the protein crown will have the unique surface characteristics of the nanosystem and proteins that determine the affinity of binding to the plane of adsorption and protein-protein interactions, the protein surface of nanomaterials, protein-water.



For each protein chain, the three logarithmic functions are unified, integrated, and represented on a 3D graph. For example, for the PAMAM G0 raw protein string, the following equation was used in graph construction:

$$\int [(2.372 \log(x) + 71.411) U(0.01076 \log(x) + 28.956) U(2.8141 \log(x) + 29.356)] dx.$$

The protein strands resulting from the interaction with dendrimers show a distinct topology. The representation obtained shows approximately the same morphology (same curvature, angle, orientation). Differences are depending on the type of dendrimer used. When the protein is computationally distorted without the dendrimer layer graph, there is an opposite "bending" compared to the dendrimer protein system. The graphical result by placing the coordinates of the integrated logarithmic function for 1ATU 3D, without being distorted, shows a relatively flat shape with some small folded regions, which correspond to the protein functional regions.

II.5.4 Conclusions

The polymer layers have a quantifiable effect on the proteins and favor different reactions. This effect is demonstrated by the variation of the topological coordinates. The variation of the topological properties determines variations of the biochemical properties that favor different types of reactions. From 1ATU PAMAM, G1.5 has probably the lowest interaction with proteins and toxic effects. This property is reflected by the analysis of the protein chain - the cluster count and the biochemical analysis of G1.5 which showed that there are no potentially binding biochemical reactions with proteins. The Zeta

potential suggests a less stable colloidal solution of 1ATU protein. Protein chain analysis demonstrates here a feasible method of exploring the interaction of polymer proteins. The protein structure is influenced by the immobilization environment. The technique used is suitable for calculating protein interactions with different media (carbon nanostructures, dendrimers).

By operating small changes in the structure of the protein (receptor), the bioactivity space of the ligand can be accessed by the ligands. When the whole protein is denatured, that is, reduced to its primary and secondary structure, the bioactivity space cannot be approached. Thus, the bioactivity space can only be accessed by a structure with tertiary and quaternary architecture.

Chapter 6

Exploring the area of bioactivity

The sample of the bioactivity space can be done computationally, using statistical methods mainly QSAR (multiple linear regression, partial least squares regression, vector regression, neural network regression) or by conventional methods - experiments: calorimetric titration (ITC), antibiograms for toxic compounds, etc.

II.6.1 Pyridine imidazole derivatives: QSAR models

II.6.1.1 Introduction

Imidazole, pyridine and dihydroxyacetone are valuable essences in drug chemistry, having a wide variety of biological activities, such as antitumor, antimicrobial, anti-inflammatory, antihypertensive, anti-neuropathic and antihistamine activities (Richaud et al., 2011; Zhang et al., 2011; Zhang et al., 2011; Zhang et al., 2011; Zhang et al. 2014). Although each scheme discussed exhibits promising drug-like properties, in terms of pharmacokinetics and pharmacodynamics, new compounds developed using these methods fail to turn into "hits." Most do not pass preclinical studies, and the others fail in Phase I or II clinical trials. The reasons are different: from the lack of in vivo bioactivity to the pharmacological interactions with other drugs.

II.6.1.2 Experimental methods

The imidazole-pyridine hybrid substances, without the dihydroxyacetophenone framework, were obtained using a three-step adjustment procedure described elsewhere (Mantu et al., 2016): N-acylation, N-alkylation, and quaternization of nitrogen-containing heterocycles. Imidazole-pyridine derivatives hybrid

with dihydroxyacetophenone backbone were obtained using an analogous procedure. The compounds were tested for antitumor and antimicrobial activity (against *Mycobacterium tuberculosis*). The anti-tumor screen used the NCI60 cell line (Shoemaker, 2006). The screening results were evaluated using the COMPARE software (<http://dtp.nci.nih.gov>). Antimicrobial activity against *Mycobacterium tuberculosis* was tested as part of the TAACF TB screening. Antimicrobial activity against *Mycobacterium tuberculosis* H37Rv, increased under aerobic conditions, was evaluated by determining the minimum inhibitory concentration (MIC) of the compounds. MICs were determined by measuring the bacterial growth after five days, in the presence of rifampicin as a control. For the potent compounds, the tests were repeated at lower initial concentrations. The analysis is based on measuring the growth in liquid medium of an H37Rv fluorescent marker strain, where the reading is done either as optical density (OD) or as fluorescence. The use of two readings minimizes the problems caused by compound precipitation or autofluorescence. A linear relationship between OD and fluorescent reads was established to justify the use of fluorescence as a measure of bacterial growth. The pharmacokinetics of these compounds was evaluated experimentally as the solubility of the compounds in a microbiological environment. The compounds were prepared as serial dilutions 2-fold in 10-point DMSO and diluted on average 7H9-Tw-OADC in 96-well plates. The highest concentration of compounds was 200 pM. For compounds with limited solubility, the highest concentration was 50 times lower than the concentration of the stock (for example, 100 pM for 5 mM DMSO stock, 20 pM for 1 mM DMSO). Turbidity was measured using a microplate reader from Nefeloskan Ascent. A compound was considered insoluble if the turbidity was 300% higher than that of the negative control. The lowest insoluble concentration is recorded. Each plate of test compounds included a positive control (1 mM haloperidol) and a negative control DMSO (2% DMSO in 7H9-Tw-OADC). Each experiment included control compounds: metoprolol tartrate, rifampicin, phenytoin, haloperidol, simvastatin, diethylstilbestrol, and tamoxifen.

II.6.1.3 Computational methods

The interactions of ligands with three target types (receptors) were considered: receptor 1, Dickerson Drew dodecamer with its medium structure (PDB id 2DAU) and atomic resolution (PDB id 436D) (Tereshko et al., 1999); Receiver 2, G-quadruplex (PDB id 2F8U) (Chen et al., 2006) (of the 10 conformers provided by the crystallographic model, the one with the lowest energy was chosen); Receptor 3, which is a complex between dioxygenase and DNA strand (PDB id 3S5A) (Chengqi and Chuan, 2013).

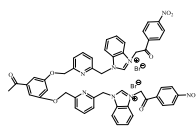
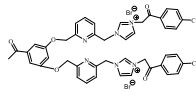
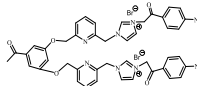
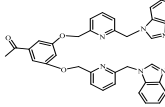
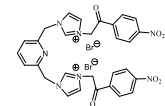
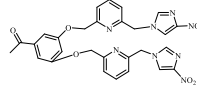
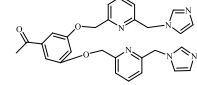
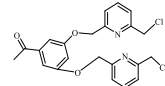
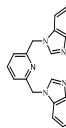


Descriptors evaluating the pharmacokinetics and drug-like properties were calculated using the QuickProp module included in the Schrodinger software package.

II.6.1.4 Results

II.6.1.4. Experimental results

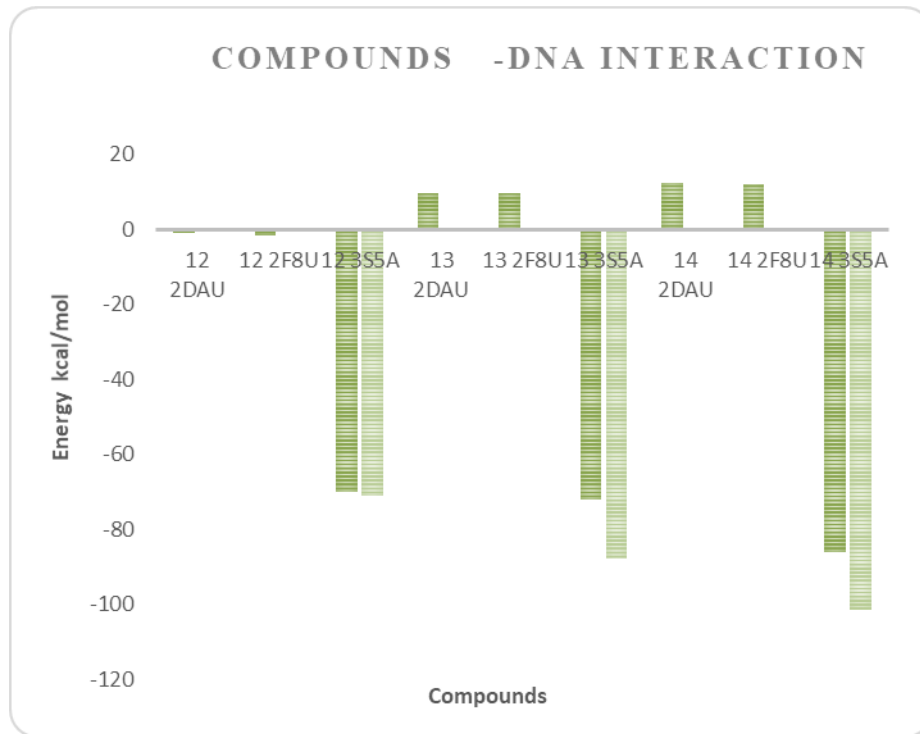
The synthesized compounds # 1 - # 11 were tested against *Mycobacterium tuberculosis*. The compounds (Table 1) showed significant anti-tuberculosis activity. The activities were expressed in IC₅₀ (minimum inhibitory concentration) and IC₉₀ (90% inhibitory concentration).

Componente testate pentru activitatea antituberculoasă

#	Formula	Formula brută	MIC	IC ₅₀	IC ₉₀
1		C₅₂H₄₂Br₂N₈O₉	9.6	3.7	10
2		C₄₄H₃₈Br₂Cl₂N₆O₅	8.4	3.4	8.6
3		C₄₄H₃₈Br₂N₈O₉	4.8	2.7	5.2
4		C₃₆H₃₀N₆O₃	39	20	42
5		C₂₉H₂₅Br₂N₇O₆	58	38	71
6		C₂₈H₂₄N₈O₇	12	9.2	11
7		C₂₈H₂₆N₆O₃	100	43	69
8		C₂₂H₂₀Cl₂N₂O₃	>200	>200	>200
9		C₂₁H₁₇N₅	>50	43	>50
10		C₁₃H₁₁N₇O₄	92	77	100
11		C₁₃H₁₃N₅	>200	>200	>200

Compound no. 3 showed the best activity against Mycobacterium tuberculosis

Docking on the R1 receptor (medium structure) for compounds # 1 - # 11 showed the following: (i) Residue A: 1 DC type, number of heavy atoms 16 (expected heavy atoms 19) (ii) Residue B: 13 DC type, docking coordinates x: -7.52, z: -10.00, y: 1.25. The binding position is shown in Figure 2.



Bar graph showing total sophisticated energy (dark green) and external interaction energy ligand (light green), in kcal/mol, of compounds # 12 - # 14 and receptors R1 to R3.

II.6.1.4. QSAR models

Using a component selection algorithm and taking into account the small size of the data set (11 molecules), three descriptors were selected from a data set of 177 descriptors: (1) Balaban distance connectivity index; (2) Balaban mass-weighted distance index and (3) shape index of the 2-Randić route.

The obtained model has $r^2 = 0.9656$, $p = 0.9068$, $q^2 = 0.9656$; model equation: $y = 0.9657\text{MIC observed} + 2.2917$ (figure 9). A list of the possible models considered is presented in Table 3.

II.6.1.5 Conclusions

The compounds discussed here showed promising pro-antitumor and anti-tuberculosis activity, as compared to standard approved drugs used as controls in our experimental tests. The divalent and larger molecules showed a higher affinity for DNA strains compared to the small, unbranched structures. Compounds with lower docking energies (negative energy) showed higher activity in terms of DNA. ADME and drug-like computational properties were shown in smaller, but lesser branched compounds.

Finally, the interplay between the DNA sequence and the peptide sequence can be explored. A peptide sequence can be constructed computationally from the selected DNA strands. Such a sequence is constructed in silico using DNA nucleotides. Starting from DNA nucleotides, the amino acid sequence can be identified computationally. Peptide inhibitors, encoded by DNA sequence, may be useful in experimental and computational methods, to be developed.

II.6.2. Oxabutan benzamides as inhibitors of UDP-3-O-acyl-N-acetylglucosamine deacetylase: a QSAR study

II.6.2.1 Introduction

UDP-3-O-acyl-N-acetylglucosamine (Jackman et al., 1999), UDP, (Uniprot ID O67648) deacetylase is an enzyme that has converted the conversion of UDP-3-O- [(3R) -3-hydroxymyristoyl] -N-acetylglucosamine in the presence of water in UDP-3-O- [-3-hydroxymistohistoyl] -D-glucosamine and acetate. UDP is a Zn protein that participates in lipid A biosynthesis (Li et al., 2008). Lipid A is a component of an endotoxin responsible for the toxicity of gram-negative bacteria. Lipid is a critical component involved in the onset of immune responses to gram-negative infection. It is a powerful stimulant of the immune system being able to activate monocytes or macrophages at pg/ml concentrations. Inhibition of lipid A synthesis leads to a decrease in the scale of the immune response, following infection with gram-negative bacteria. In this study, a QSAR model of UDP inhibitors is calculated on a series of 44 commercially available molecules, with the potential inhibitory effect on the experimentally determined UDP. The QSAR model was used in conjunction with chemical descriptors to classify compounds based on similarity scores

II.6.2.2 Computational methods

The QSAR model is based on clean calorimetric titration (ITC) data, which provides binding affinity expressed as equilibrium dissociation constants (K_d , nM). Data measured at 25°C and pH -7, were extracted from CSARDock, CD (Dunbar et al., 2011). In this respect, 54 compounds were initially taken (see Supplementary Materials), and the data were initially clean. In this process duplicates of molecules with the same structure and the same K_d (except those with different K_d for the same chemical formula) were removed. After this process, 45 compounds resulted (see Supplementary materials). On the available data K_d , the standard deviation and the standard error were calculated.

To construct the QSAR model, multiple linear regression was used. A number of chemical descriptors were used: indicators for H, C, N, O, P, S atoms, molecular mass (MW), total number of atoms, number of heavy atoms, number of rotation limits, number of donor groups of hydrogen (HD), the number of

hydrogen acceptor groups (HA), the number of rings, the minimum distance between two hydrogen donor groups, the maximum distance between two hydrogen donor groups, the minimum distance between two hydrogen acceptor groups, the maximum distance between two hydrogen acceptor groups and the Wiener index, compared to the computational and experimental ITC binding affinity, Kd obtained with UDP objective. Correspondingly, two models were built. The models were validated using the LOO "leave one out" technique. The QSAR model was used to characterize the interaction with UDP of the first three high-rank compounds, based on the binding affinity calculated after docking and on experimental Kd. For the three UDP ligand complexes, the energies were calculated and the most energetically favorable compound was chosen as the target for a similarity screening in a commercially available database. The QSAR model was used to predict Kd for compounds extracted from the commercial database.

II.6.2.3 Results

After optimizing all available molecules, a set of 22 structures was obtained, as shown in Table 1, together with Kd and the calculated BA binding affinity. The molecules are derived from benzamide and carboxamide.

The QSAR models obtained are the following:

(1) binding affinity

$$BA = -0.0788093 + 0.0986589 * X,$$

N = 22; Pearson square correlation (R2) = 0.9819; Standard error = 0.2417

(2) dissociation constant

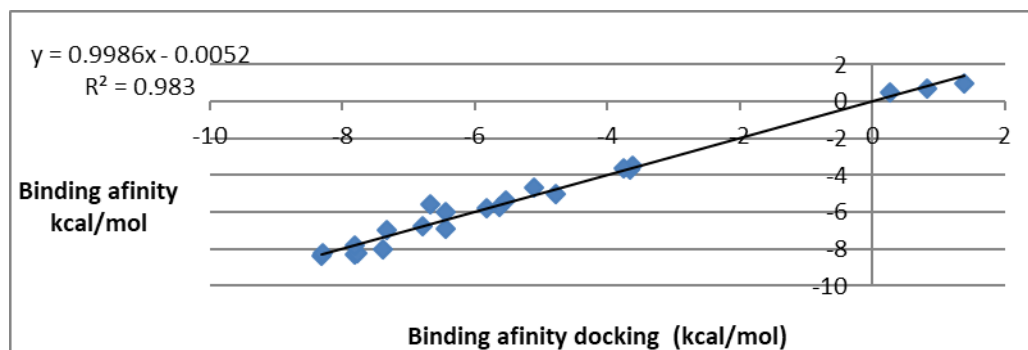
$$Kd = -6.13739 + 1.02618 * x,$$

N = 22; Pearson square correlation (R2) = 0.9855; Standard error = 0.2393

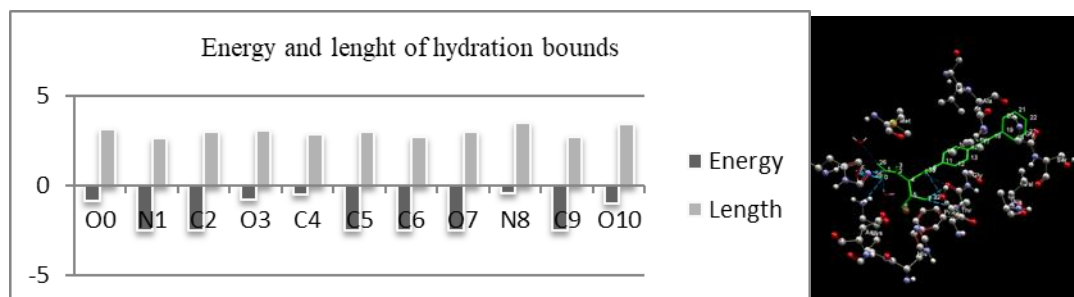
X is the observed variable. The values along the regression line are the anticipated post-test scores. The regression line falls in the center of the graph. It's the best fit line for the chart.

The data are represented graphically, as shown in fig.1-3. Details of the ligand interaction with the UDP binding site are shown in Figs. 4-9.

Considering the experimental affinity Kd and the binding affinity (obtained by docking), three molecules were chosen for a similarity study: # 6, 16, and 22.



Binding affinity, docking vs. predicted QSAR data.



Ligand # 22; The length of the hydrogen bond (Å) and the energy (kcal/mol) and the relationship of the Zn cofactor to the ligand, the black-energy, the long-gray.

II.6.2.4 Conclusions

The binding affinity calculated by docking and the K_d obtained experimentally are closely correlated. Multiple linear regression is a valuable tool in predicting binding affinity. Oxobutane benzamide in interaction with UDP gave a good predicted K_d value, suggesting a potential role as an inhibitor of UDP.

II.6.3 Famotidine as a urokinase inhibitor: a QSAR study

II.6.3.1 Introduction

A QSAR study was performed on a series of naphthalene cyclohexane amine derivatives with the experimental determination of the urokinase (K_i) inhibition constant to identify a potent human urokinase inhibitor. The QSAR model was constructed using artificial neural networks (ANN). The model was used to find compounds with strong urokinase inhibitory properties. For the compound resulting from the screening, the ANN model was used to calculate K_i . The best compound after K_i was chosen for discussion. An induced fixation docking (protein-ligand) and protein-protein docking were used to obtain an analysis of the results. A derivative of the isoquinoline-diol base resulted in inhibition of urokinase and drug-like properties.

II.6.3.2 Computational methods

To develop a QSAR model of urokinase inhibitors (Vincent, 1990), a series of compounds with the experimental determination of the inhibition constant (K_i) on urokinase was chosen. The data were processed to have a distinct K_i value for a single molecule. Also, the 3D structures were energy minimized, and the tasks were corrected. Furthermore, three Br-containing molecules were removed from the dataset due to failure in calculating partial charges.

Artificial neural networks (ANN) were used to build the model for multiple linear regression (MLR) method failures in predicting the target variable (K_i) with an $r^2 < 0.7$. The molecules in the dataset were classified based on a similarity score. The coating molecule was chosen based on the best K_i and using ANN correlation with the descriptors, and similarity screening was performed on a commercially available database. The QSAR model was applied to the set of molecules taken to calculate K_i . In order to interpret the results, a comparative docking study was performed on the ligand data established on the endogenous urokinase inhibitors API-1 and API-2.

In the first docking study, ligands were docked on the crystallographic structure of urokinase with PDB id 1C5Y. The docking port was taken from the literature and its precise coordinates were detected computationally: x 8.84Å, y 3.18Å, z 29.94Å with a radius of 15Å. Ligands 9 and 24 were docked to the urokinase using the maximum number of stable docking conformations for each ligand. A second protein docking study was performed with urokinase as the target and API-1 and API-2 as ligands. The structures that make up the ANN model were arranged based on the similarity score on the best K_i and the best correlation with the ANN model.

II.6.3.3 Results

A data set of 42 distinct molecular compounds resulted. Data were divided using a randomized algorithm from a training set (21 compounds) used to build the model, and a test set (21 compounds) used to test the model. The target variable was established as a constant inhibitor (K_i). The descriptors used to construct the model were: minimum aromatic distance between hydrogen donor groups, number of rings, number of hydrogen atoms, average distance between two hydrogen acceptor groups, Wiener index, number of C atoms that are hybridized sp^3 , tasks Andrews, the number of operating system groups, the total number of atoms. The neurons, including the learning range of the output layer used, was 0.3, the initial weight range was 0.5. The equation of the model is $y = 53.4084 + 0.956756 * x$. The Pearson correlation squared (r^2) is 0.9614, and the validated square CC (q^2) is 0.960921. The correlation between K_i used in model formation and the K_i prediction of the training set is shown in Figure 1.

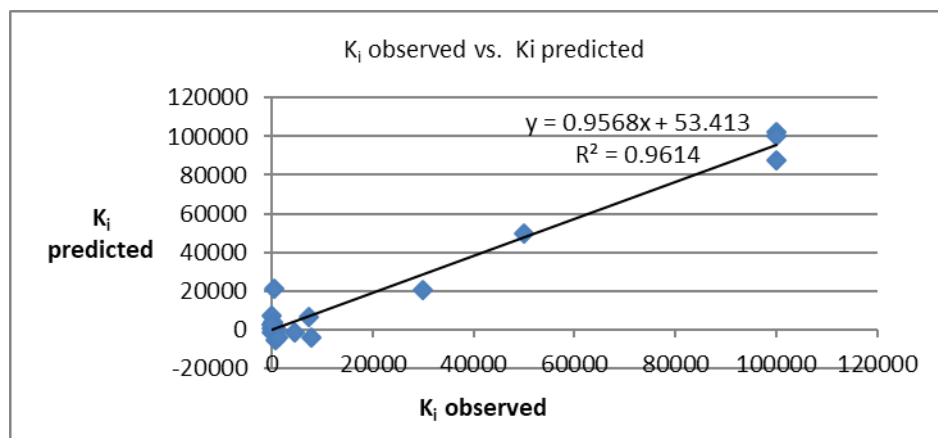
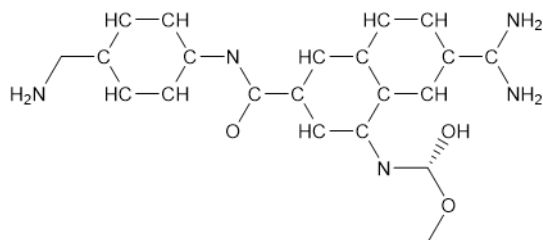


Figure 1. Inhibition constant (K_i) introduced versus ANN model prediction.

The relevance of the descriptors used in the model is as follows, in descending order: aromatic, the number of hybridized h3 carbon atoms, Andrews charges, the number of rings, the number of H atoms, the minimum distance between a donor group and a hydrogen acceptor, the Wiener index, OS number, the minimum mean distance between hydrogen donor groups.

The classification of the similarity with the compound of molecular formula C₂₁H₂₂N₅O₃ (K_i = 0.9) is shown in the figure below.



Ligand 9 hexahydro-naphthalene methanol.

Often, compounds that have been designed by drug-based design (in which the structure of the protein and the protein-ligand complex has been used to design better inhibitors) suffer from poor bioavailability and remarkable toxicity. The problem of bioavailability is often associated with low solubility: suitable inhibitors are most often "at home" in the hydrophobic majority environment of a typical enzyme active site and thus often have several polar groups on them. Increasing the number of polarized groups on the ligand will often improve its solubility by two orders of magnitude or more. The efficacy of the ligand as an inhibitor will usually be diminished - a 10 nM K_i inhibitor could be changed to a K_i = 50 nM inhibitor - but this is often seen as an acceptable compromise. Increased solubility will often reduce toxicity, as the compound is less likely to accumulate as an insoluble mass. Reducing the

toxicity of a drug is not an easy task. Some molecules recovered from the five-Lipinski rule of screening have drug-like properties # 4 known as Famotidine and # 11 as d-Phe-d-Phe-d-Nle-d-Arg-NH₂. It is a well-known peptide sequence with kappa opioid agonist effect.

II.6.3.4 Conclusions

ANNs were superior in predicting bioactivity compared to multiple linear regression (MLR). A model with r^2 of 0.9614 resulted using a library of 42 compounds. The ANN model has been successfully used to classify the data library similarity. The ANN model predicted in a realistic area that Ki did not get any error or bias value, all being predicted in the field of Ki's activity. The interaction of famotidine with urokinase was successfully predicted by Ki values that suggested a pharmacological action and the need for its IV administration in the urokinase-associated anaphylactoid reaction.

II.6.4 Cyclin-dependent kinase two inhibitors: a QSAR artificial neural regression study

II.6.4.1 Introduction

Cyclin-dependent kinase (CDK) plays a significant role in regulating cell dynamics. Kinases are present in all known eukaryotes, and their functional regulatory pathway has been evolutionarily conserved, suggesting that this pathway plays a dominant role in controlling cell growth and disrupting it can lead to cell death. 3D-QSAR simulation strategies, molecular simulation, and molecular dynamics (MD) strategies were applied to investigate the molecular interaction between active ligands and cyclin 2-dependent kinase (CDK2). A QSAR model was calculated using the regression of the neural network (ANN), with excellent predictive capacity in internal and external validation. The results were compared with the CDK inhibitory prototype, Staurosporine. A mixed analysis incorporating the QSAR model, molecular docking and molecular dynamics allowed the definition of the high-definition pharmacophore. The data provided by MD were consistent with the results of the 3D-QSAR model.

The main biological activity of Staurosporine is the competitive binding of protein kinase against ATP, being a model of the competitive ATP kinase inhibitor; however, it also has a high affinity for many other CDKs. In other words, its lack of specificity prevented the use of Staurosporine in the clinic (Chae et al., 2000). Despite the progress in discovering several CDK2 inhibitors, there is still a chemical space available for potent and selective CDK2 inhibitors. The difficulties are in achieving isomorphous specificity (Eaton et al., 1995), which targets specific cells or tissues and ensures a correct degree of inhibition (Elmore, 2007). So far, the interaction between CDK2 and ligands is not fully understood, and the associated mechanism is not yet explained. The considerable variation in the binding affinities of these compounds with CDK2 and the relationship between the biological activity and the "flap"

movement of the enzyme, as well as the conformational changes in the catalytic site of CDK2, were investigated using a combined approach that includes docking and molecular dynamics. The activity data were taken from an original series of 264 compounds (Liu et al., 2007), obtained by isothermal titration calorimetry (Pierce et al., 1999) (ITC) against CDK2. The structures were treated as follows: only very active compounds with $K_d < 10,000$ nM were considered. Structural isomers were not considered. Following these steps, 26 final compounds resulted. The above selection was made to obtain a model based on non-congeneric compounds to expand the chemical space. However, inactive compounds were not considered when constructing the model.

II.6.4.2 Computational methods

A dataset of 26 molecules with measured K_d (nM) on human cyclin dependent kinase 2 was used to calculate a QASR model by the ANN regression method. The target variable was set to K_d (nM). The dependent variables were as follows: potential energy (kcal / mol), molecular weight (GM), AlogP, polar surface area (PSA), molecular radius (MR), molecular polarity, first Zagreb index (Z11) (Khalifeh et al. , 2009), Wiener index (Wiener, 1947), Xu index (Doucet and Panaye 2010), Gutman topological descriptor (GTM = Z12) (Agnes, 2015), eccentricity (ECC) (Delilers et al., 1999), Lipinski's five rule

II.6.4.3 Results

20 hypotheses were developed and analyzed. After analyzing the alignment of the active ligands and the generated hypotheses, the AADRR19 hypothesis was selected. The selected hypothesis contained a hydrogen atom donor (D4), two hydrogen bond acceptors (A2 and A3) and two aromatic rings (R10 and R12), Figure 1.

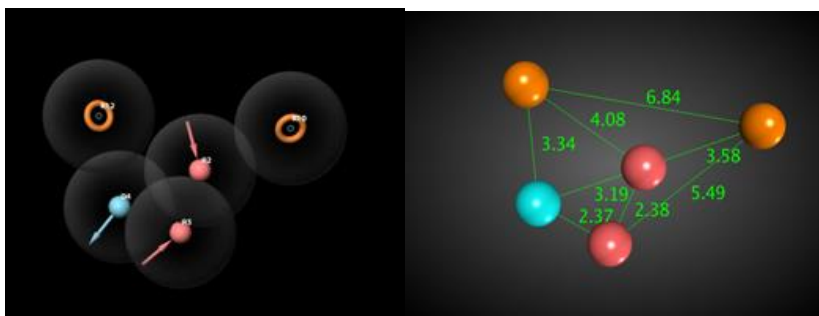


Figure 1. a. Common Pharmacophore for active ligands. The pharmacophore's characteristics are colored: blue donor H, acceptor H pink, orange aromatic ring; b. The distance between pharmacophore characteristics is presented in angstroms.

II.6.4.4 Conclusions

Briefly, a multi-calculation strategy has been used to explore the basic structure inhibition process for a number of CDK2 inhibitors. The attachment was used to generate hypothetical ligand binding modes. The "low" molecular dynamics was used to evaluate the binding mode from the receptor perspective. Both, Staurosporine and compound # 9 have a lipophilic core made of coplanar aromatic rings that fit into the binding pocket. As the QSAR model suggests, it is of minor importance. Both, Staurosporine and # 9 have an outer region that binds to marginal amino acids that act as a "hook" in fixing the compound in place. In addition, the satisfactory agreement between experiment and theory and between the QSAR model and the pharmacophore model (independently constructed) suggested that the QSAR model has a good correlation and predictive power.

II.6.5 CGMP 3', 5'-cyclic 10A phosphodiesterase inhibitors: a QSAR study.

II.6.5.1 Introduction

A series of compounds with IC₅₀, determined experimentally (nM) targeting phosphodiesterase 10A (PDE10A), have been studied. Recently, PDE10A has been proposed as a target for colon cancer. A QSAR model was constructed using compounds having as their target variable their inhibitory effect on PDE10A expressed as IC₅₀ (nM). A multiple correlation technique was used to select the appropriate descriptors for constructing a regression model. The descriptors used were the functional group based descriptors, namely centralization descriptors. The regression model was constructed using artificial neural regression (ANN). A model with $r^2 = 0.9769$ and a standard error deviation of 0.41 was built.

II.6.5.2 Computational methods

To construct a pharmacophore model, a data set of experimental determination of inhibitory activities on compound PDE10A was used (Liu et al., 2007). The compounds were used to generate a common pharmacophore hypothesis. LigPrep was used to prepare ligands using the OPLS 2005 force field, pH set to 7.4. In the construction of the pharmacophore, compounds having IC₅₀ of more than 100 nM were considered inactive for the hypothesis to have discriminating power, that is, to be able to favor compounds that specifically act on PDE10A and have an inhibitory effect, in contrast to other inhibiting compounds. The hypotheses were validated using a bank of false dockers (Bauer et al., 2013). The best hypothesis was chosen for the screening of specific PDE10A inhibitors. The ZINC database was used with an MRSD constraint of 0.7; 10 rotational limits and molecular weight between 100-500 Daltons. The resulting structures were represented in the table together with the predicted IC₅₀ (nM). The phase vector score was used to evaluate ligands (Murugesan et al., 2014).

II.6.5.3 Results

The pharmacophore hypothesis (Figure 1) recovered by Prism, selected according to the classification of the 12 hypotheses, was ARRRH 67. The pharmacophore characters are: A1 x -1.40; y 2.12; z -2.79; R7 x -2.79; and 2.37; z -0.68; R8 x 0.55; y -1.33; z 1.37; R10 x 1.36, y -2.73, z 3.04; H4 x 1.11; and -3.19; z 5.80. The ligands used are shown in Table 1.

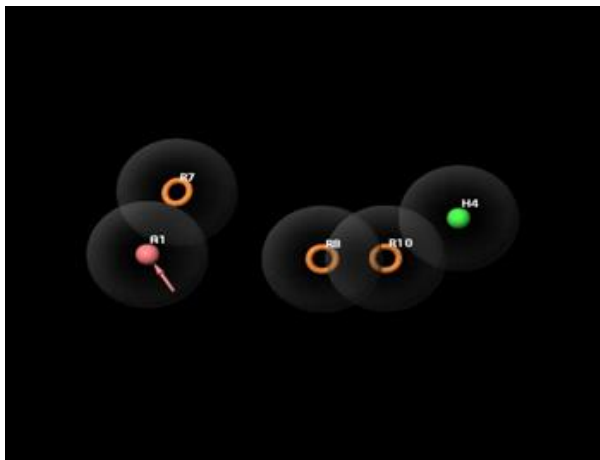
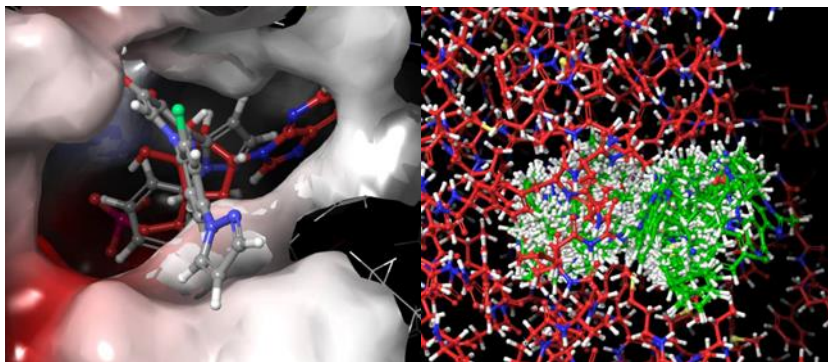


Figure 1. A1 acceptor of H, H4 green hydrophobic group, R ring aromatic group

The docking procedure was successful. None of the compounds were taken out of the binding pocket (see figure). A docking was superimposed on the crystallographic model of PDE10A (PDB id 2OUN) in complex with AMP, which confirmed the reliability of the positions (see figure).



a. PDE10A binding pocket containing AMP (red) superimposed with TAK-069; b. All compounds in the PDE10A binding pocket.

II.6.5.4 Conclusions

Both the QSAR model and the pharmacophore showed the dominant role of HA groups in the biological activity of the ligand as selective PDE10A inhibitors. The QSAR model supports this assumption with a contribution of r^2 0.7 of HA (HA, HA-HA-Max, HA-HA-Mean) to the model and the correlation of the estimated values of the model with the experimental values. The descriptors of centralization explained the rest of the correlation. The pharmacophore has resumed a hypothesis in which the HA group plays a major role. The docking study highlighted the role of Pi-Pi interactions with Phe 696.

II.6.6 Exploring bioactivity - the toxicity space

A theoretical toxicity study on "friendly" biologic macromolecules

II.6.6.1 Introduction

Polyethyleneimine (PEI) and polyamidoamine (PAMAM) are widely used in fields such as biochemistry, biotechnology, and nanotechnology. These molecules, through intravenous administration, come into contact with biological systems or whole organisms. Their ADME properties, toxicity, and reactivity are widely tested. Such molecules are subject to functionalization and derivatization.

Toxicity and reactivity of congeneric compounds can be determined by computational methods, such as QSAR. In this study, QSAR-based toxicity models and virtual toxicity screening (vHTTS) are applied to some (new) zero-generation synthesized dendrimers to assess their toxicity. Several laboratory-synthesized zero-generation "ZAC" core dendrimers have been studied computationally for toxicity assessment. Chemical synthesis of dendrimers was performed using three different derivatives of bromomethylbenzene as aromatic core and dialcanolamines as branching units with different carbon chain lengths.

II.6.6.2 Computational methods

The studied ZAC dendrimers were chemically synthesized according to the literature procedure, described by Füstös et al. 1,4-bis (bromomethyl) benzene, 1,3,5-tris (bromomethyl) benzene and 1,3,5-tris (bromomethyl) -2,4,6-trimethyl benzene were used as an aromatic core. The derivatization was performed with commercially available diethanolamine (DEA) and "ad-hoc" prepared dipropanolamine, dibutanolamine and dipentanolamine.

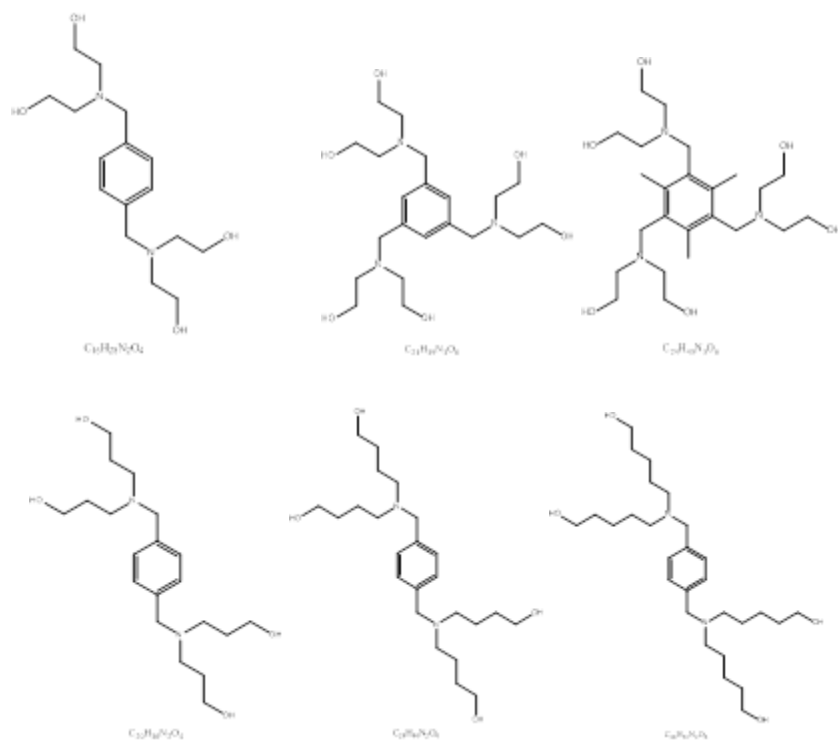


Figure 1. Synthesized ZAC dendrimers for vHTTSPS. On G0, G1, C16H28N2O4, and C20H36N2O4.

II.6.6.3 Results

PAMAM and PEI properties were analyzed. Figure 2 illustrates the space of the QSAR property (Hansch et al., 2004). The chemical shapes of the space are relatively identical for all PEI, PAMAM and ZAC dendrimers

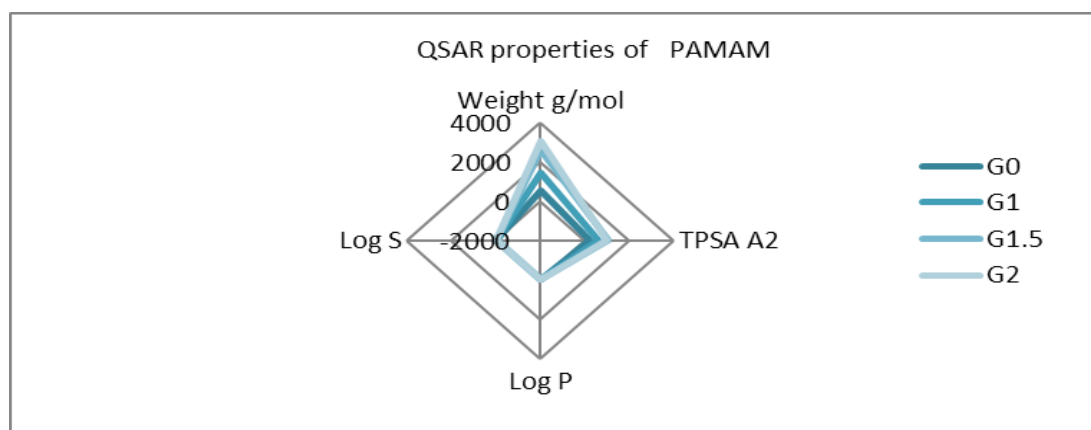
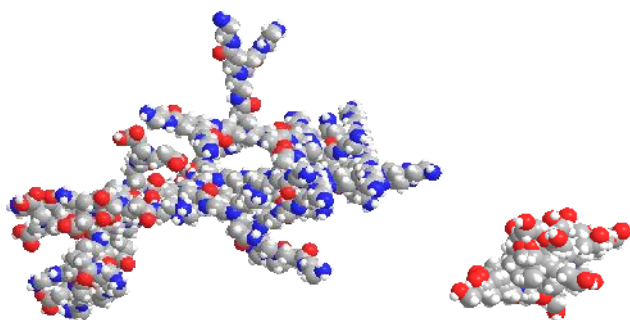


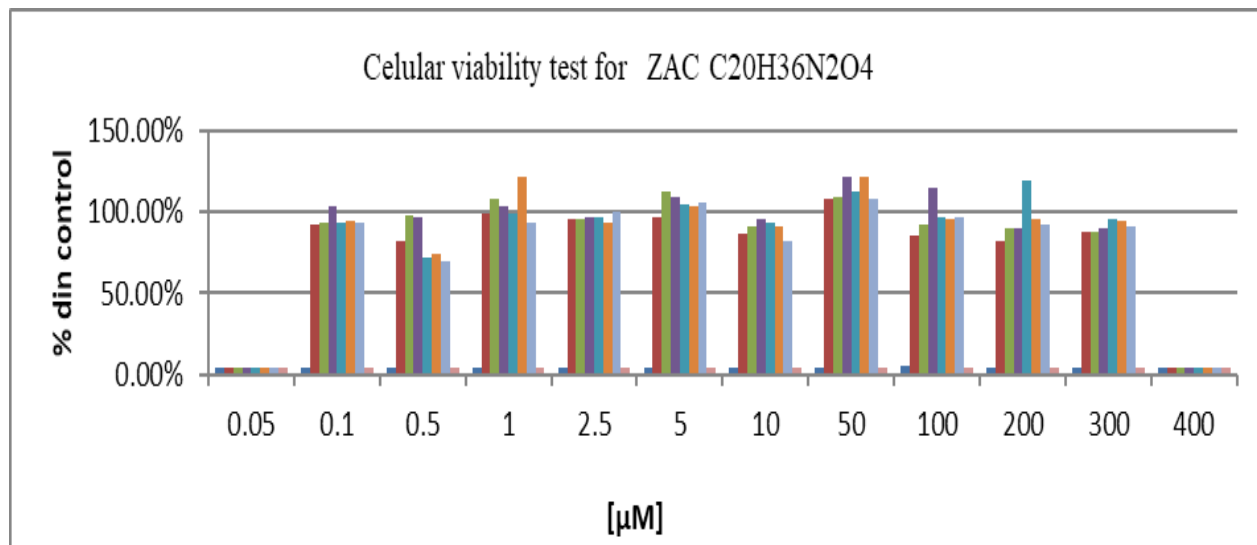
Figure 2. QSAR properties of PAMAM / PEI and ZAC dendrimers.

PAMAM and PEI analyzed for the genotoxic and carcinogenic effect were found not to have such an effect. The Ames test was negative for all compounds (performed on *S. typhimurium*). The biodegradability test of the structure showed for all compounds a persistent chemical of class 2 (tertiary amine). The tests for C 450 showed for PAMAM N-dealkylation and N-oxidation. In vitro analysis of the micronucleus showed a class 1 H acceptor pathway for the PAMAM series. A multiple linear regression model was calculated using PEI and PAMAM. The calculated pharmacophores are shown in Figure 3.

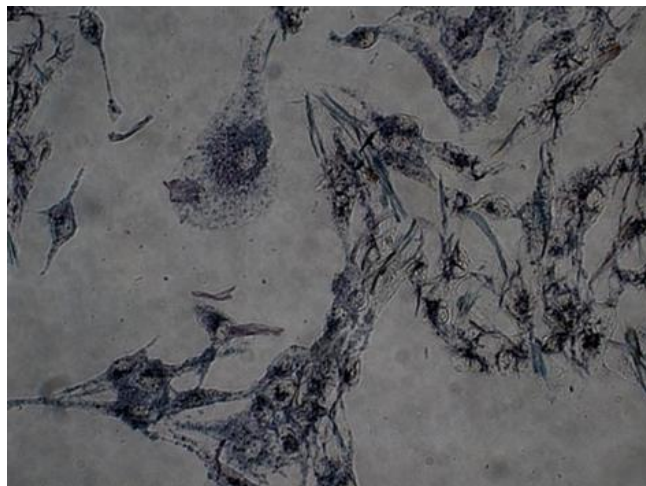


Structures of pharmacophores from left to the right: PEI / PAMAM and pharmacophore ZAC

The data show that ZAC dendrimers are less recommended than PAMAM / PEI compounds. The toxicity test performed on G0, G1, C16H28N2O4, and C20H36N2O4 showed no toxicity. All culture cells were viable. Furthermore, it was observed that ZAC also showed intracellular contribution without exhibiting any toxic effect.



Analiza viabilității celulare care arată o linie celulară neafectată de concentrațiile crescute de C20H36N2O4.



II.6.6.4 Conclusions

PEI and PAMAM have no chemical reactivity in the body, nor have genotoxic and cytotoxic properties. All compounds are class 2 persistent chemicals (tertiary amines) and are excreted as they are. Tests for C 450 showed for N-dealkylation and N-oxidation reactions PAMAM possible interactions with drugs metabolized by the same cytochrome P450. The " in vitro " test of the computational micronucleus showed a way of accepting the H-receptor 3 H-class 1 for the PAMAM series that expresses its potential interaction with the genetic material. The synthesized ZAC dendrimers are 1000 times stronger than PAMAM / PEI in terms of dose and daily dose, as predicted by the QSAR model, which showed a decrease in the dose of ZAC dendrimers compared to PEI and PAMAM. The chemical space of the QSAR properties of the model and the ZAC compounds (see Figure 2) has the same shape, and thus the model prediction is probably correct. Moreover, by the calculated topological properties of both pharmacophores, it was suggested a slight overlap of the formation pharmacophore (PAMAM / PEI) and the pharmacophore of the ZAC dendrimers.

II.6.7 Exploration of the bioactivity space - antimicrobial activity

Penicillin-binding protein in the mechanism of resistance to *Staphylococcus aureus*

II.6.7.1 Introduction

Since 1961 methicillin-resistant *Staphylococcus aureus* (MRSA) has appeared. Presence of protein penicillin-binding: protein 2A (PBP2A) is responsible for the antibiotic resistance of MRSA. Currently, the medication of choice for the treatment of MRSA infections are Vancomycin, Daptomycin, Ceftriaxone and Linezolid. Even with controlled use, strains with reduced sensitivity or direct resistance to these drugs have emerged (Stapleton and Taylor, 2002). Oxacillin is a penicillinase-resistant β -lactamine. Because it

is resistant to penicillinase enzymes, such as that produced by *Staphylococcus aureus*, it is widely used clinically in the US to treat penicillin-resistant *Staphylococcus aureus* infection (Lowy, 2003). However, with the introduction and widespread use of both oxacillin and methicillin, antibiotic-resistant *Staphylococcus aureus* strains are MRSA / ORSA (methicillin and / or oxacillin resistant) and have become increasingly widespread worldwide. MRSA / ORSA is treated using vancomycin. In this study ORSA was considered. Antibiotic sensitivity was performed by the Kirby-Bauer disc diffusion method and by determination

II.6.7.2 Experimental methods

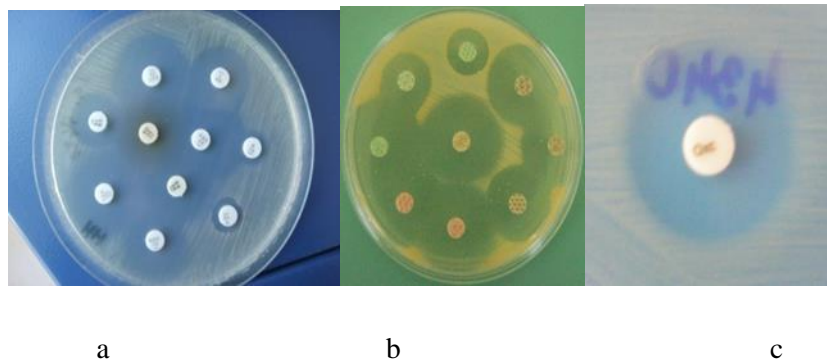
Antibiotic sensitivity of *S. aureus* strains was performed using the Kirby-Bauer disc diffusion method. All isolates were tested using Cefotaxim disk diffusion (Cefotaxim 30 μ g / disc) and MIC methods of oxacillin (1 μ g / disc). The interpretative criteria (in μ g / ml) for Oxacillin MIC according to CLSI were: ≤ 2 (μ g / ml) susceptible to oxacillin sensitivity - MSSA; 2-4 mg/ml intermediate N / AB reversible points; ≥ 4 (μ g / ml) resistant to MRSA oxacillin.

II.6.7.3 Computational methods

Oxacillin complexes with PBP 1,2a, 3,4 were compared by comparison. The complexes were obtained by docking using Autodock. Conformational analysis of the allosteric region was performed by introducing the region motion around O4414 (by deleting O4104 and O44145). A conformational energy diagram (Kcal/mol) was obtained. Molecular recognition is treated here as the amount of water displaced by the ligand (Oxacillin). The binding sites of all four types of PBP potentially found in *S. aureus* were studied. For each type of PBP, 6 binding sites were analyzed. The binding points were treated as cavities. Each cavity was populated with a maximum number of water molecules in terms of fluid characteristics (ie number/unit volume/water molecule was respected) to mimic the flooding of each cavity with water. The resulting water clusters were further analyzed. A molecular dynamics (MD) study was performed on PBP2a, to explain the binding site characteristics. Clusters were minimized using Amber 10 force field, water molecules were considered rigid, RMS was set to 0.1kcal / mol / \AA^2 . MD was performed after the protonation of the clusters at 320°K, pH 7.4 and a salt concentration of 0.0154 M. This step was necessary for the atoms to cleave and orient each water molecule properly. The temperature target for MD was set to 320° K, with a starting point of 320.15° K, and the same parameter with the same parameterization was used. The motion equation used was (Noise-Poincare-Andresen) with an equivalent equilibrium point of 250 iterations and a sample time of 0.5 picoseconds. The water molecules were considered flexible, and the time step was chosen 0.001 NPA with 57 steps. These procedures were used to obtain 57 water

clusters corresponding to the observed MICs. Theoretically for each MIC there must be an exponent in terms of clusters. At each stage the number of clusters differs.

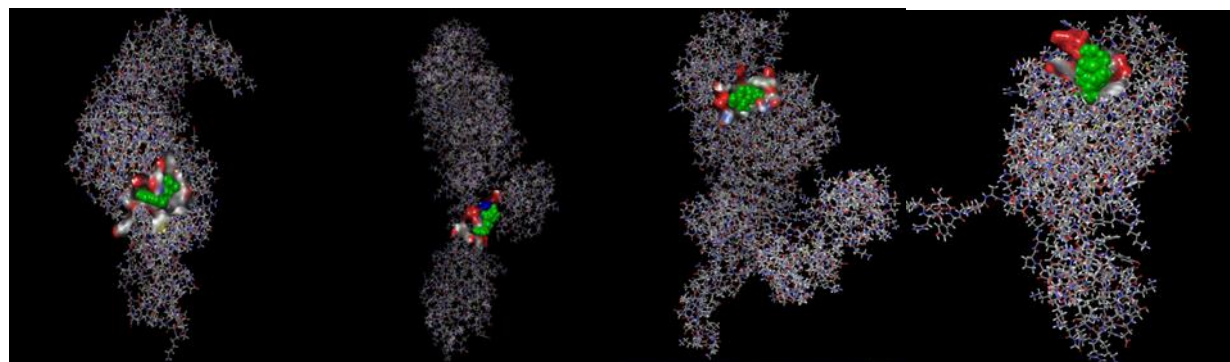
II.6.7.4 Results



From left to right a.b. *S. aureus* strain of mic. c. Oxacillin MIC for *S. aureus* MIC strains - minimum inhibitory concentrations.

Oxacillin MICs of the strains studied ranged from (I) 0.38-2 $\mu\text{g} / \text{ml}$ for 21 strains (44.68%) identified as MSSA; (II) 2-4 mg / ml for 1 strain (2.13%) identified as borderline methicillin-resistant *S. aureus* (BL MRSA); (III) $\geq 4 \mu\text{g} / \text{ml}$ for 25 strains (53.2%) identified as MRSA.

The results of oxacillin docked with each type of PBP are shown in Figure 4. It is observed that PBP2a has the largest volume cavity, but also the most profound site "which is not accessible to the ligand."



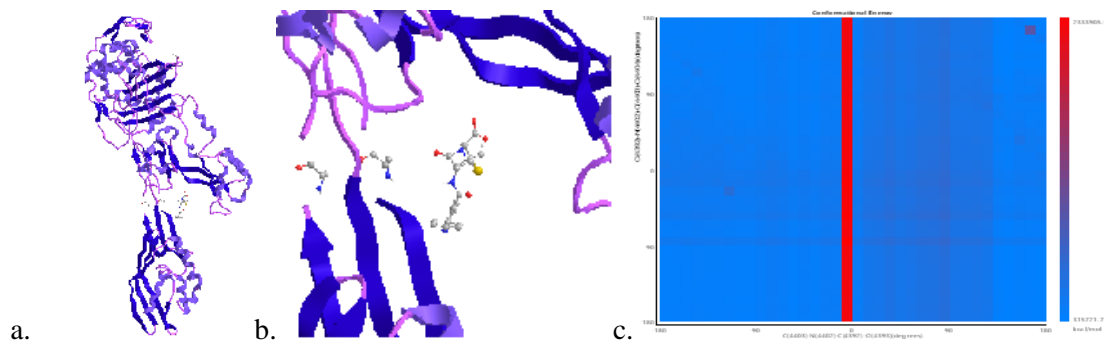
Oxacillin complex with PBP 1,2a, 3,4 from left to right. Oxacillin is shown in green.

The analysis of the cavity of a single unit (unit A was chosen) of each PBP 1,2a, 3 and 4 respectively showed the binding of Oxacillin preferentially to the level of cavity 1, which has the highest volume.

The main cavities for PBP *S. aureus* represented as groups of water are presented below.

Molecular dynamics data showed that they are strongly correlated with MIC values. Several linear regression analyses showed the following:

Moreover, a conformational study carried out on PBP2a by mobilizing the loop responsible for the resistance to Oxacillin revealed a conformational energy map, divided into two areas: a vast energy advantageous domain (blue) and a conformational energy domain, negative (red).



Penicillin-binding protein in complex with Oxacillin Overview (a) and detail (b). Conformational energy map of the allosteric region of PBP2a (c).

II.6.7.5 Conclusions

The study of molecular dynamics and multiple linear regression explains, using a group of water models, the mechanism of action of Oxacillin on *S. aureus* and, consequently, its resistance. These results correlate with the experimental data. The calculated model showed an $r^2 = 0.9119$.

General conclusions

The topological space of the bioactivity is formed around the process of molecular recognition. The bioactivity space is regulated by the ligand space and the chemical receptor. The complete virtual application of the bioactivity space of a particular bioactive complex means that the ligand will have 100% specificity and selectivity for the receptor. Moreover, the complementary conclusion is that the active site of the receptor of the cavity will "react" only to this specific single ligand and there will be no other binding sites for other molecules for the respective ligand. Further, in connection with these

hypothetical assumptions, the biological effect of such a complex formed on living tissue will be precisely localized, very specific and permanent; the tissue involved will not show any "saturation" of the complex.

The concept of molecular recognition was developed in 1990 from the "Zauberkuugel" ligand - the magic bullet (Heynick, 2009). This was the primary paradigm used in the design and discovery of therapeutic molecules. Over the years, this concept has been changed to "methyl magic" when QSAR / QSPR approaches were developed. In the last decades, this concept of modeling the ligand's bioactivity space has evolved. The ligand is now regarded as a complex that will interact with its target.

In the last year, progress has been made in designing molecular devices - nanodevices that are composed of a target recognition section (usually a specific antibody), an active element (enzyme, proper ligand) and a carrier (nanoparticle), parts in an active bio-nano device. These complicated machines are built to access the bioactivity space exclusively from the ligand. Contrary to expected results that increase the specificity of the ligand for its target up to a certain threshold, it results in increased adverse effects. Clonidine is one such example. The molecule is the prototype of the "perfect" agonist for the adrenergic α_2 agonist (Cristea, Medical Publishing, 2012). As presented in Chapters 1 and 2 (original contributions), both the ligand and the receptor can access the bioactivity space. In essence, the ligand can model the receptor cavity and as shown in the case of the FAD enzyme, the amino acid fragment around the FAD determines the complex function. Accordingly, the total bioactivity space (absolutely) to have perfect "bioactivity" must be accessed simultaneously by both components (ligand and receptor). Under real physiological conditions, the bioactivity space is not fully accessible. The shape of the bioactivity resembles a topological space formed by both ligand and receptor. Therefore, this space is not a homeomorph with the space of the central cavity of the receiver (the binding site). It is an intersection of interactions. As shown in Chapter Four, this topological surface of bioactivity can be represented by the topological surface generated using the Cartesian coordinates of the ligands. As surface receivers are expected to have a rigorous modeling effect, they must be considered. Knowing the topology of bioactivity, any property related to it can be predicted.

In Chapter Three, a small part of the bioactivity space, namely the conformational space, is presented as being used to predict the potency of a particular ligand. Chapter four uses bioactivity surface descriptors for accurate prediction of bioactivity that correlates with experimental data.

Recently, receptor optimization is being considered to maximize the accessibility of the bioactive space. Due to the technical impossibility of modeling all the receptors, this optimization tends to be designed on the ligand itself. So, a last-minute ligand has to access not only the bioactivity space but also the receptor

space. Remarkably in absolute terms, the "perfect" ligand as a drug must have the ability to access the bioactivity space of the receptors as well.

Such experimental or computational techniques for exploring the bioactivity space of the receptors are homology modeling and protein folding as discussed in Chapter Five. As demonstrated in Chapter Four, Cartesian coordinates can be used to locate the coordinates of bioactivity, namely, to describe its location.

On both molecules, there are privileged regions with known Cartesian coordinates, which are crucial in accessing the bioactivity space. In the case of the ligand, these regions are located at the level of H / acceptor groups, aromatic ring regions, hydrophobic regions, heteroatoms. At the receptor level, that point is localized to the amino acids that bind to the ligand, to the amino acid motifs. These points (regions) with known position are used to develop the hypothesis regarding a particular property of the molecule (ligand, receptor). These assumptions are used to construct pharmacophores (virtual constructs that are probably "perfect" tools for accessing the bioactivity space). Considering lately the reciprocity of the bioactivity topology, such constructs are available to the receptor itself, the pharmacophore construction algorithm using the surface form of the binding of a particular receptor ligand is already implemented.

The topology of the bioactivity is transposed into the real world using the QSAR / QSPR methodology which allows, as shown in Chapter Six, the rational exploration of the space of the topological bioactivity and explains the basic concept of QSAR according to which similar molecules have similar bioactivity. Excellent examples are given by Patany and Brown (Patany, 1996 and Brown, Wiley, 2012) in reviews on bioisosterism.

The exploration of the bioactivity space can be performed using computational or experimental methods. Frequently, these methods are combined, as shown in Chapter Six, where toxicity and antimicrobial activity is explored, to improve the topology of the bioactivity image.

The fundamental problem is to define a small difference at the molecular level - and to avoid the SAR paradox that "not all similar molecules have similar activities" (Chen et al., 2013). The paradox could be solved by using the topological surface (including ligand and receptor) as a tool for defining molecular recognition and predicting a particular topological property (bioactivity topology).

Finally, the following conclusions are drawn:

1. Ligand adaptability is a crucial step in molecular recognition. Adaptability depends on the structure of the ligand. Linear ligands are favored compared to branched ligands.

2. The amino acid portion has a crucial function in an active complex function. Variation of the fragment around the same ligand determines different enzymatic properties. The function of GOx is dependent on temperature and pH.
3. The conformational energy (of the ligand) is a critical factor in predicting the binding conformation and the potency of a ligand. Cluster-based molecular dynamics is a valuable tool in evaluating the power of a conformation based on a scoring algorithm.
4. Cartesian coordinates can be used to generate a topological surface of a ligand-receptor complex. The polynomial discriminant and the Laplacian operator are good predictors of bioactivity.
5. Protein folding plays a crucial role in molecular recognition. Nanostructures interact with the bioactivity of the enzymes by modifying the folding and consecutive capacity of the amino acid fragment.
6. QSAR / QSPR is a valuable tool in molecular design.
7. Multiple linear regression is a relevant methodology in exploring the bioactivity space of Oxobutane benzamide derivatives.
8. Artificial neural networks have accurately predicted the bioactivity of urokinase inhibitors.
9. The regression methods can be used in the process of exploiting the "drug". These methods used showed novel compounds with a potential inhibitory role, CDK2, and PDE10A.
10. The methods based on the structural activity used as predictive tools have been validated by the experimental results in the case of PEI / PAMAM in which, both the computational and the experimental methods, have shown that the polymers do not show toxicity.
11. Molecular descriptors, topological and energetic, have helped to optimize the DNA properties of the "DNA intercalary" ligands.
12. Molecular dynamics simulation, combined with docking studies and experimental results, successfully explained the resistance mechanism of *S. aureus*. The multiple regression model used established a correlation between site binding water groups and MIC of Oxacillin.

List of publications

Some of the results presented in this thesis are the subject of several research articles, and some of the results have been exposed to international conferences in the form of communications or posters.

Publications:

1. **Lungu CN.**, Bratanovici BI., Antoci GM., Mangalagiu II., Hybrid imidazole-pyridine derivatives: an approach to novel anticancer DNA intercalators. *Curr. Med. Chem.* 2018, doi 10.2174/0929867326666181220094229
2. **Lungu C N.**, Diudea M V., Putz M V., Grudziński I P., Linear and Branched PEIs (Polyethylenimines) and Their Property Space. *Int. J. Mol. Sci.* **2016**, *17*, 555. doi 10.3390/ijms17040555
3. **Lungu C N.**, Diudea M V., Putz M V., Ligand Shaping in Induced Fit Docking of MraY Inhibitors. Polynomial Discriminant and Laplacian Operator as Biological Activity Descriptors. *Int. J. Mol. Sci.* **2017**, *18*, 1377., doi :10.3390/ijms18071377
4. **Lungu C N.**, Diudea M V., Binding site and potency prediction of Teixobactin and other Lipid II ligands by statistical base scoring of conformational space maps, *CCADD* **2018**, *14* (1): 29-34. doi: 10.2174/1573409913666170927113813
5. **Lungu CN.**, C-C CHEMOKINE RECEPTOR TYPE 3 INHIBITORS: BIOACTIVITY PREDICTION USING LOCAL VERTEX INVARIANTS BASED ON THERMAL CONDUCTIVITY LAYER MATRIX. *Studia UBB Chemia.* **2018.**, *63*(1):177-188., doi10.24193/subbchem.2018.1.13
6. **Lungu CN.**, Ersali S., Szeffler B., Moldovan AP., Basak SC., Diudea MV. DIMENSIONALITY OF BIG DATA SETS EXPLORED BY CLUJ DESCRIPTORS. *Sudia UBB Chemia.***2017**, *62*(3): 197-204., doi: 10.24193/subbchem.2017.3.16
7. **Lungu CN.**, Paisz C., Füstös ME.,Orza A., Diudea VM., Grudzinski IP., A PREDICTIVE TOXICITY STUDY OF peiS, PAMAMA AND ZAC DENDRIMERS. *Studia UBB Chemia.* **2019**, *64* (2):499-508., doi: 10.24193/subbchem.2019.2.42
8. Majumdar S., Basak CS., **Lungu CN.**, Diudea MV., Grunwald D., Finding needles in a haystack: determining key molecular descriptors associated with the blood-brain barrier

entry of chemical compounds using machine learning. *Mol. Inform.* **2019**, doi: 10.1002/minf.201800164.

9. Majumdar S, Basak SC, **Lungu CN.**, Diudea MV., Grunwald GD. Mathematical structural descriptors and mutagenicity assessment: a study with congeneric and diverse datasets. *SAR QSAR Environ. Res.* **2018**, 29(8):579-590., doi: 10.1080/1062936X.2018.1496475
10. Diudea MV., **Lungu CN.**, Nagy CL. Cube-rhombellane related bioactive structures. *Molecules.* **2018**, 23(10) E2533., doi: 10.3390/molecules23102533

References

1. Carlson A H., McCammon A J., (1999), Accomodating protein flexibility in computational drug design, *Molecular Pharmacology* 57: 213-218
2. Berg J M., Tymoczko L J., Lubert Stryer L., (2002), *Biochemistry* 5th edition, New York, W.H.Freeman, ISBN-10:0-7167-3051-0
3. Atkins P., De Paula J., (2006). *Atkins' Physical Chemistry* (8th ed.). W. H. Freeman.:200–202.
4. Lo K M., Tilgner M., Shi1 P-Y.,(2003), Potential High-Throughput Assay for Screening Inhibitors of West Nile Virus Replication, *J Virol.* 77(23): 12901–12906.
5. Nantasenamat C., Isarankura-Na-Ayudhya C., Naenna T., Prachayasittikul V., (2009), A practical overview of quantitative structure-activity relationship". *Excli J.* 8: 74–88.
6. Todeschini R., Gramatica P., (1998), New 3D Molecular Descriptors: The WHIM and QSAR Applications. *Perspectives in Drug Discovery and Design*, 9-11, 355-380.
7. Mukherjee S., Zhang Y.,(2011), Protein-protein complex structure predictions by multimeric threading and template recombination, *Structure.*13; 19(7): 955–966.
8. Diudea M V., (1997a) Cluj matrix invariants. *J Chem. Inf Comput Sci* 37:300-305
9. Diudea M V., Florescu M S., Khadikar P V., (2006), *Molecular Topology and Its Applications*, Eficon, Bucharest
10. Leskovic V., Trivić S., Wohlfahrt G., Kandrac J., Pericin D., (2005), *Int J Biochem Cell Biol* 37(4):731-50.
11. Metzler D E., Metzler C M., Sauke D J.,(2003), *Biochemistry*, Academic Press, ISBN 978-0-12—492541-0
12. Henrichfreise B., Schiefer A., Schneider T., (2009), Functional conservation of the lipid II biosynthesis pathway in the cell wall-less bacteria *Chlamydia* and *Wolbachia*: why is lipid II needed., *Mol Microbiol.*73 (5): 913–23.
13. Gualerzi O., Brandi L., Fabbretti A., Pon L C., *Antibiotics Targets, Mechanism and Resistance* Claudio, ISBN: 978-3-527-33305-9 Wiley: 137
14. Kang E S., Ford K., Grokulsky G., Wang Y B., Chiang T M., Acchiardo S R., (2000), Normal circulating adult human red blood cells contain inactive NOS proteins, *J Lab Clin Med.*135(6):444-51.