



SAPIENZA
UNIVERSITÀ DI ROMA

Facoltà di Farmacia e Medicina

Dipartimento di Chimica e Tecnologie del Farmaco

Dottorato di Ricerca in Scienze Farmaceutiche XXVI Ciclo

*Application of Medicinal Chemistry Methods
on Different Classes of Drugs*

Tutor: Prof. Rino Ragno

PhD student: Dr. Flavio Ballante

This thesis is dedicated to my wife,

Emanuela

Index

Summary

Preface	iv
Introduction	v
Chapter I.....	1
3-D QSAutogrid/R: An alternative procedure to build 3-D QSAR models. Methodologies and applications	1
Chapter II.....	21
Comprehensive model of wild-type and mutant HIV-1 reverse transcriptases.....	21
Chapter III.....	34
2-(Alkyl/Aryl)amino-6-benzylpyrimidin-4(3 H)-ones as inhibitors of wild-type and mutant HIV-1: Enantioselectivity studies.....	34
Chapter IV	38
Histone deacetylase inhibitors: Structure-based modeling and isoform-selectivity prediction	38
Chapter V	45
Design, synthesis and biological evaluation of new classes of thieno[3,2-d]pyrimidinone and thieno[1,2,3]triazine as inhibitor of vascular endothelial growth factor receptor-2 (VEGFR-2).....	45
Chapter VI	59
Pharmacophore assessment through 3-D QSAR: Evaluation of the predictive ability on new derivatives by the application on a series of antitubercular agents	59
Chapter VII	78
Hsp90 Inhibitors (I). Definition of 3-D QSAutogrid/R Models as a Tool for Virtual Screening	78
Chapter VIII	106
Hsp90 Inhibitors (II). Combining ligand-based and structure-based approaches for Virtual Screening application.....	106
Chapter IX.....	114
<i>Research Period Abroad I</i>	114
Synthesis of compounds 7a, 7b, 7c as new potential VEGFR-2 inhibitors	116
Biological evaluation for compounds 7a, 7b, 7c.....	130
Chapter X.....	133
<i>Research Period Abroad II</i>	133
HDACIs activity investigation.....	135
Chapter XI.....	164

Scientific Production during the PhD scholarship.....	164
Courses, seminars, conferences attended during the PhD scholarship	166
Chapter XII.....	168
Publications.....	168
Acknowledgements.....	251

Preface

Fascinating: this is the word that best expresses my PhD scholarship.

Fascinating for the research activity, fascinating for the friendships arisen from scientific collaborations, fascinating for the chance to work with different people and in foreign countries, fascinating for the impact that all of this had and still has on me. Express feelings, satisfaction, hope and even the frustrations that the research activity involves is not simple, and transfer them to a thesis is certainly arduous; however, I hope to convey the mix of passion, dedication, perseverance and romance, which have accompanied me through this experience.

During my PhD scholarship I worked mainly on computational chemistry, but also on wet chemistry, like organic synthesis and enzymatic inhibition assays. This has made me aware of the various aspects that characterize the different medicinal chemistry applications, and how these are linked together, allowing not only to immerse myself in the specific laboratory practice, but also to gain an overview of the manifold scientific research process.

I couldn't have asked for anything better.

Introduction

The present doctoral thesis is the result of the work carried out during the three years of my PhD scholarship at the Rome Center for Molecular Design laboratory (RCMD, Department of Chemistry and Drug Technologies, Sapienza University of Rome), under the supervision of Prof. Rino Ragno. The research activity was focused mainly on the design, optimization and application of computational strategies to derive quantitative structure-activity relationships (QSAR, 3-D QSAR, and COMBINE) on different molecular classes of current interest, such as: opioid receptor antagonists (OPAs), Hepatitis C Virus NS5B-Polymerase Inhibitors (NS5B-NNIs), Hystone Deacetylase Inhibitors (HDACIs), Anti-tubercular agents, vascular endothelial growth factor receptor-2 (VEGFR-2) inhibitors, HSP90 inhibitors, HIV-1 reverse transcriptase inhibitors (NNRTIs), Bovine Serum Amine Oxidase (BSAO) substrates, etc... Moreover two research periods abroad were performed: the first framed in a LLP Erasmus program collaboration, was conducted for six months at the Laboratoire d'Ingénierie et Moléculaire Pharmacologique Biochimie (LIMBP) of the Université de Lorraine Metz (France), directed by Prof. Gilbert Kirsch, and characterized by the application of organic synthesis to obtain new thienopyrimidinone derivatives as potential inhibitors of vascular endothelial growth factor receptor-2 (VEGFR-2); the second took place, for three months, at the Department of Biochemistry and Molecular Biophysics in Washington University School of Medicine in Saint Louis (MO, USA), under the supervision of Prof. Garland R. Marshall, investigating the activity profile of new Histone Deacetylases (HDACs) inhibitors by the application of the Mobility Shift Assay Technology.

Main purpose of this doctoral thesis is to highlight the activities carried out in the different research projects, the applied methodologies and the obtained results. The

text starts describing those studies whose results were published in scientific journals (chapters I-VI): the author decided to omit some procedural details, completely reported in the published papers, that would make the text too long, tedious and redundant; therefore readers who want to delve these aspects can also refer to Chapter XII in which is possible to read the original papers; on the contrary for studies that have not yet been published, as those characterizing the Chapters VII and VIII, discussion is adequately detailed. Chapters IX and X report the scientific activities carried out in France and in USA respectively; Chapter XI summarizes all the scientific activities accomplished during the entire PhD course, whereas Chapter XII, as mentioned, contains the published articles.

Chapter I

3-D QSAutogrid/R: An alternative procedure to build 3-D QSAR models. Methodologies and applications

Flavio Ballante and Rino Ragno

Journal of Chemical Information and Modeling 2012 52 (6), 1674-1685

Preamble

This chapter refers to a work that I care particularly, a constantly evolving project which began few years ago: the 3-D QSAutogrid/R procedure.¹ The 3-D QSAutogrid/R is a tool, designed and developed to be freely used from academics, capable to perform optimized three dimensional quantitative structure-activity relationship (3-D QSAR) studies by the means of the PLS algorithm². Respect to the classic 3-D QSAR approaches (like CoMFA³ and GRID⁴/GOLPE⁵) allows to:

- *minimize human-machine interactions, costs, and calculation time*
- *select the most appropriate pretreatment parameters (CAPP¹ procedure)*
- *conduct MPGRS¹ (Multi Probe Guided Region Selection) analyses*

The 3-D QSAutogrid/R procedure is, actually, the default 3-D QSAR methodology used at the Rome Center for Molecular Design laboratory (RCMD, Department of Chemistry and Drug Technologies Sapienza University of Rome), and the core engine to prepare thousands of 3-D QSAR models, based on either ligand based (LB) or structure based (SB) alignment methodologies, forming the very first 3-D QSAR server (www.3d-qsar.com), which allows to model the desired ligand and to retrieve the predicted biological affinity for the chosen model. Several studies,⁶⁻¹⁰ characterizing my PhD scholarship, were accomplished using this procedure confirming its usefulness in drug discovery.

The Procedure

The 3-D QSAutogrid/R methodology¹ is based on the integration of the molecular interaction fields (MIFs) as calculated by the AutoGrid program (Autodock Suite¹¹) and the R statistical environment,¹² a freely available language for statistical computing and graphics. The AutoGrid software (based on the AMBER united-atom Force Field) is used to generate either training and test sets' MIFs: in particular 8 different probes (Table 1) were chosen in agreement with the most common residue atomic composition, and currently implemented.

Table 1. List of the AutoGrid probes employed for MIF calculation.

Probe Type	Description
A	Aromatic Carbon
C	Aliphatic (sp ³) Carbon
OA	Hydrogen-bond-accepting oxygen
HD	Hydrogen bonded to heteroatom
NA	Hydrogen-bond-accepting amine nitrogen
N	Amide nitrogen
e	Electrostatic
d	Desolvation

Several R-based packages (Table 2, Figure 1) were developed to carry out complete 3-D QSAR studies to obtain a comprehensive statistical and graphical report (see Computational methods section) for each different used probe and principal component (PC). The procedure needs only the pre-aligned training and test sets molecules.

What makes the 3-D QSAutogrid/R procedure innovative is the capability to determine, through a combinatorial calculation (CAPP procedure),¹ the most appropriate pretreatment values to get optimized 3-D QSAR models; but especially the possibility to derive multiprobe 3-D QSAR models (MPGRS procedure)¹

highlighting the most informative regions around the ligands, using all the probes together, and reducing the chance of missing important correlations when using single probe 3-D QSARs.

The whole procedure was validated on three data sets, covering both ligand-based (LB) and structure-based (SB) alignment methodologies, previously reported using CoMFA¹³ and GRID/GOLPE,¹⁴ respectively.

Table 2. List of the R compiled packages.

R package	Description
D2M	Data to Model
CAPP	Combinatorial Analysis of Pretreatment Parameters
MDP	Model Data Pretreatment
CV	Cross Validation
VS	Variable Selection
GRS	Guided Region Selection
MPGRS	Multi Probe Guided Region Selection
ESP	External Set Prediction
YS	Y-scrambling

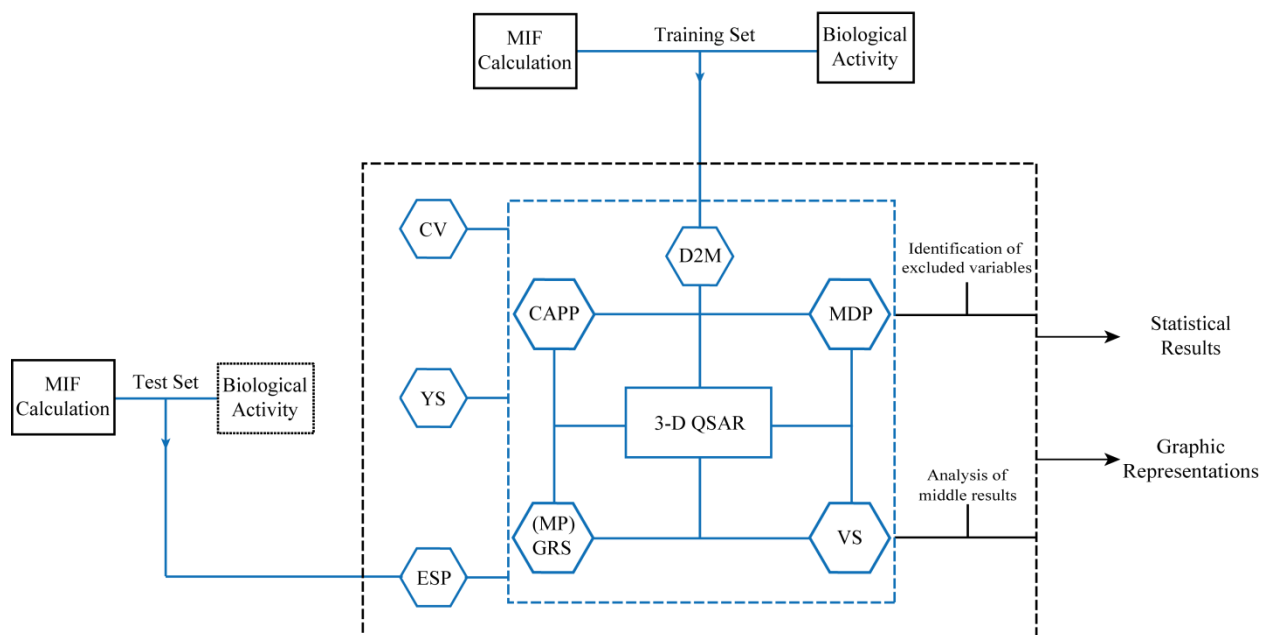


Figure 1. 3-D QSAutogrid/R process workflow. Acronyms inside the blue meshed square refer to the packages (steps) that effectively build or optimize the statistical PLS 3-DQSAR models; others are related to packages that perform analysis on these created models (see below for the description of each package). D2M: “Data to Model”; CAPP: “Combinatorial Analysis of Pretreatment Parameters”; MDP: “Model Data Pretreatment”; CV: “Cross-Validation”; VS: “Variable Selection”; (MP)GRS: “(Multi Probe) Guided Region Selection”; ESP: “External Set Prediction”; YS: Y-scrambling.

Computational Methods

As reported,¹ each R-package is able to perform a specific 3-D QSAR analysis process releasing all the statistical and graphical results (Figure 2):

- 1) *D2M package* (“Data to Model”): allows to import the training set biological data and MIF information (even MIFs generated with different tools, i.e. GRID or CoMFA) to build a raw PLS 3-D QSAR model for each used probe and user-defined principal component (PC).

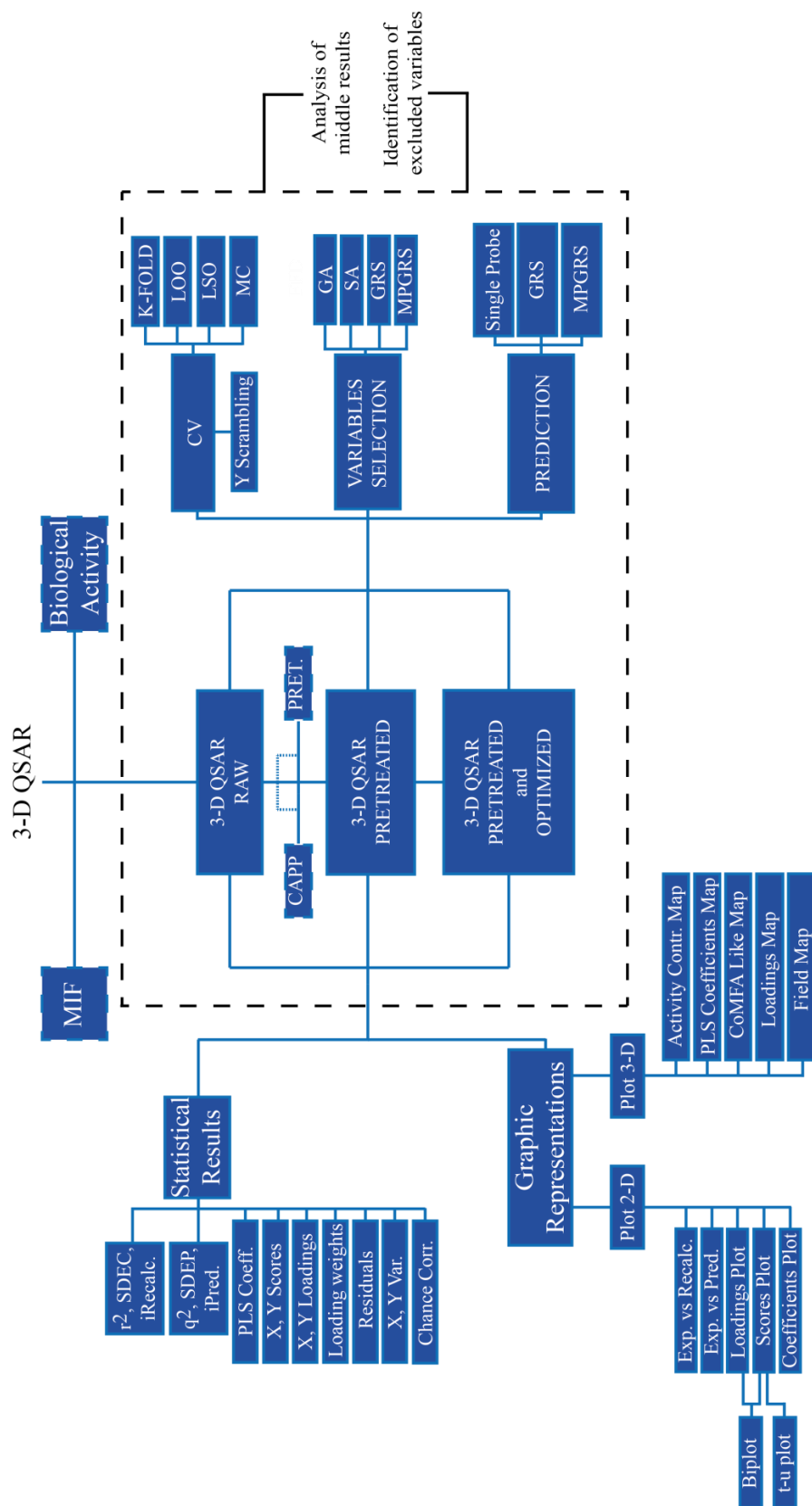


Figure 2. Overview of the information released by the 3-D QSAutogrid/R procedure.

- 2) *CV package* (“Cross Validation”): allows to perform different kind of internal validation methods like LOO (leave one-out), LSO (leave-some-out), KF (k-fold), and MC (Monte Carlo) to:
- assess the chance correlation;¹⁵
 - select the optimal model dimensionality (number of PCs);
 - measure the internal predictive ability by means of statistical coefficients such as cross-validated correlation coefficient (q^2) and standard deviation error of prediction (SDEP);
- 3) *CAPP package* (“Combinatorial Analysis of Pretreatment Parameters”): represents a new 3-D QSAR feature, since it systematically seek the more efficient data pretreatment values (energy cutoff, zeroing of very low data points, and minimum standard deviation cutoff) through a combinatorial analysis: for each combination the pre-treated model is generated and then its q^2 evaluated using LOOCV, LSOCV or KFCV. The optimal pretreatment combination is then selected according to the maximum q^2 and the percentage decrement of sPRESS values between subsequent PCs.¹⁶
- 4) *MDP package* (“Model Data Pretreatment”): to pretreat the molecular descriptors (MIFs), in an arbitrary way or as found by the CAPP procedure. A further data filter, that performs the 1N kind of 2-level variable elimination (variables which take only 2 values in all of the data file, one of which appears only in one object) can be selected.
- 5) *VS package* (“Variable Selection”) allows to reduce redundant data improving the predictability of the statistical model. Different variable-selection algorithms are implemented as: genetic algorithm¹⁷ (GA), simulated annealing¹⁸ (SA) and fractional factorial design¹⁹ (FFD).
- 6) *GRS package* (“ q^2 -Guided Region Selection”): this package, inspired on the previously reported q^2 -guided region selection (q^2 -GRS)²⁰ and smart-region

definition^{20, 21} (SRD),^{20, 21} allows to extract, for each probe and PC, only those sub-areas endowed with a q^2 value greater than a specified threshold.

- 7) *MPGRS package* (“Multi Probe Guided Region Selection”): the MPGRS approach is a new powerful tool capable to conduct advanced 3-D QSAR analyses. Indeed, this procedure is able to sum all the most defining chemical information, obtained by all the mono-probe models, deriving final quantitative pharmacophoric models.
- 8) *YS package* (*Y-Scrambling*): this package allows to detect if the 3-D QSAR model is characterized by chance correlation¹⁵ by means of scrambling procedure²²
- 9) *ESP package* (*External Set Prediction*): this tool is of fundamental importance, in fact: 1) allows to test the predictive capability of the 3-D QSAR models against an external test set of molecules with known biological responses; 2) predicts the activity of untested or not yet synthesized compounds.

Results and discussion

The procedure was validated¹ on three data sets, covering both ligand-based and structure-based alignment methodologies:

- a data set of aligned opioid-receptor antagonists (LB data set)¹³
- two data sets of Hepatitis C Virus NS5B-Polymerase Inhibitors (SB data sets)¹⁴

LB data set (opioid-receptor antagonists):

A data set composed of 74 pre-aligned compounds with associated δ , μ , κ opioid-binding affinities, previously described in a CoMFA application,¹³ was used to build 24 3-D QSAR models with the new procedure, maintaining the original¹³ training set and test set composition. All the 3-D QSARs were built, pretreated

through the CAPP procedure, and optimized by means of the genetic algorithm (GA, package VS). All the models were cross-validated through Leave-One-Out (LOO), Leave-Two-Out (LTO), k-Fold (KF) and Monte Carlo (MC) methodologies and submitted to the YS package to test the presence of chance correlation. To directly compare the two methodologies, double probe models (DP), similarly to those generated by CoMFA, were built. Statistical results (Tables 3 and 4), were similar to those originally obtained (compare Tables 3 and 4 with table 5), confirming the robustness of the methodology. Moreover, the obtained contour maps were in agreement with those obtained from CoMFA. As for example, activity contribution plots obtained from the δ DP models are shown (compare Figures 3 and 4 with Figures 5 and 6).

Table 3. Opioid-receptor antagonists: Autogrid/R PLS models statistical results (CAPP and GA processes were applied).

model	OR	P	PC	r^2	q^2_{LOO}	q^2_{K5FCV}	r^2_{YS}	q^2_{YS}
1	δ	A	2	0.81	0.73	0.70	0.27	-0.37
2	δ	C	2	0.82	0.74	0.71	0.32	-0.35
3	δ	HD	2	0.83	0.75	0.72	0.33	-0.34
4	δ	NA	2	0.83	0.75	0.73	0.31	-0.34
5	δ	N	2	0.83	0.76	0.72	0.29	-0.32
6	δ	OA	2	0.83	0.74	0.71	0.32	-0.37
7	δ	e	3	0.69	0.58	0.56	0.22	-0.19
8	δ	d	3	0.70	0.59	0.55	0.24	-0.30
9	μ	A	3	0.91	0.82	0.76	0.57	-0.50
10	μ	C	3	0.90	0.81	0.78	0.59	-0.50
11	μ	HD	3	0.90	0.81	0.75	0.47	-0.49
12	μ	NA	3	0.91	0.81	0.78	0.59	-0.50
13	μ	N	3	0.91	0.83	0.78	0.52	-0.61
14	μ	OA	3	0.91	0.83	0.77	0.51	-0.61
15	μ	e ^a	1	0.31	0.21	0.20	0.06	-0.10
16	μ	d	3	0.72	0.60	0.52	0.27	-0.39
17	κ	A	2	0.78	0.58	0.49	0.42	-0.37
18	κ	C	3	0.81	0.62	0.55	0.54	-0.53
19	κ	HD	3	0.82	0.72	0.65	0.34	-0.41
20	κ	NA	3	0.80	0.62	0.54	0.55	-0.47
21	κ	N	3	0.80	0.61	0.52	0.54	-0.48
22	κ	OA	3	0.82	0.65	0.59	0.54	-0.44
23	κ	e ^a	2	0.35	0.20	0.18	0.13	-0.18
24	κ	d	3	0.58	0.38	0.34	0.29	-0.36

OR: Opioid-receptor data, P:Autogrid Probe, PC: optimal number of principal components/latent variables, r^2 : conventional square-correlation coefficient; q^2_{LOO} : cross-validation correlation coefficient using the leave-one-out method; q^2_{K5FCV} : cross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations; r^2_{YS} : average square correlation coefficient obtained after Y-scrambling process using 100 iterations; q^2_{YS} : average cross-validation correlation coefficient using the leave-one-out method obtained after Y-scrambling process using 100 iterations ^aThe e models 15 and 23 reported were only pretreated due to too few variables after GA selection.

Table 4. Opioid-receptor antagonists: Autogrid double-probe (DP) PLS models statistical results (only the CAPP process was applied).

model	OR	P	PC	r^2	q^2_{LOO}	q^2_{K5FCV}	r^2_{YS}	q^2_{YS}
28	δ	Autogrid DP	3	0.83	0.70	0.67	0.41	-0.50
29	μ	Autogrid DP	4	0.85	0.65	0.63	0.52	-0.53
30	κ	Autogrid DP	3	0.84	0.67	0.63	0.50	-0.53

OR: Opioid-receptor data; P: Autogrid double probe (DP, C and e probes), PC: optimal number of principal components/latent variables, r^2 : conventional square-correlation coefficient; q^2_{LOO} : cross-validation correlation coefficient using the leave-one-out method; q^2_{K5FCV} : cross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations; r^2_{YS} : average square correlation coefficient obtained after Y-scrambling process using 100 iterations; q^2_{YS} : average cross-validation correlation coefficient using the leave-one-out method obtained after Y-scrambling process using 100 iterations

Table 5. Opioid-receptor antagonists: original CoMFA models statistical results.

model	OR	P	PC	r^2	q^2_{LOO}	q^2_{K5FCV}
25	δ	CoMFA	4	0.91	0.69	-
26	μ	CoMFA	4	0.92	0.67	-
27	κ	CoMFA	6	0.96	0.60	-

OR: Opioid-receptor data; P: standard CoMFA Probe $C_{\text{sp}3}^+$, PC: optimal number of principal components/latent variables, r^2 : conventional square-correlation coefficient; q^2_{LOO} : cross-validation correlation coefficient using the leave-one-out method; q^2_{K5FCV} : cross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations.

By means of the ESP package,¹ all the 3-D QSAutogrid/R models were externally validated, as in the original CoMFA application,¹³ confirming the robustness of the methodology even in predicting the external test sets (compare Table 6 with Table 7).

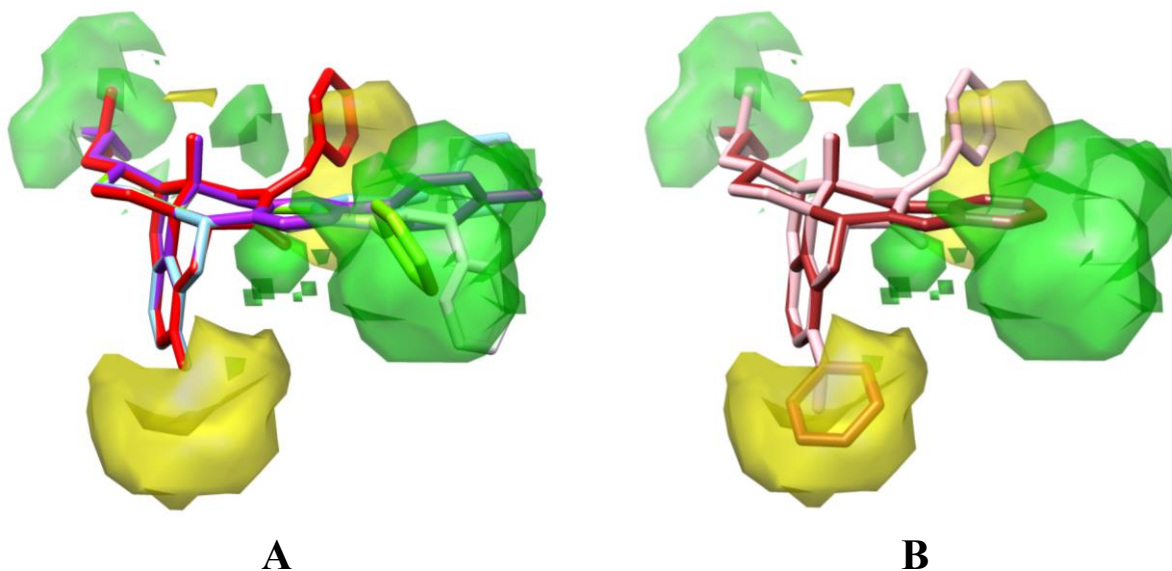


Figure 3. CoMFA-like steric-contour map derived from the C probe for the δ -opioid receptors. A: compounds **18** (sky blue), **20** (white), **22** (green), **50** (purple) and **67** (red). B: compounds **30** (brown) and **68** (pink). Contour levels: 85% (positive green, negative yellow). Hydrogen atoms are omitted for the sake of clarity. In A and B are reported similar contour maps for the new procedure and CoMFA.

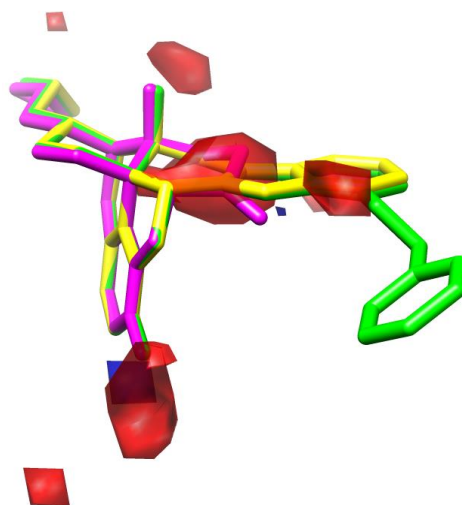


Figure 4. CoMFA-like electrostatic contour map derived from e probe for the δ -opioid receptors. Compounds: **Naltrexone** in magenta, **NTI** in yellow, **24** in green. Contour levels: 85% (positive blue, negative red). Hydrogen atoms are omitted for clarity.

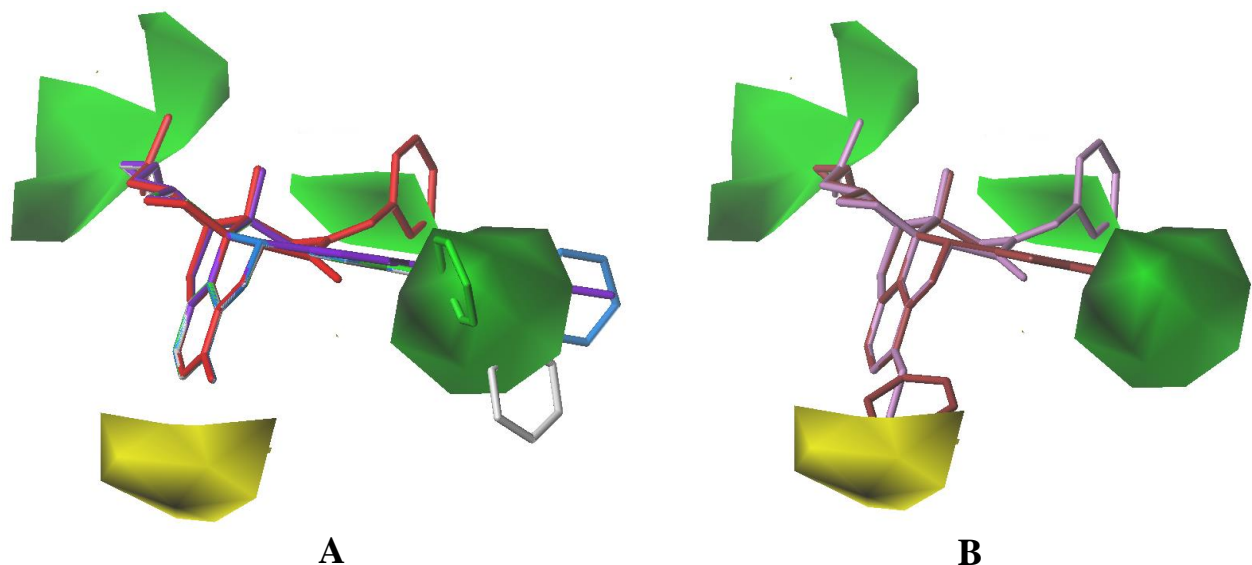


Figure 5. CoMFA steric contour map for the δ opioid receptors. A: compound **18** in sky blue, **20** in white, **22** in green, **50** in purple and **67** in red. B: compound **30** in brown, **68** in pink. Green (favored): 80%, yellow (disfavored) 20%. Hydrogen atoms are omitted for the sake of clarity.

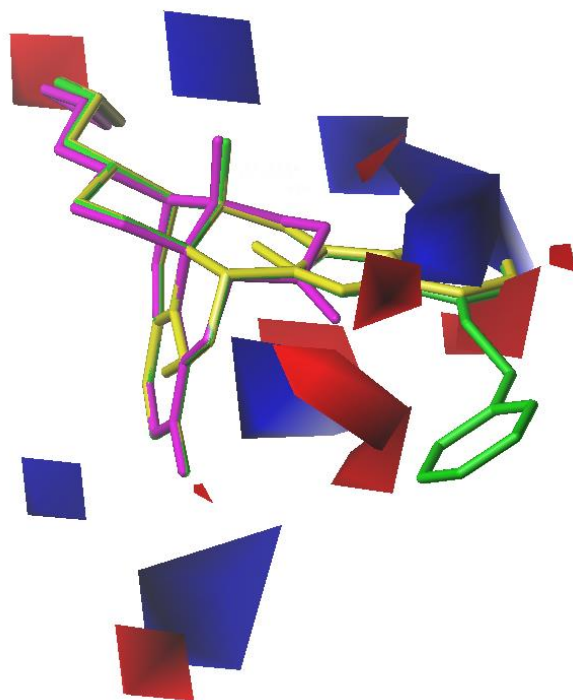


Figure 6. CoMFA electrostatic contour map for the δ opioid receptors. **Naltrexone** in magenta, **NTI** yellow, **24** green, **Blue** favored 80%, red (disfavored) 20%. Hydrogen atoms are omitted for the sake of clarity.

Table 6. δ Test Set predictions indicated by SDEP values.

OR Model	P	PC	SDEP _{TS1}	SDEP _{TS2}
1	A	2	0.66	0.80
2	C	2	0.64	0.77
3	HD	2	0.62	0.74
4	NA	2	0.64	0.82
5	N	2	0.64	0.76
6	OA	2	0.65	0.75
7	e	3	0.81	1.20
8	d	3	0.90	1.12

OR Model: Opioid-receptor model of Table 6; P: Autogrid probe, PC: optimal number of principal components/latent variables; SDEP_{TS1}: standard deviation error of prediction for the original test set; SDEP_{TS2}: standard deviation error of prediction for the external test set.

compd	pK _i			compd	pK _i		
	exptl	pred	res		exptl	pred	res
6	7.10	7.63	-0.53	42	6.29	6.94	-0.55
11	8.33	8.33	0.00	47	9.11	8.24	0.87
13	7.28	7.18	0.10	48	8.75	8.11	0.64
19	7.74	8.28	-0.54	52	6.90	7.00	-0.10
27	7.46	7.83	-0.39	64	7.80	7.94	-0.14
37	7.66	7.41	0.25	70	5.62	6.09	-0.47
39	8.05	8.64	-0.59				

A

compd	pK _i			compd	pK _i		
	exptl	pred	res		exptl	pred	red
6	7.1	7.98	-0.88	42	6.29	7.23	-0.94
11	8.33	8.33	0.00	47	9.11	8.09	1.02
13	7.28	7.23	0.05	48	8.75	8.25	0.5
19	7.74	8.47	-0.73	52	6.90	6.77	0.13
27	7.15	7.43	-0.28	64	7.80	8.28	-0.48
37	7.66	8.31	-0.65	70	5.62	6.37	-0.75
39	8.05	8.81	-0.76				

B

Table 7. Opioid receptor antagonists: Experimental and Predicted pK_i values of the δ Test Set compounds; A: CoMFA values from the reference,¹³ experimental activity for compd 27 was mistyped, the right one is equal to 7.15; B: AutoGrid/R δ CAPP pretreated double probe (DP) model values.

SB data set (Hepatitis C Virus NS5B-Polymerase Inhibitors):

In this application two structure based datasets composed of HCV NS5B non-nucleoside inhibitors (thumb and palm NNI), previously investigated¹⁴ through the GRID/GOLPE methodology,⁵ were submitted to the 3-DQSAutoGrid/R protocol.¹ As for the LB study, the models were built, and optimized via the CAPP procedure. Since no variable selection was originally applied,¹⁴ no further optimization processes were performed on the models. As for the LB case study, the 3-DQSAutoGrid/R procedure proved to derive models comparable to those originally reported¹⁴ concerning either the statistical (Table 8) and graphical results. In this case was decided to show the similarities between the PLS-coefficients plots obtained from the two methodologies (Figure 7).

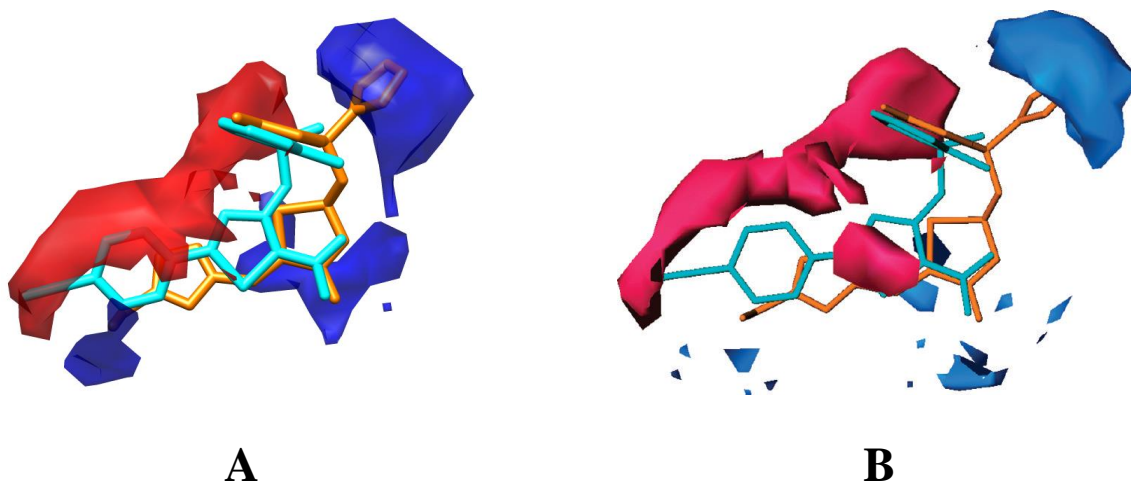


Figure 7. PLS-coefficients contour maps using the thumb-training set; only the highest active (**6** in cyan) and one of the lowest active (**11** in orange) compounds are shown. A: AutoGrid/R PLS coefficients contour maps derived from A probe analysis (Contour levels: 60%, positive red, negative blue). B: GRID/GOLPE PLS coefficients contour maps derived from C1= GRID probe analysis (contour levels: 0.0008 red, -0.0008 blue).

Table 8. PLS Analysis Results for the Thumb- and the Palm-Structure Based Autogrid/R and original GRID/GOLPE C1= 3-D QSAR Models.

Dataset	P	PC	r^2	q^2_{LOO}	q^2_{K5FCV}	r^2_{YS}	q^2_{YS}
Thumb	A	2	0.90	0.67	0.64	0.70	-0.63
Thumb	C	2	0.90	0.68	0.65	0.70	-0.60
Thumb	HD	2	0.92	0.75	0.73	0.68	-0.69
Thumb	NA	3	0.95	0.75	0.73	0.79	-0.66
Thumb	N	3	0.95	0.76	0.73	0.78	-0.67
Thumb	OA	3	0.95	0.77	0.73	0.77	-0.54
Thumb	e	3	0.98	0.58	0.52	0.92	-0.55
Thumb	d	1	0.58	0.36	0.36	0.27	-0.38
Thumb	GRID/GOLPE/C1=	3	0.99	-	0.69	-	-
Palm	A	3	0.96	0.73	0.62	0.68	-1.62
Palm	C	3	0.96	0.73	0.62	0.69	-1.59
Palm	HD	1	0.90	0.75	0.71	0.44	-0.76
Palm	NA	2	0.97	0.62	0.52	0.84	-0.76
Palm	N	2	0.97	0.62	0.55	0.85	-0.87
Palm	OA	1	0.86	0.67	0.64	0.32	-0.66
Palm	e	3	0.96	0.85	0.82	0.73	-1.01
Palm	d	3	0.93	0.62	0.39	0.73	-1.80
Palm	GRID/GOLPE/C1=	3	0.99	-	0.55	-	-

P:Autogrid Probe or GRID C1= probe; PC: optimal number of principal components/latent variables, r^2 : conventional square-correlation coefficient; q^2_{LOO} : cross-validation correlation coefficient using the leave-one-out method; q^2_{K5FCV} : cross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations; r^2_{YS} : average square correlation coefficient obtained after Y-scrambling process using 100 iterations; q^2_{YS} : average cross-validation correlation coefficient using the leave-one-out method obtained after Y-scrambling process using 100 iterations.

As for the original application,¹⁴ all the 3-DQSAR models were then externally tested, showing statistical values comparable with those obtained¹⁴ applying the GRID/GOLPE approach (Table 9).

Table 9. Thumb- and Palm-External Test Set prediction obtained from Structure Based AutoGrid/R and original GRID/GOLPE C1= 3-D QSAR Models.

Dataset	P	PC	SDEP_{ext}
Thumb	A	2	0.69
Thumb	C	2	0.69
Thumb	HD	2	0.76
Thumb	NA	3	0.66
Thumb	N	3	0.66
Thumb	OA	3	0.67
Thumb	e	3	0.63
Thumb	d	1	0.67
Thumb	GRID/GOLPE/C1=	3	0.59
Palm	A	3	1.14
Palm	C	3	1.11
Palm	HD	1	1.29
Palm	NA	2	1.04
Palm	N	2	1.04
Palm	OA	1	1.03
Palm	e	3	1.18
Palm	d	3	1.18
Palm	GRID/GOLPE/C1=	3	1.08

P:Autogrid Probe or GRID C1=probe; PC: optimal number of principal components/latent variables; SDEP_{ext}: standard deviation error of prediction for the external test set.

Multi-Probe Guided Region-Variable Selection

The Multi Probe Guided Region Selection methodology (MPGRS), represents a new powerful technique able to condense all the most informative interactions (from the mono probe models) in a single 3-D QSAR PLS model, leading to more comprehensive interpretations and allowing to derive useful three-dimensional pharmacophoric quantitative models (as explained in Chapters VI and VII). As reported,¹ each probe information is color coded according to Table 10. To test the procedure, the MPGRS was applied to both the case studies (LB and SB). All the multiprobe models (MP) were characterized by similar statistical coefficients as those obtained from the mono probe models (to avoid redundancy only the results from the SB case study are reported, Table 11) but enhancing the interpretation of the models (as in other studies like those characterizing Chapters VI and VII).

Table 10. List of the AutoGrid probes employed for MIF calculation and MPGRS Subregion color coding.

Probe	Description	MPGRS Colour
A	Aromatic Carbon	Gray
C	Aliphatic (sp ³) Carbon	Dark Gray
HD	Hydrogen bonded to heteroatom	Green
NA	Hydrogen-bond-accepting amine nitrogen	Cyan
N	Amide nitrogen	Blue
OA	Hydrogen-bond-accepting oxygen	Red
e	Electrostatic	Orange
d	Desolvation	Yellow

It was very interesting to overlap the information obtained from multiprobe models built for the SB case study, with the SB alignments, in order to compare the proposed pseudo receptor with experimental information. As for example, Figure 8 shows the pharmacophoric model (key interactions) derived from the multiprobe palm model. An high agreement between the selected regions and the HCV NS5B-palm binding pocket surface was observed,¹ demonstrating the usefulness of this application. More specific analyses can be developed using either these informative interactions with the associated PLS-coefficients (as applied during the studies characterizing Chapters VI and VII).

Table 11. Statistical Results Obtained from MPGRS Analysis for the Thumb- and the Palm-HCV Training Sets.

Dataset	MPGRS 3-D QSAR					
	$PC_{FL:SL}$	r^2	q^2_{K5FCV}	r^2_{YS}	q^2_{YS}	$SDEP_{ext}$
Thumb	2:2	0.95	0.90	0.50	-0.67	0.74
Palm	1:2	0.99	0.91	0.61	-0.93	1.06

$PC_{FL:SL}$: optimal number of principal first level (FL) and second level (SL) components/latent variables for the MPGRS model; r^2 : conventional square-correlation coefficient; q^2_{LOO} : cross-validation correlation coefficient using the leave-one-out method; q^2_{K5FCV} : cross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations; r^2_{YS} : average square correlation coefficient obtained after Y-scrambling process using 100 iterations; q^2_{YS} : average cross-validation correlation coefficient using the leave-one-out method obtained after Y-scrambling process using 100 iterations.

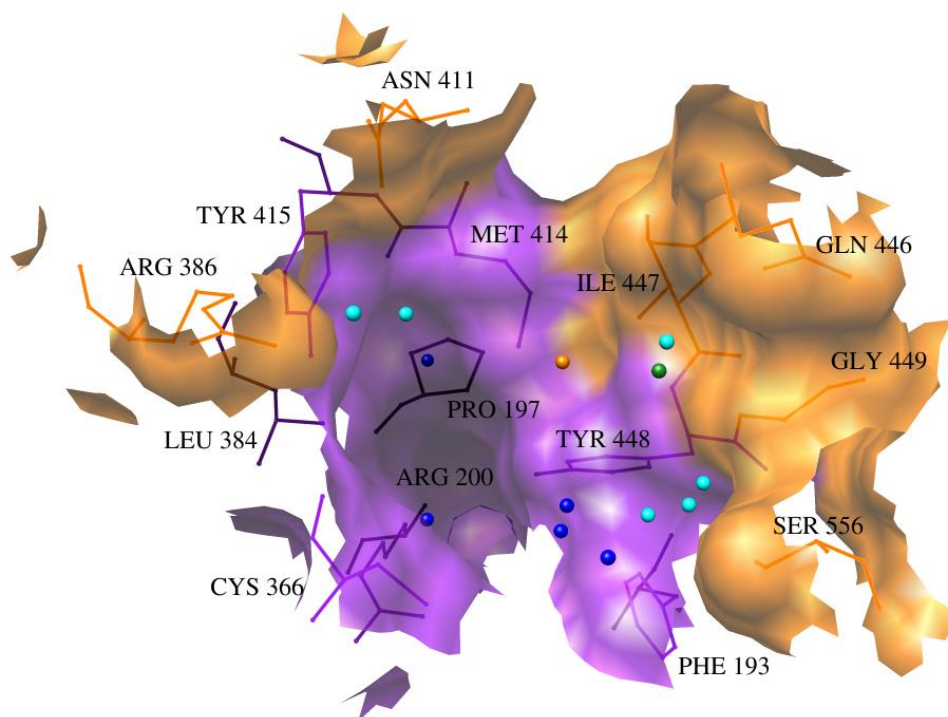


Figure 8. MPGRS 3-D QSAR palm model key points. The points are color coded: in blue N (amidic nitrogen) probe key points, in cyan those from NA (hydrogen acceptor nitrogen) probe, in green and orange those from HD (hydrogen donor) and e (electrostatic) probes respectively.

Bibliography

1. Ballante, F.; Ragno, R., 3-D QSAutogrid/R: an alternative procedure to build 3-D QSAR models. Methodologies and applications. *Journal of chemical information and modeling* **2012**, 52, 1674-85.
2. S. Wold, A. R., H. Wold, and W. J. Dunn, III, The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. and Stat. Comput.* **1984** 5, 735-743
3. Cramer, R. D.; Patterson, D. E.; Bunce, J. D., Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* **1988**, 110, 5959-67.
4. Goodford, P. J., A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* **1985**, 28, 849-57.
5. Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S., Generating Optimal Linear Pls Estimations (Golpe) - an Advanced Chemometric Tool for Handling 3d-Qsar Problems. *Quant Struct-Act Rel* **1993**, 12, 9-20.
6. Ballante, F.; Musmuca, I.; Marshall, G. R.; Ragno, R., Comprehensive model of wild-type and mutant HIV-1 reverse transcriptases. *Journal of computer-aided molecular design* **2012**, 26, 907-19.
7. Friggeri, L.; Ballante, F.; Ragno, R.; Musmuca, I.; De Vita, D.; Manetti, F.; Biava, M.; Scipione, L.; Di Santo, R.; Costi, R.; Feroci, M.; Tortorella, S., Pharmacophore assessment through 3-D QSAR: evaluation of the predictive ability on new derivatives by the application on a series of antitubercular agents. *Journal of chemical information and modeling* **2013**, 53, 1463-74.
8. Perspicace, E.; Jouan-Hureau, V.; Ragno, R.; Ballante, F.; Sartini, S.; La Motta, C.; Da Settimo, F.; Chen, B.; Kirsch, G.; Schneider, S.; Faivre, B.; Hesse, S., Design, synthesis and biological evaluation of new classes of thieno[3,2-d]pyrimidinone and thieno[1,2,3]triazine as inhibitor of vascular endothelial growth factor receptor-2 (VEGFR-2). *European journal of medicinal chemistry* **2013**, 63, 765-81.
9. Rotili, D.; Samuele, A.; Tarantino, D.; Ragno, R.; Musmuca, I.; Ballante, F.; Botta, G.; Morera, L.; Pierini, M.; Cirilli, R.; Nawrozkij, M. B.; Gonzalez, E.; Clotet, B.; Artico, M.; Este, J. A.; Maga, G.; Mai, A., 2-(Alkyl/aryl)amino-6-benzylpyrimidin-4(3H)-ones as inhibitors of wild-type and mutant HIV-1: enantioselectivity studies. *J Med Chem* **2012**, 55, 3558-62.
10. Silvestri, L.; Ballante, F.; Mai, A.; Marshall, G. R.; Ragno, R., Histone deacetylase inhibitors: structure-based modeling and isoform-selectivity prediction. *Journal of chemical information and modeling* **2012**, 52, 2215-35.

11. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **2009**, 30, 2785-91.
12. Team, R. D. C. The R Foundation for Statistical Computing. <http://www.r-project.org/>
13. Peng, Y.; Keenan, S. M.; Zhang, Q.; Kholodovych, V.; Welsh, W. J., 3D-QSAR comparative molecular field analysis on opioid receptor antagonists: pooling data from different studies. *J Med Chem* **2005**, 48, 1620-9.
14. Musmuca, I.; Caroli, A.; Mai, A.; Kaushik-Basu, N.; Arora, P.; Ragno, R., Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors. *Journal of chemical information and modeling* **2010**, 50, 662-76.
15. Clark, M.; Cramer, R. D., The Probability of Chance Correlation Using Partial Least-Squares (Pls). *Quant Struct-Act Rel* **1993**, 12, 137-145.
16. Wold S, E. J. a. M. C., *PLS : Partial Least Squares Projections to Latent Structures in 3D QSAR in Drug Design: Theory, Methods and Applications*. ESCOM Science Publishers: 1993.
17. Holland, J., *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. A Bradford Book: 1992.
18. Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P., Optimization by simulated annealing. *Science* **1983**, 220, 671-80.
19. George Box, S. H. a. W. G. H., *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. John Wiley & Sons: 1978; p 653.
20. Cho, S. J.; Tropsha, A., Cross-validated R²-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results. *J Med Chem* **1995**, 38, 1060-6.
21. Pastor, M.; Cruciani, G.; Clementi, S., Smart region definition: a new way to improve the predictive ability and interpretability of three-dimensional quantitative structure-activity relationships. *J Med Chem* **1997**, 40, 1455-64.
22. S. Wold, L. E., *Chemometrics Methods in Molecular Design*. VCH, Weinheim: 1995.

Chapter II

Comprehensive model of wild-type and mutant HIV-1 reverse transcriptases

Flavio Ballante, Ira Musmuca, Garland R. Marshall, Rino Ragno

Journal of Computer-Aided Molecular Design. 2012 Aug;26(8):907-19.

Preamble

This study is characterized by the application of the COMBINER procedure,¹ developed by us, to obtain a comprehensive 3-D QSAR model of 7 different HIV-1 reverse transcriptase enzymes (RT, wild-type and drug-resistant mutants) complexed with niverapine (NVP) and efavirenz (EFV), able to define those mutations responsible for the different activity profiles. The model predictive ability was assessed using an external test set of novel chiral 2-(alkyl/aryl)amino-6-benzylpyrimidin-4(3H)-ones (DABOs).² The COMBINER model was able to correctly predict either the experimental activities and the right eudismic ratio of the test set derivatives, rationalizing the experimentally observed inhibitory activity and confirming to be a useful tool in drug discovery.

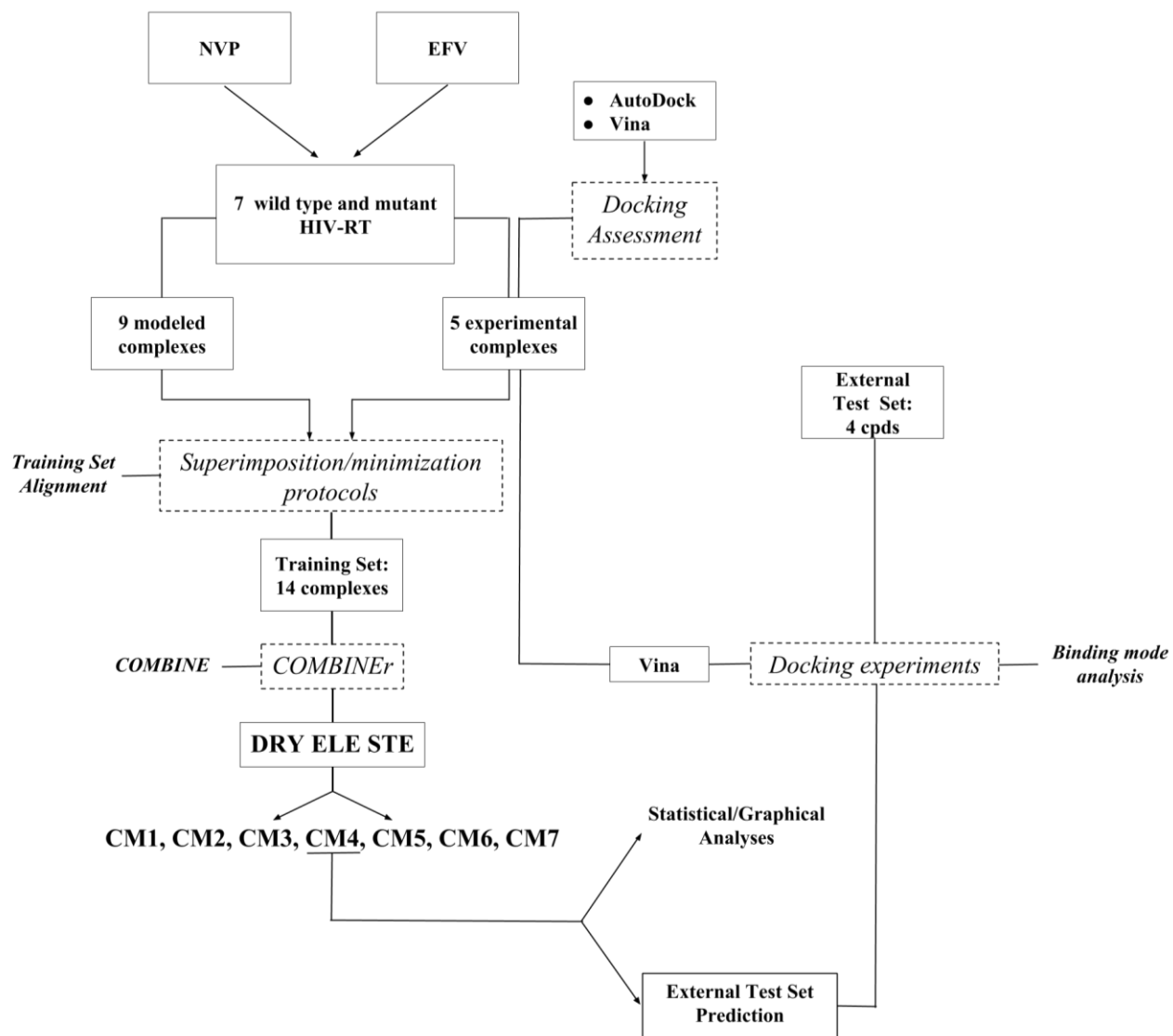


Figure 1. Computational procedure.

Introduction

The COMBINER procedure,^{1, 3} an enhanced version of the original COMBINE methodology,⁴⁻⁶ is a structure based (SB) 3-D QSAR application that uses ligand-receptor complexes to quantify, by molecular mechanics, their interaction energies.⁷ The novelty of COMBINER is to use the AutoGrid program (AutoDockTools)⁸ to calculate ligand-residues interaction energies for each ligand-enzyme complex and derive, through the PLS^{9, 10} multivariate analysis, descriptive and predictive models through the R environment.¹¹ Since its capability to directly compute the ligand/enzyme per residue interaction it represents a powerful tool to highlight the most important ligand-receptor interactions explaining the effect of single points mutation, on the ligands' inhibitory activity, as in the case of the HIV-1 reverse transcriptase (RT) mutant enzymes that characterize the application characterizing this chapter.

Applied Methodology

The choice of the training set was based considering both the availability of homogeneous inhibition data² for Nevirapine (NVP) and Efavirenz (EFV) (Figure A1), and co-crystal structures. This approach, led to the selection of 7 different HIV-RT wild-type and mutant enzymes complexed with the two inhibitors (Appendix A Table A1). Therefore, a total of 14 complexes composed the training set: since co-crystal structures were available for only five complexes, the other nine were modeled using as template structural information available from other complexes present in the PDB.³ As reported,³ a superimposition/minimization protocol was then applied to get these complexes ready for the COMBINER¹ investigation. By means of Autogrid4,⁸ three kinds of ligand/residue interactions: the electrostatic (ELE), the steric (STE) and the desolvation (DRY) were derived

for each complex to compose seven different fields combinations: the monoprobe fields (DRY, ELE and STE) and the multiprobe fields (DRY+STE, ELE+STE, DRY+ELE, and DRY+ELE+STE), and derive, finally, seven different COMBINER PLS¹⁰ models (CM1-CM7, Table 1). As shown in Table 1, all these models were characterized by good statistical coefficients, but only two, CM1 and CM4 (showing the best statistical-values profiles), were selected for further investigations. PLS coefficients, weighted PLS coefficients (PLS coefficients x standard deviation values) and activity contribution plots (PLS coefficients x interaction energies), were of fundamental importance to detect which residues are relevant for differences in activity and quantify their relative importance. By analyzing these plots, has appeared how a similar profile was characterizing the DRY field in both CM1 and CM4 models,³ therefore the attention was focused on the latter (DRY+STE fields) in order to consider more data. Analyses of PLS coefficients allowed to identify the residues mainly involved in the model definition³ (Figure 2) while weighted PLS coefficients (Figure 3) were useful to derive the global importance of each interaction. Results of these investigations suggested that interactions with residues Leu100(Ile100), Lys101 and Tyr188 (Leu188) are desirable, while should be avoided with residues Trp229 and Leu224 (Figures 2 and 3).

Table 1. Statistical coefficients of the COMBINE models. CM: COMBINE Model Number; r^2 : conventional squared correlation coefficient; SDEC: standard error of calculation; q^2 : cross-validation coefficient; LOO: leave one out; SDEP: standard error of prediction; LSO5 and LSO2: leave some out using 5 and 2 groups respectively.

CM	Model	r^2	SDEC	q^2_{LOO}	SDEP_{LOO}	q^2_{LSO5}	$\text{SDEP}_{\text{LSO5}}$	q^2_{LSO2}	$\text{SDEP}_{\text{LSO2}}$
1	DRY	0.91	0.31	0.82	0.43	0.79	0.46	0.63	0.58
2	ELE	0.80	0.45	0.51	0.71	0.49	0.72	0.37	0.79
3	STE	0.81	0.44	0.69	0.57	0.65	0.60	0.52	0.68
4	DRY_STE	0.88	0.35	0.78	0.48	0.75	0.50	0.61	0.61
5	ELE_STE	0.82	0.43	0.58	0.66	0.53	0.69	0.44	0.75
6	DRY_ELE	0.89	0.34	0.66	0.59	0.63	0.62	0.48	0.70
7	DRY_ELE_STE	0.86	0.38	0.66	0.59	0.62	0.62	0.50	0.70

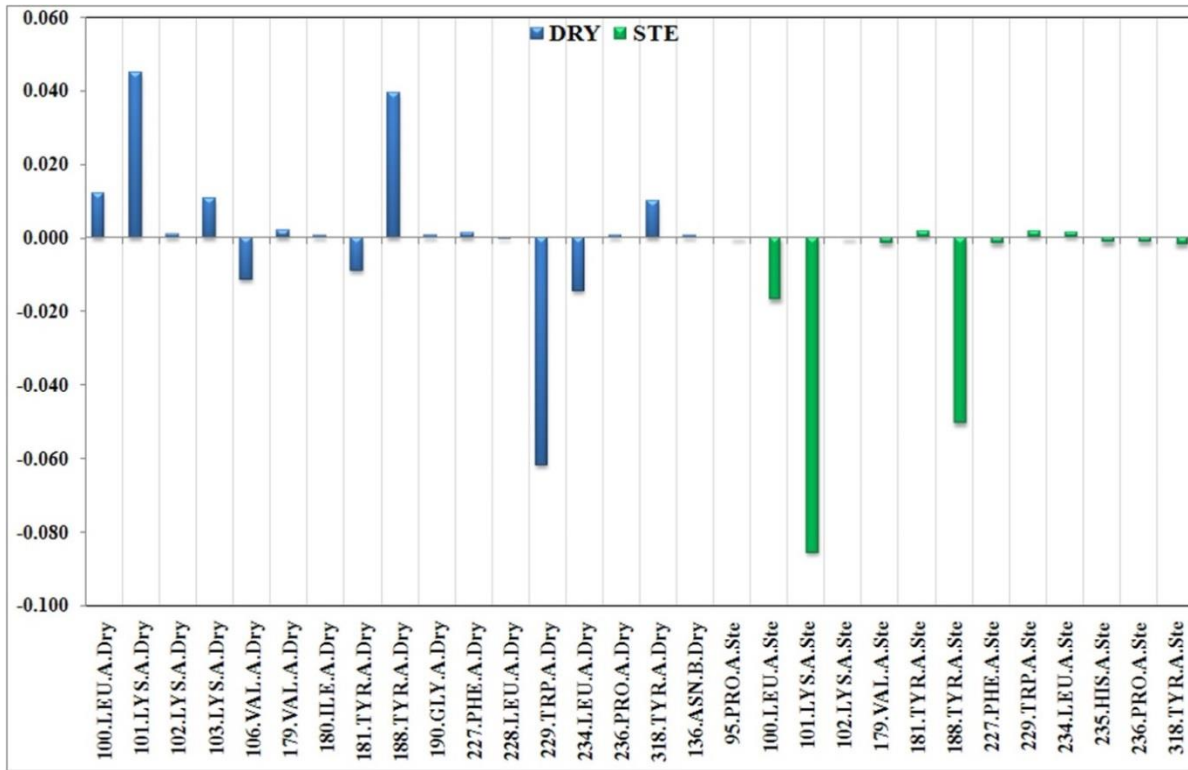


Figure 2. PLS coefficients obtained from the CM4 model. Only bars with values higher than 0.001 and lower than -0.001 are shown.

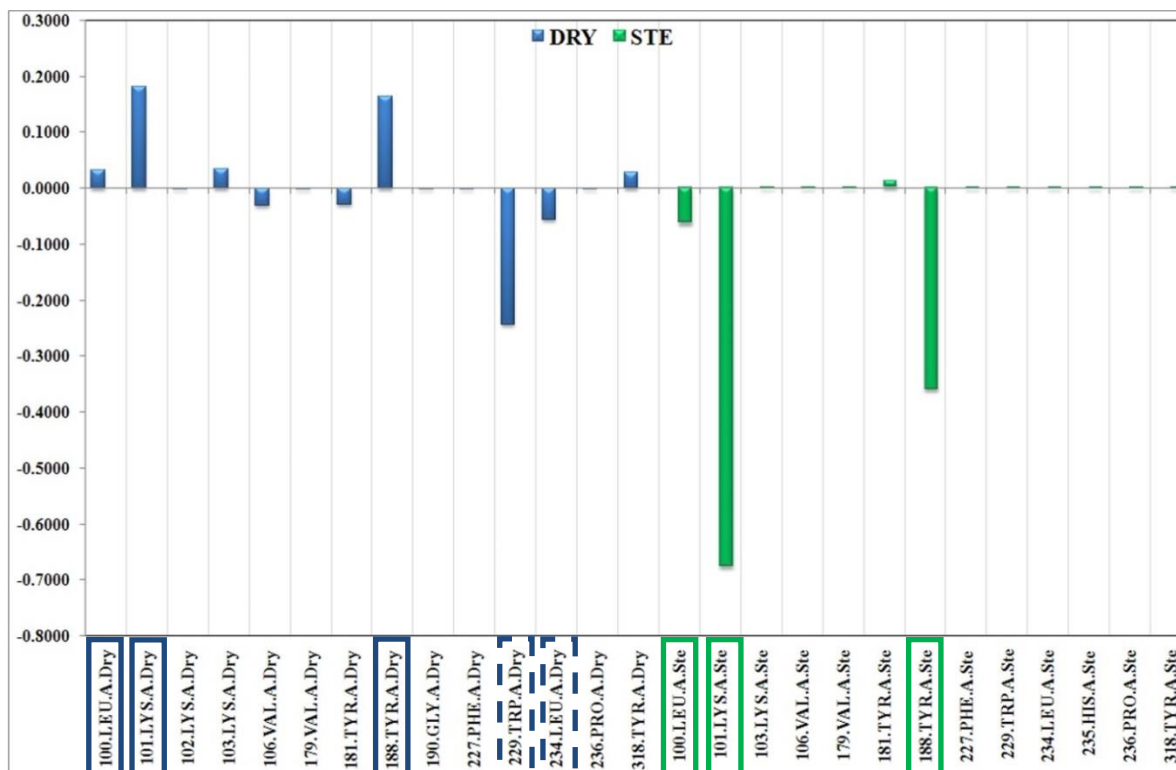


Figure 3. PLS coefficients x SD obtained from the CM4 model. Only bars with values higher than 0.001 and lower than -0.001 are shown.

The CM4 model was finally applied to an external test set composed of four DABO derivatives² (Appendix A Figure A2) with the purpose to rationalize the role of the mutations on their activity profile. To build the non-experimental complexes, with those isoforms used to test these compounds,² a reported¹² cross-docking protocol was applied by means of the AutoDockVina.¹³ This docking program was chosen after a docking assessment investigation in which either AutodockVina¹³ and AutoDock⁸ were tested for their reliability.³ Binding mode analyses suggested similar poses for R-conformations of MC-1501 and MC2082, in agreement with previous studies,¹⁴⁻¹⁶ while for (*S*) derivatives, results confirmed the role of the methyl group at the C6-benzylic position to prevent similar interactions.¹⁶ The COMBINER¹ model was able to predict with an acceptable average absolute error of prediction the test set experimental activities (Table 2), tracing the right eudismic ratio for the two *R/S* pairs. As extensively reported in the

paper,³ straightforward was the possibility to investigate the activity contributions (Figure 4) of each inhibitor/residue couple, allowing to:

- a) confirm that the residues mutations were responsible for the different activity profiles of EFV and NVP
- b) derive a final three-dimensional scheme of contributions for each inhibitor/residue pair (Figure 5)
- c) identify what interactions determine the different activity profiles:
 - interaction with Lys101 was identified, from both DRY and STE fields analysis, to be the mainly responsible for the higher activity of (*R*)-MC2082 respect (*R*)-MC1501;
 - DRY and STE interactions with Trp229 and Lys101 respectively, contribute for the higher activity of (*R*)-MC1501 versus (*S*)- MC1501
 - Lys101 and Trp229 DRY interactions, and mostly Lys101 STE interaction, determine the higher activity of (*R*)-MC2082 respect (*S*)-MC2082;
 - residue 188 is capable to maintain interactions with ligands in its wild-type (Tyr188) and in the Leu188 mutation, and to compensate for loss of activities due to other single-point mutations.

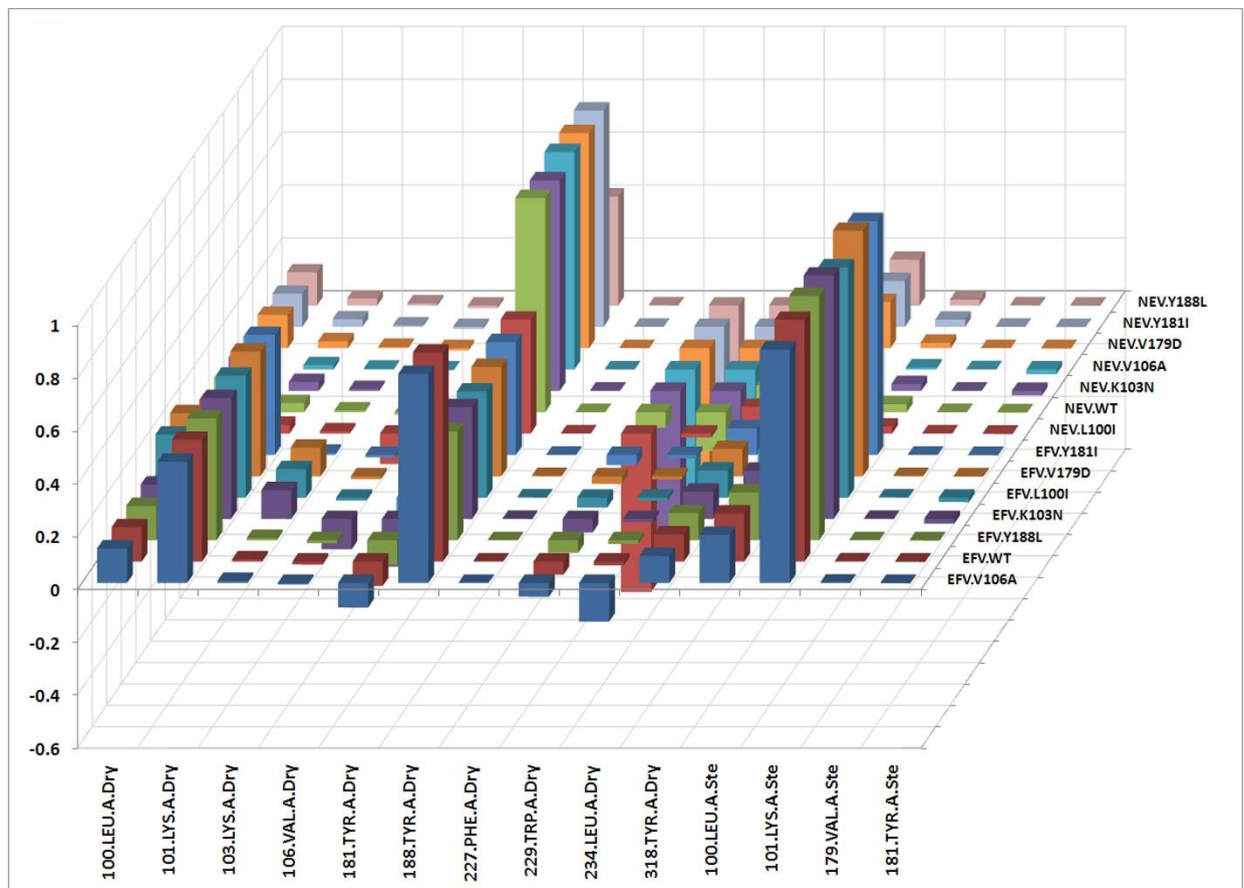


Figure 4. Activity contribution histograms obtained from the CM4 model. Only bars with values higher than 0.001 and lower than -0.001 are shown.

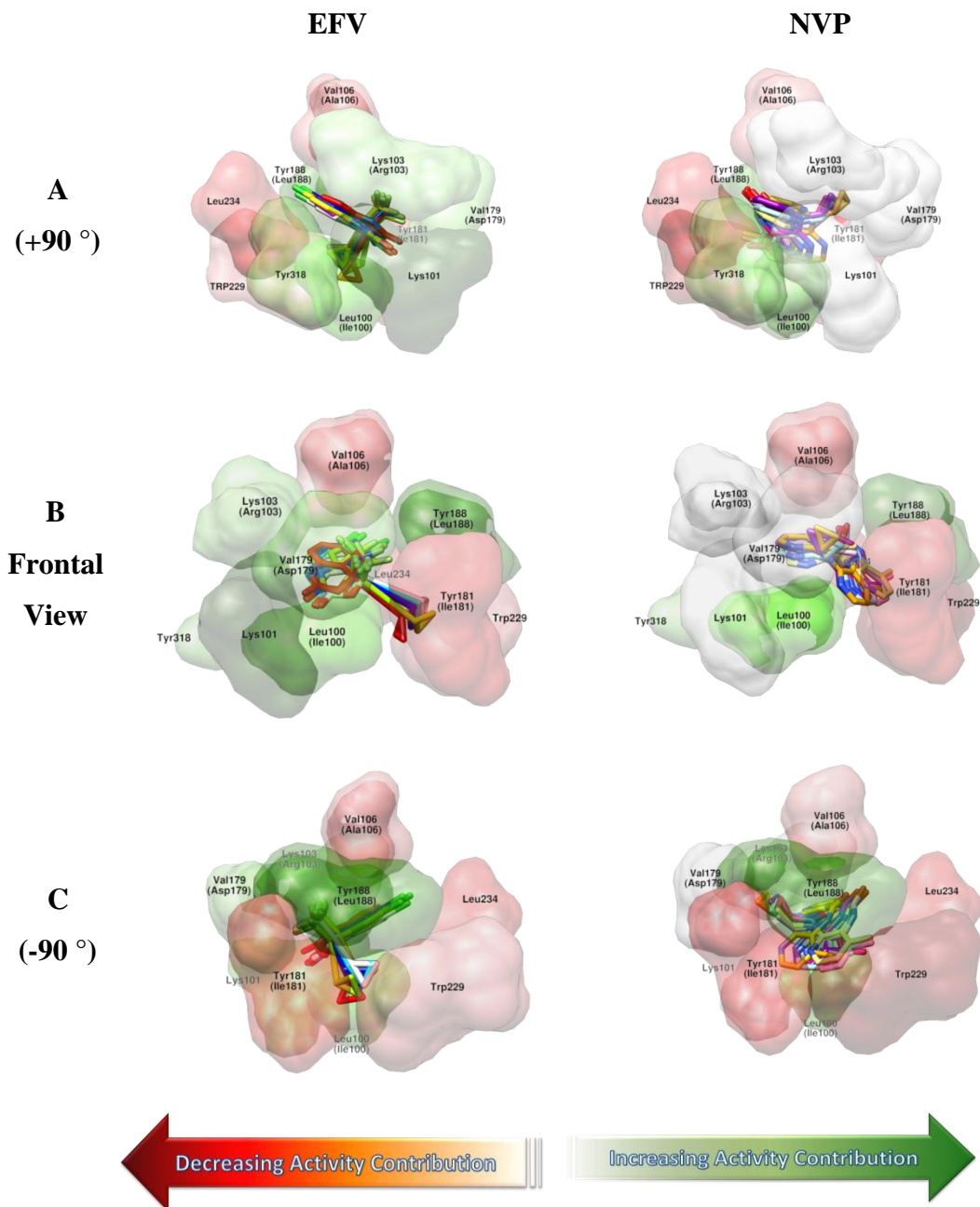


Figure 5. Efavirenz (EFV) and Nevirapine (NVP) with the surrounding residues surfaces as in the experimental complexes. The surfaces are colored by activity contribution. Three view of the complexes which are rotated along the X axes by $\pm 90^\circ$ are shown.

Table 2. Experimental and COMBINE model CM4 predicted activities of MC compounds.

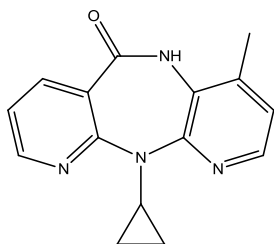
	MC1501				MC2082			
	<i>R</i>		<i>S</i>		<i>R</i>		<i>S</i>	
	Exp	Pred	Exp	Pred	Exp	Pred	Exp	Pred
WT	8.70	7.46	6.93	7.20	6.81	7.21	4.52	5.77
V106A	8.52	9.19	6.45	5.78	9.52	9.43	6.62	7.51
K103N	7.02	7.17	6.01	7.52	8.52	9.11	7.19	7.52
L100I	7.02	6.69	4.40	7.11	8.10	7.49	6.74	6.03
Y188L	6.71	7.51	4.40	5.11	8.10	7.09	4.40	5.95
Y181I	6.35	6.05	4.40	6.12	6.12	6.25	6.29	5.48

APPENDIX A

Table A1. Anti-RT activities (μM) of NVP and EFV used to build the COMBINER model.

RT	NVP	EFV
WT	0.4	0.03
L100I	9.0	0.12
K103N	7.0	0.16
V106A	10.0	0.04
V179D	2.0	0.10
Y181I	36.0	0.15
Y188L	18.0	0.38

Nevirapine (NVP)



Efavirenz (EFV)

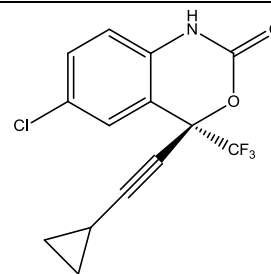
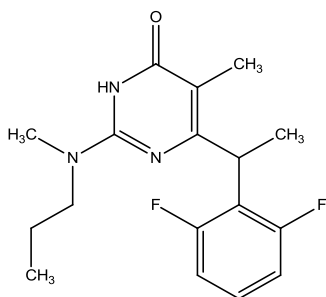


Figure A1. Nevirapine and Efavirenz.

(*R/S*) **MC1501**
(*N,N*-DABO)



(*R/S*) **MC2082**
(DAPY-DABO Hybrid)

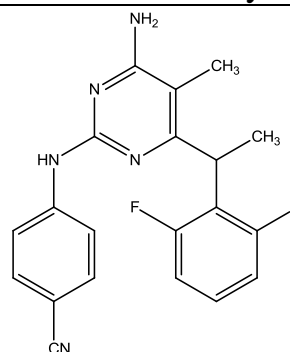


Figure A2. Racemic DABO derivatives used as external test set.

Bibliography

1. Silvestri, L.; Ballante, F.; Mai, A.; Marshall, G. R.; Ragno, R., Histone deacetylase inhibitors: structure-based modeling and isoform-selectivity prediction. *Journal of chemical information and modeling* **2012**, 52, 2215-35.
2. Rotili, D.; Samuele, A.; Tarantino, D.; Ragno, R.; Musmuca, I.; Ballante, F.; Botta, G.; Morera, L.; Pierini, M.; Cirilli, R.; Nawrozkij, M. B.; Gonzalez, E.; Clotet, B.; Artico, M.; Este, J. A.; Maga, G.; Mai, A., 2-(Alkyl/aryl)amino-6-benzylpyrimidin-4(3H)-ones as inhibitors of wild-type and mutant HIV-1: enantioselectivity studies. *J Med Chem* **2012**, 55, 3558-62.
3. Ballante, F.; Musmuca, I.; Marshall, G. R.; Ragno, R., Comprehensive model of wild-type and mutant HIV-1 reverse transcriptases. *Journal of computer-aided molecular design* **2012**, 26, 907-19.
4. Lozano, J. J.; Pastor, M.; Cruciani, G.; Gaedt, K.; Centeno, N. B.; Gago, F.; Sanz, F., 3D-QSAR methods on the basis of ligand-receptor complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 ligands. *Journal of computer-aided molecular design* **2000**, 14, 341-53.
5. Perez, C.; Pastor, M.; Ortiz, A. R.; Gago, F., Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design. *J Med Chem* **1998**, 41, 836-52.
6. Rodriguez-Barrios, F.; Gago, F., Chemometrical identification of mutations in HIV-1 reverse transcriptase conferring resistance or enhanced sensitivity to arylsulfonylbenzotrioles. *J Am Chem Soc* **2004**, 126, 2718-9.
7. Ortiz, A. R.; Pastor, M.; Palomer, A.; Cruciani, G.; Gago, F.; Wade, R. C., Reliability of comparative molecular field analysis models: effects of data scaling and variable selection using a set of human synovial fluid phospholipase A2 inhibitors. *J Med Chem* **1997**, 40, 1136-48.
8. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **2009**, 30, 2785-91.
9. S. Wold, A. R., H. Wold, and W. J. Dunn, III, The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. and Stat. Comput.* **1984** 5, 735-743
10. Wold S, E. J. a. M. C., *PLS : Partial Least Squares Projections to Latent Structures in 3D QSAR in Drug Design: Theory, Methods and Applications*. ESCOM Science Publishers: 1993.
11. Team, R. D. C. The R Foundation for Statistical Computing. <http://www.r-project.org/>
12. Musmuca, I.; Caroli, A.; Mai, A.; Kaushik-Basu, N.; Arora, P.; Ragno, R., Combining 3-D quantitative structure-activity relationship with ligand based and

structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors. *Journal of chemical information and modeling* **2010**, 50, 662-76.

13. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **2010**, 31, 455-61.

14. Mai, A.; Sbardella, G.; Artico, M.; Ragno, R.; Massa, S.; Novellino, E.; Greco, G.; Lavecchia, A.; Musiu, C.; La Colla, M.; Murgioni, C.; La Colla, P.; Loddo, R., Structure-based design, synthesis, and biological evaluation of conformationally restricted novel 2-alkylthio-6-[1-(2,6-difluorophenyl)alkyl]-3,4-dihydro-5-alkylpyrimidin-4(3H)-ones as non-nucleoside inhibitors of HIV-1 reverse transcriptase. *J Med Chem* **2001**, 44, 2544-54.

15. Quaglia, M.; Mai, A.; Sbardella, G.; Artico, M.; Ragno, R.; Massa, S.; del Piano, D.; Setzu, G.; Doratiotto, S.; Cotichini, V., Chiral resolution and molecular modeling investigation of rac-2-cyclopentylthio-6-[1-(2,6-difluorophenyl)ethyl]-3,4-dihydro-5-methylpyrimidin-4(3H)-one (MC-1047), a potent anti-HIV-1 reverse transcriptase agent of the DABO class. *Chirality* **2001**, 13, 75-80.

16. Ragno, R.; Mai, A.; Sbardella, G.; Artico, M.; Massa, S.; Musiu, C.; Mura, M.; Marturana, F.; Cadeddu, A.; La Colla, P., Computer-aided design, synthesis, and anti-HIV-1 activity in vitro of 2-alkylamino-6-[1-(2,6-difluorophenyl)alkyl]-3,4-dihydro-5-alkylpyrimidin-4(3H)-ones as novel potent non-nucleoside reverse transcriptase inhibitors, also active against the Y181C variant. *J Med Chem* **2004**, 47, 928-34.

Chapter III

2-(Alkyl/Aryl)amino-6-benzylpyrimidin-4(3 H)-ones as inhibitors of wild-type and mutant HIV-1: Enantioselectivity studies

Dante Rotili, Alberta Samuele, Domenico Tarantino, Rino Ragno, Ira Musmuca, Flavio Ballante, Giorgia Botta, Ludovica Morera, Marco Pierini, Roberto Cirilli, Maxim B. Nawrozkij, Emmanuel Gonzalez, Bonaventura Clotet, Marino Artico, José A. Esté, Giovanni Maga, and Antonello Mai

Journal of Medicinal Chemistry 2012 55 (7), 3558-3562

Preamble

This study is strictly related with that characterizing Chapter II. In fact the COMBINER CM4 model, previously developed, was used to predict the biological activity of two potent anti-HIV-1 derivatives: MC1501 and MC2082, which are the object of the present study, to quantify the influence of single-point HIV-RT mutations on their activity. As discussed in the previous Chapter, the COMBINER methodology confirmed to be a useful tool to explain the role of single point mutations and predict the different activity profiles.

Introduction

F₂-N,N-DABO (MC1501)¹ and DABO-DAPY (MC2082)² are two pyrimidine-based non-nucleoside reverse transcriptase (RT) inhibitors (NNRTI) endowed with high anti-HIV-1 activity and characterized by a stereogenic center at the C-6 benzylic position (Figure 1). In this study, these two highly potent anti-HIV-1 agents were systematically investigated for their enantioselective anti-HIV-1 activity. The availability of such homogeneous biological results together with the protein X-ray information allowed to quantify the influence of single point HIV-1-RT mutations by means of the previously developed COMBINER CM4 model.³

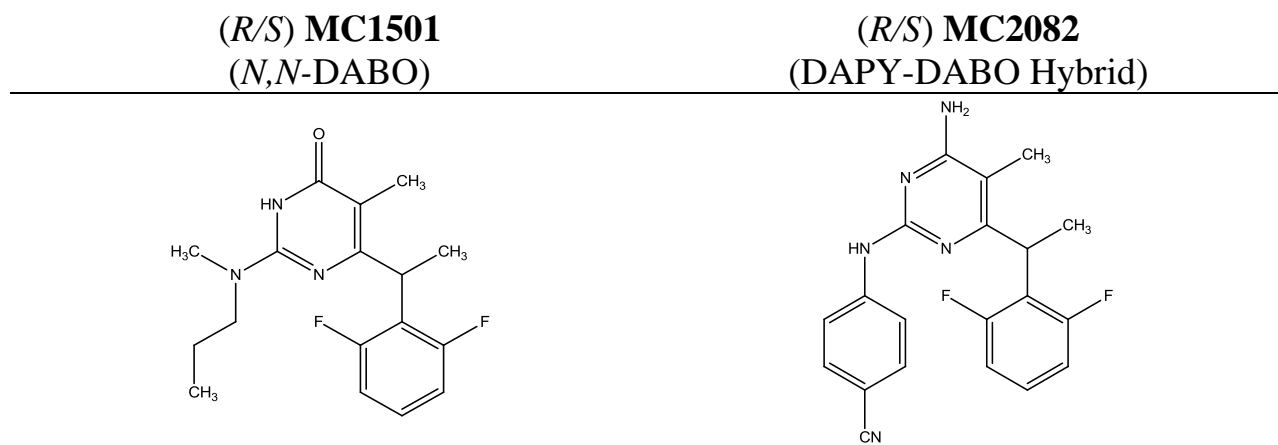


Figure 1. Racemic *N,N*-DABO and DAPY-DABO Hybrid.

Results and Discussion

The single enantiomers and the corresponding racemic mixtures of MC1501 and MC2082 were tested to evaluate their anti-HIV activity against WT HIV-1 and clinically relevant HIV-1 mutant strains (K103N, L100, Y181I, V106A and Y188L); nevirapine (NVP), efavirenz (EFV) and dapivirine (TMC210) were tested as reference drugs. Biological results highlighted that, all the *R* forms were the more active, followed by the racemic mixtures and lastly by the *S* forms.

Moreover (*R*)-MC2082 resulted to be generally (except for Y181C) more active than (*R*)-MC1501. Biological results allowed to define, for the tested compounds, the different activity profiles against the different HIV-1 isoforms. Due to the availability of protein structures, both MC1501 and MC2082 binding conformations were derived applying a reported⁴ cross-docking protocol and externally evaluated by the COMBINER CM4 model (see Chapter II for further specifications), in order to better understand the role of the different enzyme mutations. As reported in Chapter II, the COMBINER CM4 model was able to predict the higher activity of (*R*) enantiomers respect the respective (*S*) form as well as the higher potency of MC2082 respect MC1501, in agreement with the experimental results. Moreover was possible to highlight the role of Lys101, Trp229 and Tyr188 as described in Chapter II.

Bibliography

1. Mai, A.; Artico, M.; Rotili, D.; Tarantino, D.; Clotet-Codina, I.; Armand-Ugon, M.; Ragno, R.; Simeoni, S.; Sbardella, G.; Nawrozkij, M. B.; Samuele, A.; Maga, G.; Este, J. A., Synthesis and biological properties of novel 2-aminopyrimidin-4(3H)-ones highly potent against HIV-1 mutant strains. *J Med Chem* **2007**, *50*, 5412-24.
2. Rotili, D.; Tarantino, D.; Artico, M.; Nawrozkij, M. B.; Gonzalez-Ortega, E.; Clotet, B.; Samuele, A.; Este, J. A.; Maga, G.; Mai, A., Diarylpyrimidine-dihydrobenzylloxypyrimidine hybrids: new, wide-spectrum anti-HIV-1 agents active at (sub)-nanomolar level. *J Med Chem* **2011**, *54*, 3091-6.
3. Ballante, F.; Musmuca, I.; Marshall, G. R.; Ragno, R., Comprehensive model of wild-type and mutant HIV-1 reverse transcriptases. *Journal of computer-aided molecular design* **2012**, *26*, 907-19.
4. Musmuca, I.; Caroli, A.; Mai, A.; Kaushik-Basu, N.; Arora, P.; Ragno, R., Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors. *Journal of chemical information and modeling* **2010**, *50*, 662-76.

Chapter IV

Histone deacetylase inhibitors: Structure-based modeling and isoform-selectivity prediction

Laura Silvestri, Flavio Ballante, Antonello Mai, Garland R. Marshall, and Rino Ragno

Journal of Chemical Information and Modeling 2012 52 (8), 2215-2235

Preamble

In the last decade, considerable interest has developed towards those phenomena, called epigenetics,¹ capable of influencing gene expression, without changing the structure of DNA. Actually, epigenetics plays a role of primary importance in scientific research, due to the fact that different diseases are related with its dysregulation, which may be potentially rebalanced or prevented by the targeted use of chemical agents^{2, 3} The zinc-dependent mammalian histone deacetylases (HDACs) are a family of proteins comprising 11 enzymes (isoforms), which are fundamental for tissue's development and homeostasis. Considering the link between misregulated HDAC activity, carcinogenesis and other human diseases, the design of selective HDAC inhibitors for therapeutic treatment is mandatory, to develop less toxic drugs clarifying the biological role of these enzymes. This study was characterized by the application of the COMBINER^{4, 5} procedure on a series of class I/II HDACs complexes. During the investigation both ligand based (LB) and structure based (SB) alignment were used, considering all the eleven human zinc-based enzyme isoforms, to define the most influent chemical interactions for activity and isoform selectivity, allowing to rationalize the design of novel compounds endowed with both high potency and selectivity. Moreover, the

COMBINER models represent a useful tool to predict the bioactive profile of a given molecule towards the 11 HDAC isoforms, helping in the selection and development of new selective inhibitors. Due to the complexity, the comprehensiveness and the amount of used data characterizing this work, it is impossible to summarize all the aspects and results in a way different from that published by us. Therefore, the reader may refer to the next paragraph (Introduction on HDACs) and then directly to the scientific publication (Chapter XII).

Introduction on HDACs

A fundamental epigenetic mechanism is represented by chromatin remodeling, which appears to be controlled by the acetylation state of histone proteins (Figure 1).⁶

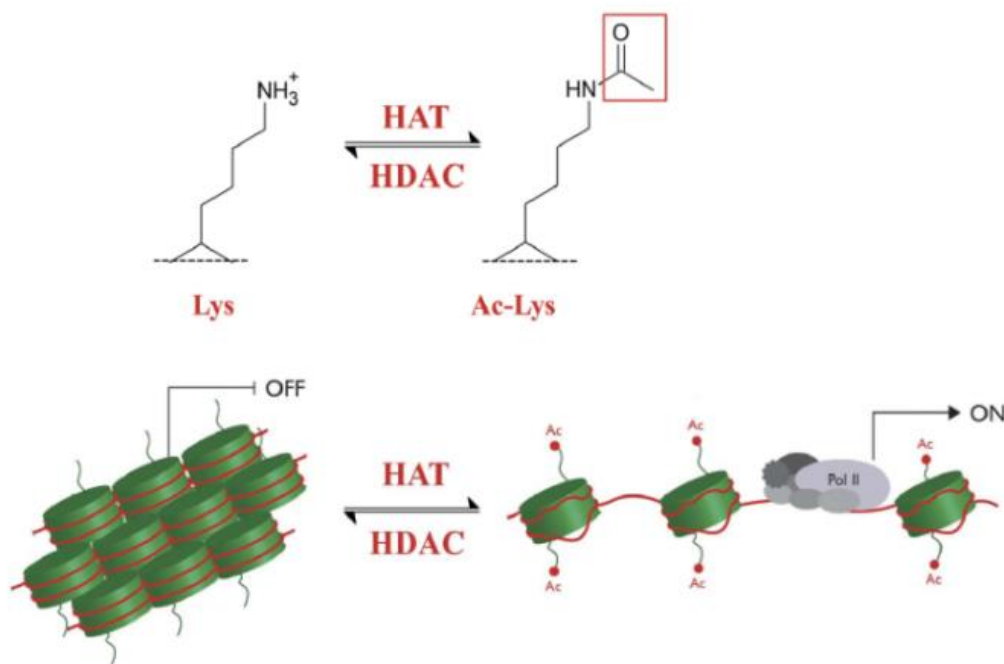


Figure 1. Histone proteins acetylation/deacetylation state is guaranteed by the activity of HATs and HDACs enzymes.

Two families of enzymes: histone deacetylase (HDAC) and histone acetyl transferase (HAT) are responsible to maintain the homeostatic acetylation state of histone proteins. In particular, histone deacetylases (HDACs) are a class of zinc metalloproteases that catalyze the deacetylation of acetylated histones by removing the acetyl moiety from the ϵ -amino groups of lysine residues in the N-terminal extensions of nucleosomal histones, causing transcriptional repression; on the contrary histone acetyl transferases (HATs) catalyze the acetylation of histone tails determining transcriptional activation. Over the last years, it has been found that overexpression of histone deacetylases (HDACs) plays an important role in carcinogenesis and other diseases (i.e. development of HIV latency and Malaria)⁷⁻¹⁰. Thus, HDAC inhibitors have emerged as new therapeutic agents for multiple human diseases. To date, there are 18 known isoforms, distinguished in 4 classes according to homology with yeast deacetylases, cellular localization and enzymatic activity.¹¹ The eleven human zinc-dependent isoforms (Figure 2 shows the catalytic mechanism) of class I (HDAC-1, -2, -3, and -8), II (which includes class IIa: HDAC-4, -5, -7, and -9, and class IIb: HDAC-6 and -10) and IV (HDAC- 11) are mainly considered since their involvement in different diseases like: cancer, viral and parasitic infections, neurodegenerative diseases and inflammation. Therefore, is of considerable interest to obtain and optimize selective inhibitors to: 1) further clarify the biological effect of the different isoforms, 2) achieve a targeted therapeutic effect, 3) reduce adverse reactions characterizing non-selective therapies. Currently, HDAC inhibitors (HDACI) are divided according to their chemical structure, as follows: short chain fatty acids, hydroxamic acids, benzamides, ketones and cyclic peptides. Currently all the HDACI share a common pharmacophore, characterized by three fundamental structural elements: 1) a zinc binding group, 2) a capping group and 3) a linker domain group. This pharmacophore can be represented using as reference the structure of trichostatin

A¹² (TSA, Figure 3), among the first identified HDACIs. The first HDAC inhibitors approved by FDA are: Merck's Zolinza, vorinostat (SAHA) and Celgene's Istodax romidepsin (Figure 4), currently used for the treatment of cutaneous T-cell lymphoma. However, other derivatives are in clinical trials for the ability to block cell proliferation, promote differentiation and induce apoptosis.¹³

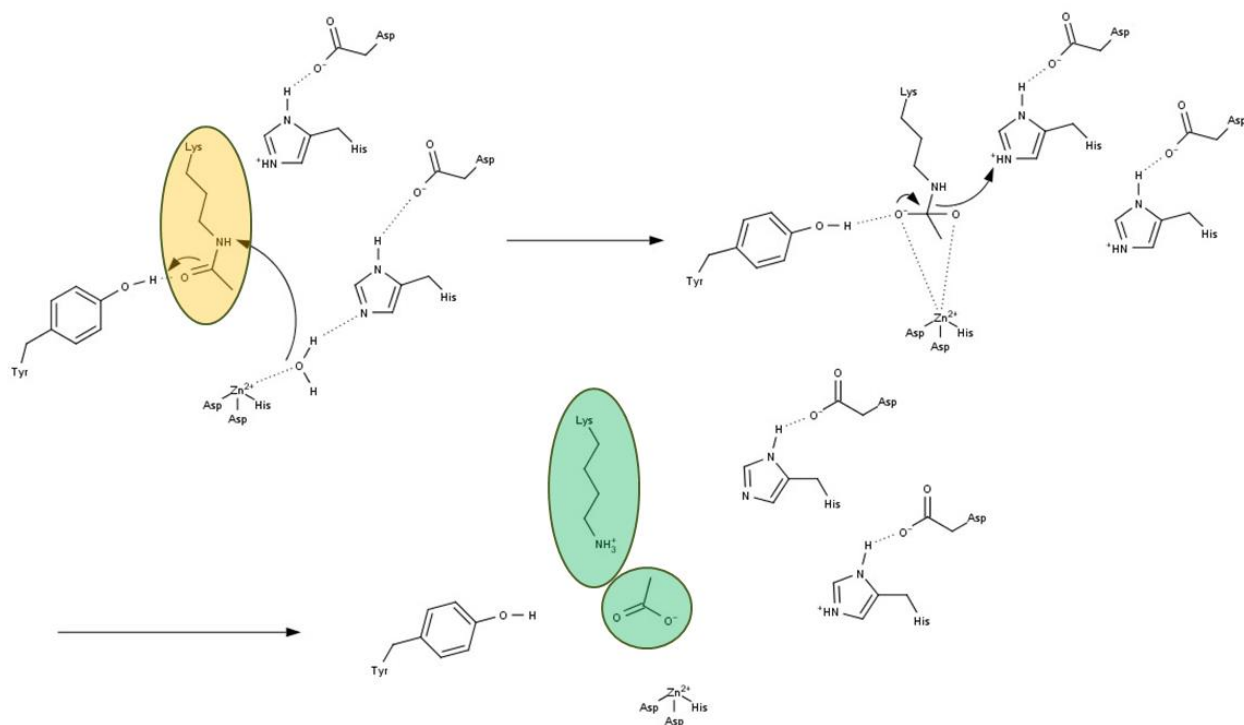


Figure 2. HDAC zinc-dependent catalytic mechanism. Highlighted in yellow and green are the acetylated substrate and the products.

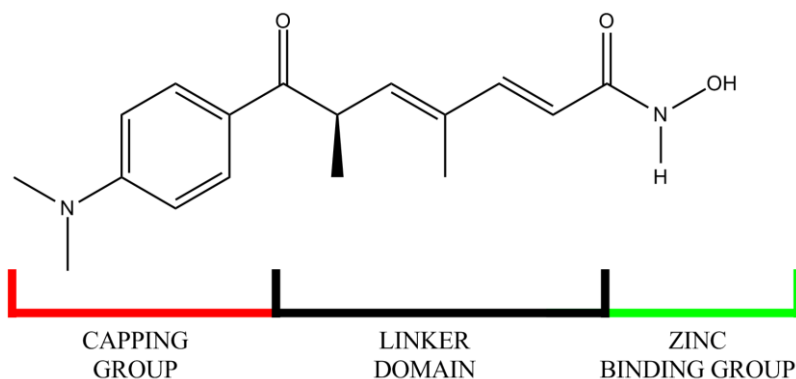
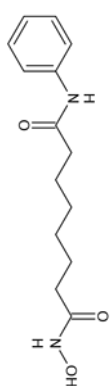
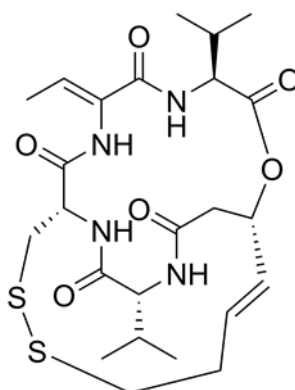


Figure 3. Common HDACi pharmacophore, TSA is depicted.

To date, unfortunately, most of the derived HDACi are not capable to inhibit specifically the different HDAC isoforms. Therefore, a main challenge is to design new selective compounds with the aim to elucidate the functions of each isoform and apply new efficient and less toxic therapeutic treatments.



SAHA



Romidepsin

Figure 4. Structure of SAHA and Romidepsin.

The purpose of this study, was to build and validate a 3-D QSAR Structure-Based (SB) model, using the COMBINER^{4,5} method (an enhanced version of the original comparative binding energy analysis, COMBINE)¹⁴⁻¹⁷ over the 11 HDAC isoforms able to detect the most discriminant chemical interactions among the different enzyme isoforms, and predict the activity/selectivity of new derivatives (see the relative article in Chapter XII).

Bibliography

1. Egger, G.; Liang, G.; Aparicio, A.; Jones, P. A., Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **2004**, 429, 457-63.
2. Lengauer, C.; Issa, J. P., The role of epigenetics in cancer. DNA Methylation, Imprinting and the Epigenetics of Cancer--an American Association for Cancer Research Special Conference. Las Croabas, Puerto Rico, 12-16 1997 December. *Molecular medicine today* **1998**, 4, 102-3.
3. Jones, P. A.; Laird, P. W., Cancer epigenetics comes of age. *Nature genetics* **1999**, 21, 163-7.
4. Ballante, F.; Musmuca, I.; Marshall, G. R.; Ragno, R., Comprehensive model of wild-type and mutant HIV-1 reverse transcriptases. *Journal of computer-aided molecular design* **2012**, 26, 907-19.
5. Silvestri, L.; Ballante, F.; Mai, A.; Marshall, G. R.; Ragno, R., Histone deacetylase inhibitors: structure-based modeling and isoform-selectivity prediction. *Journal of chemical information and modeling* **2012**, 52, 2215-35.
6. Grayson, D. R.; Kundakovic, M.; Sharma, R. P., Is there a future for histone deacetylase inhibitors in the pharmacotherapy of psychiatric disorders? *Molecular pharmacology* **2010**, 77, 126-35.
7. Huber, K.; Doyon, G.; Plaks, J.; Fyne, E.; Mellors, J. W.; Sluis-Cremer, N., Inhibitors of histone deacetylases: correlation between isoform specificity and reactivation of HIV type 1 (HIV-1) from latently infected cells. *The Journal of biological chemistry* **2011**, 286, 22211-8.
8. Agbor-Enoh, S.; Seudieu, C.; Davidson, E.; Dritschilo, A.; Jung, M., Novel inhibitor of Plasmodium histone deacetylase that cures P. berghei-infected mice. *Antimicrobial agents and chemotherapy* **2009**, 53, 1727-34.
9. Andrews, K. T.; Gupta, A. P.; Tran, T. N.; Fairlie, D. P.; Gobert, G. N.; Bozdech, Z., Comparative gene expression profiling of P. falciparum malaria parasites exposed to three different histone deacetylase inhibitors. *PloS one* **2012**, 7, e31847.
10. Andrews, K. T.; Tran, T. N.; Lucke, A. J.; Kahnberg, P.; Le, G. T.; Boyle, G. M.; Gardiner, D. L.; Skinner-Adams, T. S.; Fairlie, D. P., Potent antimalarial activity of histone deacetylase inhibitor analogues. *Antimicrobial agents and chemotherapy* **2008**, 52, 1454-61.
11. Mai, A.; Massa, S.; Rotili, D.; Cerbara, I.; Valente, S.; Pezzi, R.; Simeoni, S.; Ragno, R., Histone deacetylation in epigenetics: an attractive target for anticancer therapy. *Medicinal research reviews* **2005**, 25, 261-309.
12. Vigushin, D. M.; Ali, S.; Pace, P. E.; Mirsaidi, N.; Ito, K.; Adcock, I.; Coombes, R. C., Trichostatin A is a histone deacetylase inhibitor with potent antitumor activity against breast cancer in vivo. *Clinical cancer research : an official journal of the American Association for Cancer Research* **2001**, 7, 971-6.

13. Glozak, M. A.; Seto, E., Histone deacetylases and cancer. *Oncogene* **2007**, 26, 5420-32.
14. Ortiz, A. R.; Pastor, M.; Palomer, A.; Cruciani, G.; Gago, F.; Wade, R. C., Reliability of comparative molecular field analysis models: effects of data scaling and variable selection using a set of human synovial fluid phospholipase A2 inhibitors. *J Med Chem* **1997**, 40, 1136-48.
15. Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C., Prediction of drug binding affinities by comparative binding energy analysis. *J Med Chem* **1995**, 38, 2681-91.
16. Perez, C.; Pastor, M.; Ortiz, A. R.; Gago, F., Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design. *J Med Chem* **1998**, 41, 836-52.
17. Lozano, J. J.; Pastor, M.; Cruciani, G.; Gaedt, K.; Centeno, N. B.; Gago, F.; Sanz, F., 3D-QSAR methods on the basis of ligand-receptor complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 ligands. *Journal of computer-aided molecular design* **2000**, 14, 341-53.

Chapter V

Design, synthesis and biological evaluation of new classes of thieno[3,2-d]pyrimidinone and thieno[1,2,3]triazine as inhibitor of vascular endothelial growth factor receptor-2 (VEGFR-2)

Enrico Perspicace, Valérie Jouan-Hureau, Rino Ragno, Ballante Flavio, Stefania Sartini, Concettina La Motta, Federico Da Settimo, Binbin Chen, Gilbert Kirsch, Serge Schneider, Béatrice Faivre, Stéphanie Hesse

European Journal of Medicinal Chemistry 2013 May;63:765-81

Preamble

*The following study, is a clear example of how the integration of different chemical approaches could lead to the achievement of a new lead compound. By the application of Structure Based (SB) Three-Dimensional Quantitative Structure Activity relationships (3-D QSAR), molecular modeling, organic chemistry and biological investigations, it was possible to detect new VEGFR-2 (KDR) and human umbilical vein endothelial cell (HUVEC) proliferation inhibitors. My contribution to this study focused on the computational-chemistry investigation, mainly characterized by molecular docking simulations and 3-D QSAR analyses. Among the discovered compounds, **2f** showed the highest inhibitory activity (at μM concentration) representing a new lead compound and therefore a starting point to obtain more active derivatives after chemical modifications. This perspective characterized my first research activity abroad in France (Chapter IX): where I've practiced organic synthesis to obtain new thieno [2,3-d] pyrimidinones derivatives as promising VEGFR-2 inhibitors.*

Introduction

Angiogenesis is the process of new blood vessels growth, creating new capillaries from existing vasculature. Angiogenic dys-regulation may be involved in various diseases development and progression including inflammation,¹ tumor growth² and metastasis.³ Among the different protagonists involved in this process, of particular importance are:

- 1) the vascular endothelial growth factors (VEGFs):^{4, 5} VEGF-A (commonly named VEGF), VEGF-B, VEGF-C, VEGF-D, VEGF-E, VEGF-F and the Placental Growth Factor (PlGF);
- 2) the VEGF tyrosine kinase receptors:⁶ VEGFR-1 (Flt-1), VEGFR-2 (KDR) and VEGFR-3 (Flt-4).

So far, many molecules have been approved by FDA and many others are currently in clinical trials as anticancer agents, capable to block VEGF (like bevacizumab:⁷ a humanized monoclonal antibody) or inhibit VEGFRs (like sorafenib (BAY 43-9006)⁸ and sunitinib (SU11248),⁹ Figure 1). Considering that the VEGFR-2 receptor (kinase insert domain receptor, KDR) seems to play a key role in tumour angiogenesis,^{10, 11} the VEGF/VEGFR-2 pathway provides several opportunities by which small molecules can be used as inhibitors of endothelial proliferation and migration and thus as anticancer agents.¹²⁻¹⁴ Considering the impact of tumor angiogenesis and the resistance mechanism to anti-angiogenic compounds, the development of new VEGFR-2 inhibitors is mandatory. This study is characterized by the discovery of new thienopyrimidines, as new class of VEGFR-2 inhibitors, through the application of a multidisciplinary approach (Figure 2) composed of by:

- 1) 3-D QSAR studies
- 2) Molecular modeling applications
- 3) Organic chemistry
- 4) Biological investigations

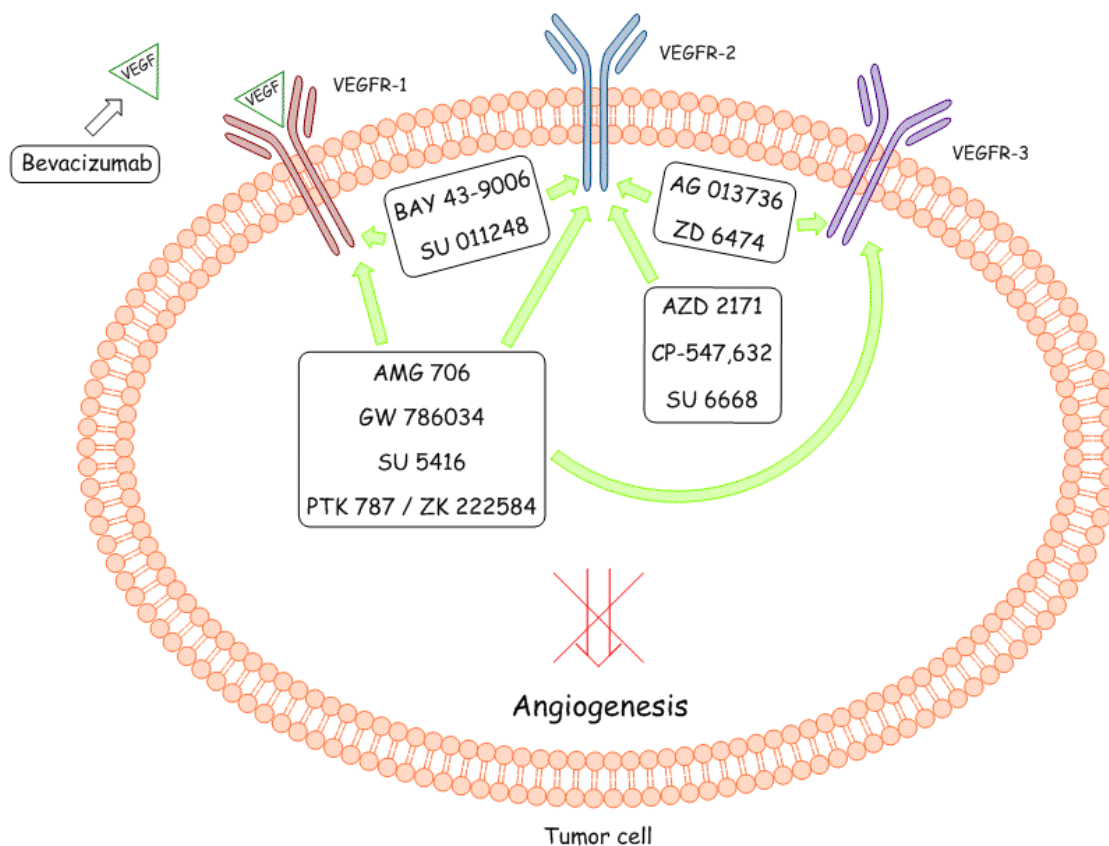


Figure 1. Inhibition pathways of angiogenesis.

Computational studies were developed by means of a reported molecular docking protocol¹⁵ and previously developed Structure-Based (SB) 3-D QSAR models.¹⁶ Molecular Docking allowed to predict the binding poses of compounds for which co-crystal structures were not available with the target protein. Previously developed 3-D QSAR models,¹⁶ characterized by a training set composed of co-crystallized compounds in the ATP-binding site of KDR, were used as a tool 1) to represent three-dimensionally what kind of interactions should be increased, introduced or minimized; 2) to predict the activity of yet untested molecules. Thus, the combination of these two applications composed a useful computational protocol to develop new VEGFR-2 inhibitors.¹⁷

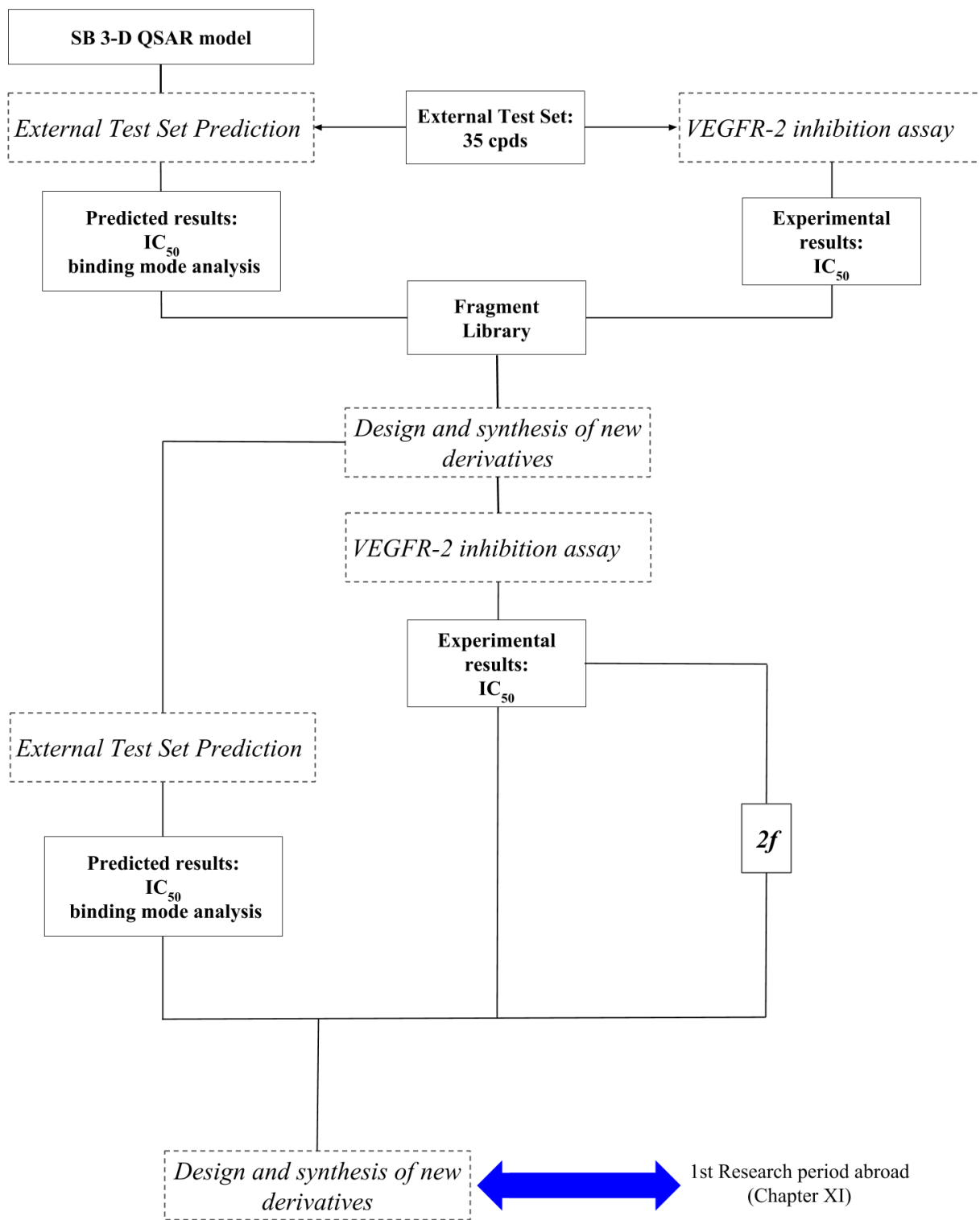


Figure 2. Overview of the multidisciplinary approach.

At the beginning of the study thirty-five compounds (**4-38**),¹⁷ characterized by thienopyrimidine, thiazolotriazine and selenolotriazine scaffolds, were chosen according to their structural similarities in comparison to the most common anti-angiogenic compounds and then biologically assayed to determine their inhibitor potency against VEGFR-2. This set was submitted to a previously reported¹⁵ molecular docking procedure and binding mode analyses were conducted highlighting that:

- 1) all the compounds' binding poses showed the characterizing scaffolds overlapped with the central benzimidazole moiety of the reference structure (PDB 2qu5);
- 2) as a consequence of the previous statement, they bind preferentially in the first part of the binding site
- 3) concerning thienopyrimidines, an increased steric hindrance seems to increase the ligand/VEGFR-2 interactions as for those derivatives characterized by the presence of a third fused cycle
- 4) regarding thiazolotriazines, steric hindrance has a detrimental effect
- 5) selenolotriazines showed binding modes similar to those of the thienopyrimidines

For the four most active compounds (**14**, **30**, **33**, **55**, Appendix A Table A1) we decided to perform a deeper binding mode analysis (Figure 3) revealing that:

- 1) all the predicted poses established strong interactions with a specific region characterized by Val27, Ala45, Lys47, Val93, Val95, Leu164 and Phe176 residues,
- 2) the presence of a morpholino substituent, as in the case of compound **35** (the least active compounds of this selected series) determines an overturning of the selenotriazine group.

The thirty-five compounds were then used as an external test set for the previously developed SB 3-D QSAR protocol.¹⁶ Although over-predictive, 3-D QSAR models proved their predictiveness, since compounds **4-38** were experimentally recognized with a certain rate of inhibition activity.¹⁷ Moreover, useful information was obtained, by superimposing the 3-D QSAR plots with the most active compound (**30**, Figure 4) suggesting that increasing the steric hindrance (i.e. by introducing further sterical groups into the ATP binding site), should enhance the inhibitory potency. Following the above indications new thienopyrimidines/thienotriazines (**1-3**, Appendix A Table A2) were designed and subjected to both molecular docking and 3-D QSAR simulations. 3-D QSARs predicted the new compounds' activity in a range comprised between 0.01 and 0.1 μM (Table 1). Since the 3-D QSAR models were over-predictive, taking into account the observed prediction errors for compounds **4-38**, it was expected that the new derivatives were characterized, at least, by a submicromolar activity. Among these compound **2f** was predicted, as the most active, to be in a low nanomolar range and particular attention was given to its binding mode (Figures 5 and 6): which showed its ability to establish a greater network of steric interactions, especially at the entrance of the active site. Compounds **1-3** were then synthesized and biologically assayed, to detect their inhibitory activity against VEGFR-2, at 200 μM fixed doses. Interestingly, as predicted by the 3-D QSARs, **2f** resulted to be the most active compound (99.2% of inhibition) showing an IC_{50} value equal to $2.25 \pm 0.1 \mu\text{M}$. Moreover, the **2f-tartaric acid salt**, was biological tested on endothelial cell tube formation induced by VEGF, resulting¹⁷ to be more than Sunitinib. Binding mode analysis, suggested that more active compounds could be obtained by introducing further substituents on piperidine positions 3 and 4 (Figure 7), in order to enhance the interactions with the ATP binding pocket. Design of new compounds, as new potential VEGFR-2 inhibitors, was then

performed obtaining a series of new molecules to be synthesized. Three of them were synthesized by myself at the Laboratoire d'Ingénierie et Moléculaire Pharmacologique Biochimie (LIMBP) of the Université de Lorraine Metz (France), directed by Prof. Gilbert Kirsch during my first research period abroad (see Chapter IX).

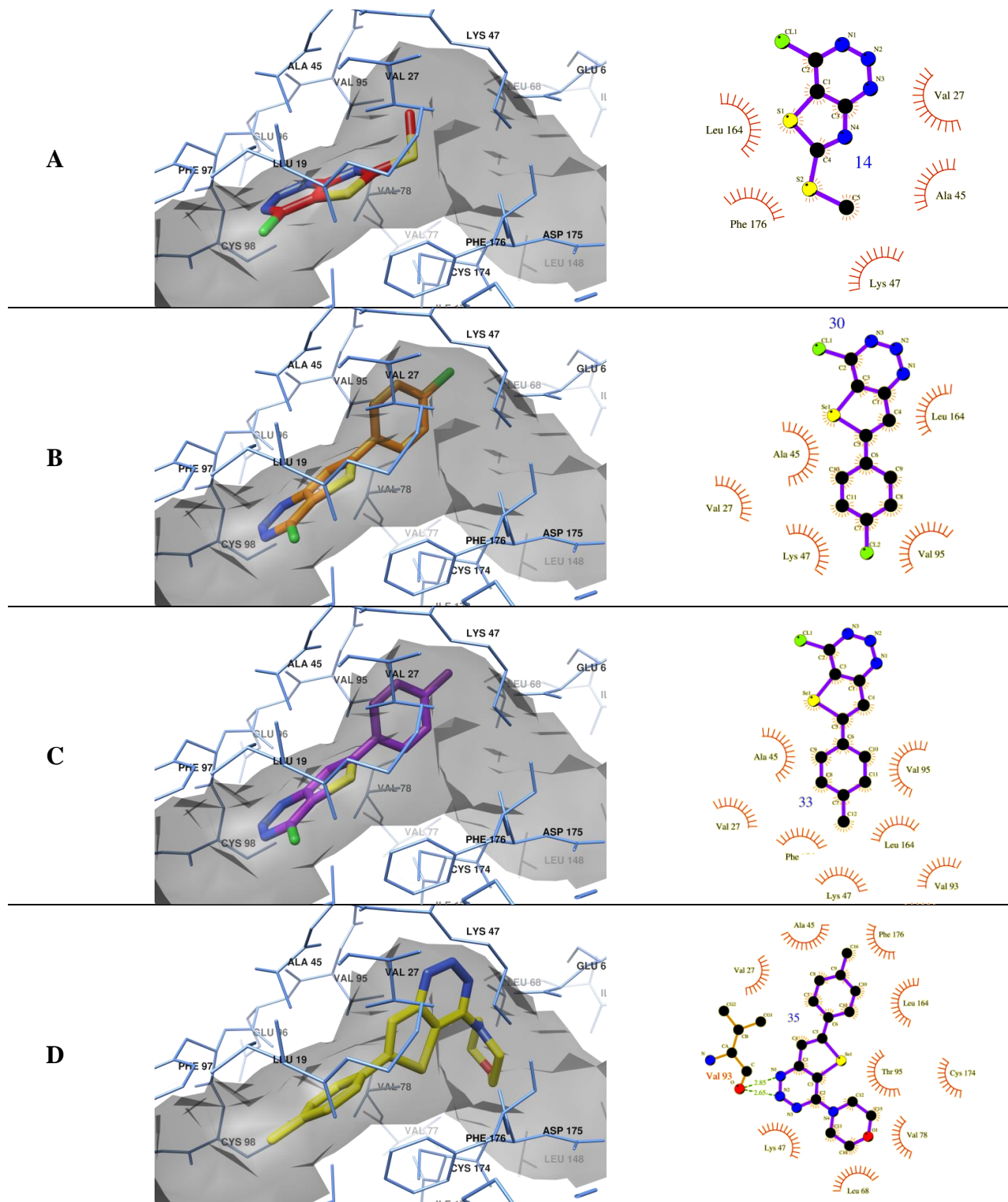


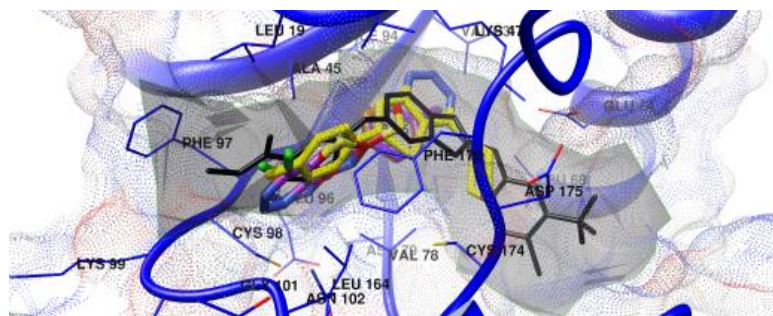
Figure 3. Details of the binding for compounds **14** (A), **30** (B), **33** (C) and **35** (D). To the right are shown the ligand/VEGFR-2 interactions profile as obtained with the ligplot program.

Table 1. Experimental and Predicted pIC₅₀ for compounds **1-3**.

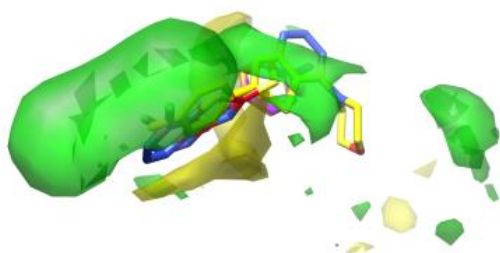
#	% inhibition	IC ₅₀ μM	Predicted pIC ₅₀ by Probe			
	@ 200 μM ^a		A	C	HD	e
1a	38.8±3.5		7.01	7.01	7.02	7.56
1b	n.t. ^b		6.90	6.90	6.80	7.30
1c	1.7±0.1		6.88	6.88	7.03	7.14
2a	5.7±0.4		6.69	6.70	6.99	7.66
2b	n.a. ^c		6.85	6.85	7.08	7.17
2c	3.9±0.2		6.50	6.50	6.70	8.23
2d	2.1±0.1		6.65	6.66	6.62	8.03
2e	n.a. ^c		6.68	6.68	7.02	7.60
2f	99.2±3.9	2.25±0.1	8.12	8.13	8.28	7.15
2g	52.6±3.7		7.96	7.96	7.87	7.87
2h	n.a. ^c		7.34	7.34	7.31	8.65
2i	19.0±0.7		7.16	7.16	7.13	8.72
2j	n.a. ^c		6.57	6.57	6.72	7.17
3a	35.2±2.8		6.90	6.91	7.16	7.43
3b	40.9±2.8		6.11	6.10	6.22	7.54
3c	52.0±3.6		6.74	6.73	6.74	7.66
3d	73.4±6.5		6.30	6.30	6.46	7.51

^aValues are means±SEM of two determinations carried out in triplicate. ^bNot tested.

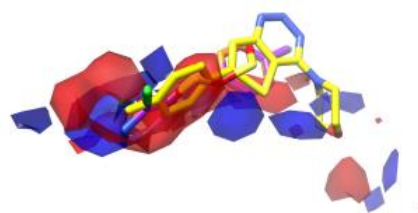
^cNot active. No inhibition was observed up to 200 μM of the test compound.



A

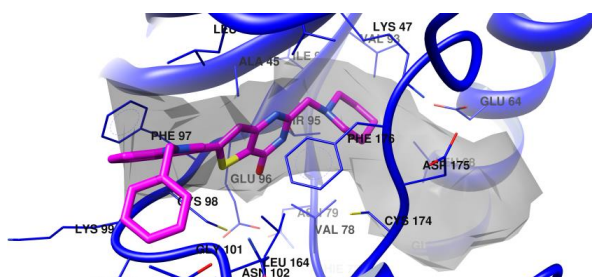


B



C

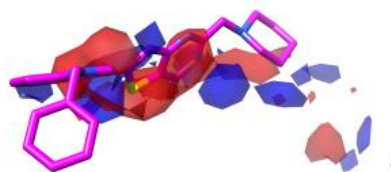
Figure 4. Docked conformation of **14** (red), **30** (orange), **33** (purple) and **35** (yellow) (A) into VEGFR-2 (PDB ID 2qu5, blue ribbon). As reference the co-crystallized ligand is also displayed in black. The surface of ATP and co-crystallized inhibitor binding site is also shown in light gray. Merged in the steric (B) and electrostatic (C) 3-D QSAR maps.



A



B



C

Figure 5. Docked conformation of **2f** (magenta) (A) into VEGFR-2 (PDB ID 2qu5, blue ribbon). The ATP binding site is also shown in light gray. Merged in the steric (B) and electrostatic (C) 3-D QSAR maps.

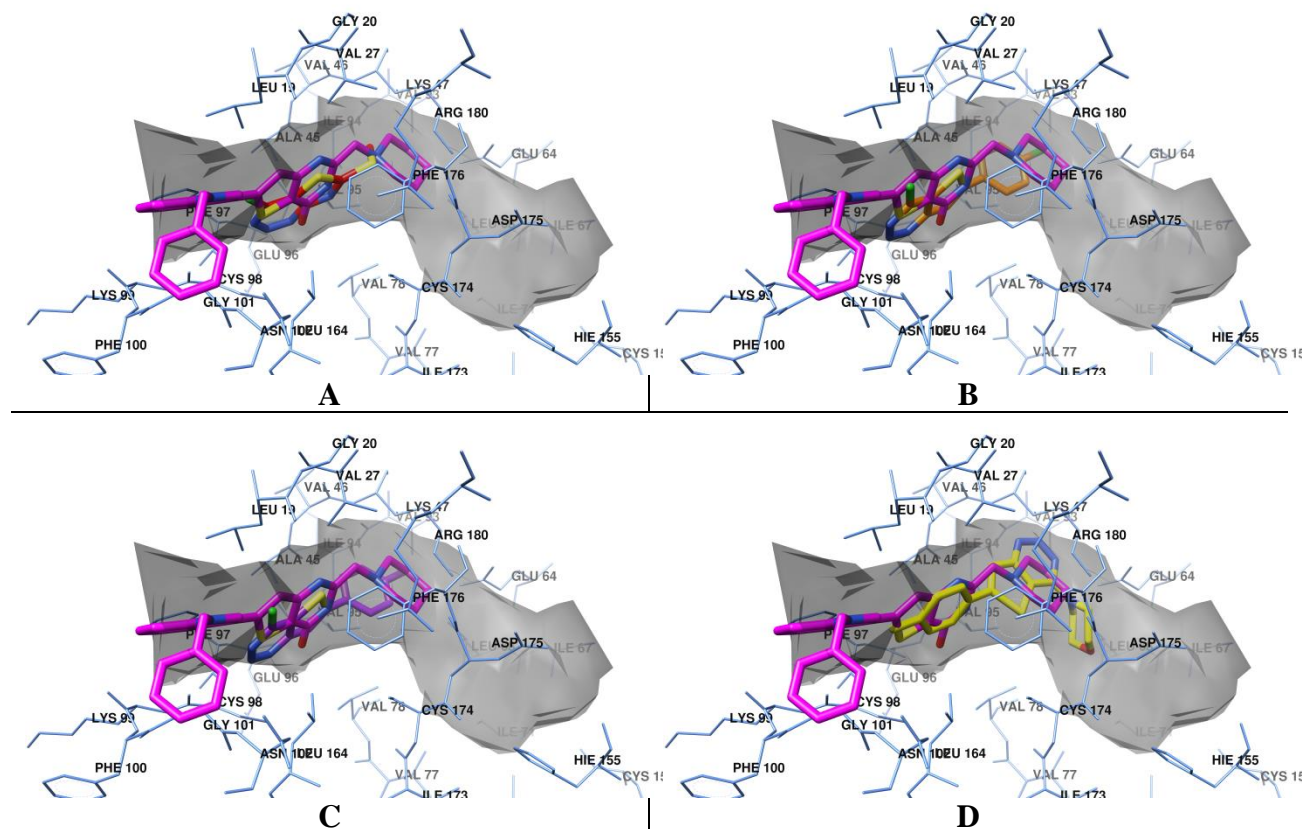


Figure 6. Details of the binding for compounds **2f**, overlapped to **14** (A), **30** (B), **33** (C) and **35** (D).

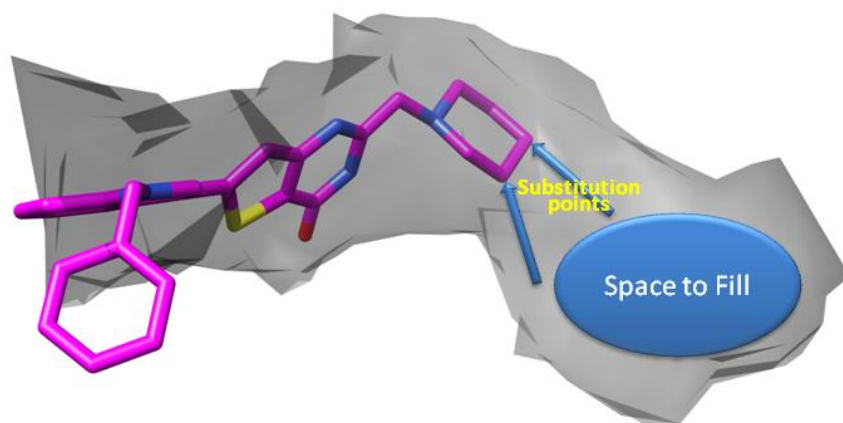


Figure 7. Binding mode analyses suggest to introduce further substituent on the **2f** piperidine ring. **2f** is shown in magenta. The grey background represent the ATP binding site.

APPENDIX A

Table A1. Predicted pIC_{50} for compounds **14**, **30**, **33** and **35**.

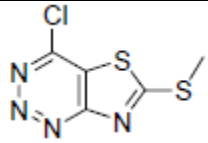
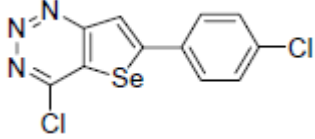
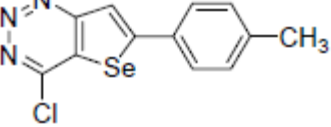
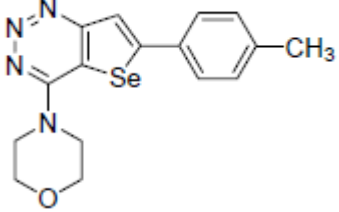
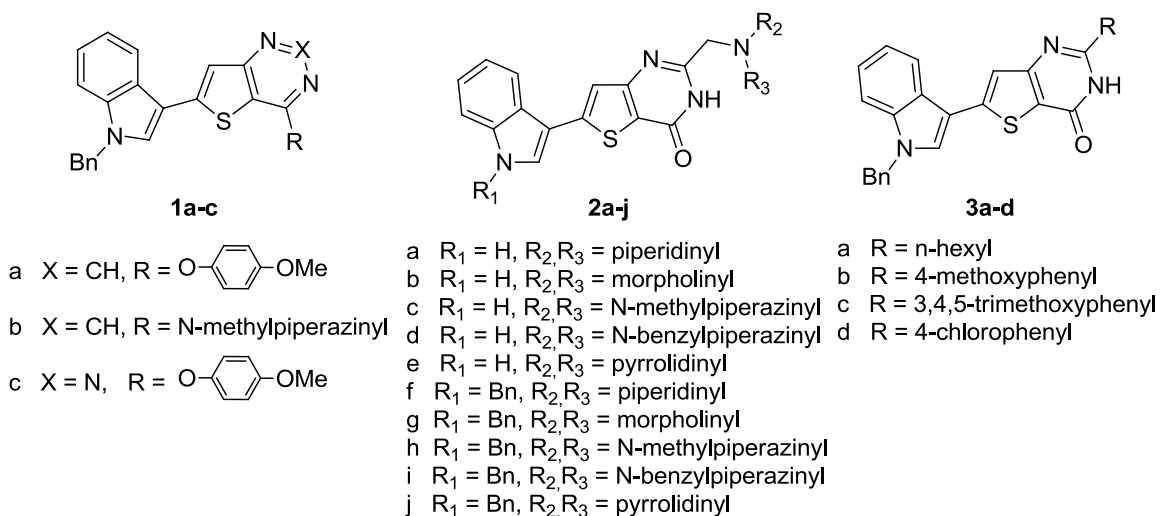
#	Molecular Structure	Exp.	3-D QSAR Probes			
			A	C	HD	e
14		4.34	5.98	5.99	6.31	7.27
30		5.39	6.00	6.01	6.35	7.27
33		4.78	6.00	6.00	6.35	7.28
35		4.58	7.11	7.12	7.48	7.19

Table A2. Designed Compounds.



Bibliography

1. Fava, G. A., Affective disorders and endocrine disease. New insights from psychosomatic studies. *Psychosomatics* **1994**, 35, 341-53.
2. Folkman, J., Anti-angiogenesis: new concept for therapy of solid tumors. *Annals of surgery* **1972**, 175, 409-16.
3. Liotta, L. A.; Steeg, P. S.; Stetler-Stevenson, W. G., Cancer metastasis and angiogenesis: an imbalance of positive and negative regulation. *Cell* **1991**, 64, 327-36.
4. Ferrara, N., VEGF and the quest for tumour angiogenesis factors. *Nature reviews. Cancer* **2002**, 2, 795-803.
5. Ferrara, N.; Kerbel, R. S., Angiogenesis as a therapeutic target. *Nature* **2005**, 438, 967-974.
6. Ferrara, N.; Gerber, H. P.; LeCouter, J., The biology of VEGF and its receptors. *Nature medicine* **2003**, 9, 669-76.
7. Ferrara, N.; Hillan, K. J.; Gerber, H. P.; Novotny, W., Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nature reviews. Drug discovery* **2004**, 3, 391-400.
8. Wilhelm, S. M.; Carter, C.; Tang, L.; Wilkie, D.; McNabola, A.; Rong, H.; Chen, C.; Zhang, X.; Vincent, P.; McHugh, M.; Cao, Y.; Shujath, J.; Gawlak, S.; Eveleigh, D.; Rowley, B.; Liu, L.; Adnane, L.; Lynch, M.; Auclair, D.; Taylor, I.; Gedrich, R.; Voznesensky, A.; Riedl, B.; Post, L. E.; Bollag, G.; Trail, P. A., BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. *Cancer research* **2004**, 64, 7099-109.
9. Motzer, R. J.; Michaelson, M. D.; Redman, B. G.; Hudes, G. R.; Wilding, G.; Figlin, R. A.; Ginsberg, M. S.; Kim, S. T.; Baum, C. M.; DePrimo, S. E.; Li, J. Z.; Bello, C. L.; Theuer, C. P.; George, D. J.; Rini, B. I., Activity of SU11248, a multitargeted inhibitor of vascular endothelial growth factor receptor and platelet-derived growth factor receptor, in patients with metastatic renal cell carcinoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **2006**, 24, 16-24.
10. Almog, N.; Ma, L.; Raychowdhury, R.; Schwager, C.; Erber, R.; Short, S.; Hlatky, L.; Vajkoczy, P.; Huber, P. E.; Folkman, J.; Abdollahi, A., Transcriptional switch of dormant tumors to fast-growing angiogenic phenotype. *Cancer research* **2009**, 69, 836-44.
11. Gimbrone, M. A., Jr.; Leapman, S. B.; Cotran, R. S.; Folkman, J., Tumor dormancy in vivo by prevention of neovascularization. *The Journal of experimental medicine* **1972**, 136, 261-76.

12. Baka, S.; Clamp, A. R.; Jayson, G. C., A review of the latest clinical compounds to inhibit VEGF in pathological angiogenesis. *Expert opinion on therapeutic targets* **2006**, 10, 867-76.
13. Sepp-Lorenzino, L.; Thomas, K. A., Antiangiogenic agents targeting vascular endothelial growth factor and its receptors in clinical development. *Expert opinion on investigational drugs* **2002**, 11, 1447-65.
14. Supuran, C. T.; Scozzafava, A., Protein tyrosine kinase inhibitors as anticancer agents. *Expert Opinion on Therapeutic Patents* **2004**, 14, 35-53.
15. Musmuca, I.; Caroli, A.; Mai, A.; Kaushik-Basu, N.; Arora, P.; Ragno, R., Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors. *Journal of chemical information and modeling* **2010**, 50, 662-76.
16. Ragno, R. VEGFR-2 Inhibitors. Ligand-Based, Structure-Based and 3-D QSAR Studies as Tools to Design New Small Molecules. In XXIII Congresso Nazionale della Società Chimica Italiana, Sorrento, Italy, 2009; Sorrento, Italy, **2009**.
17. Perspicace, E.; Jouan-Hureau, V.; Ragno, R.; Ballante, F.; Sartini, S.; La Motta, C.; Da Settimo, F.; Chen, B.; Kirsch, G.; Schneider, S.; Faivre, B.; Hesse, S., Design, synthesis and biological evaluation of new classes of thieno[3,2-d]pyrimidinone and thieno[1,2,3]triazine as inhibitor of vascular endothelial growth factor receptor-2 (VEGFR-2). *European journal of medicinal chemistry* **2013**, 63, 765-81.

Chapter VI

Pharmacophore assessment through 3-D QSAR: Evaluation of the predictive ability on new derivatives by the application on a series of antitubercular agents

Laura Friggeri, Flavio Ballante, Rino Ragno, Ira Musmuca, Daniela De Vita, Fabrizio Manetti, Mariangela Biava, Luigi Scipione, Roberto Di Santo, Roberta Costi, Marta Feroci, and Silvano Tortorella

Journal of Chemical Information and Modeling 2013 53 (6), 1463-1474

Preamble

In this study, a series of 71 published anti-tubercular agents,¹⁻³ was used as a training set to build, through the 3-D QSAutogrid/R procedure, 8 monoprobe 3-D QSAR PLS models, and a final multiprobe (MP) 3-D QSAR PLS model, able to quantitatively correlate the pharmacophoric features required for antitubercular (anti-TB) activity with molecular structures. All the 3-D QSARs were assessed by comparing their results with a previously published qualitative pharmacophoric model for anti-TB activity, with particular attention for the MP derivation. The obtained 3-D QSAR models were also tested for their predictive ability on a series of new synthesized R-4-amino-3-isoxazolidinone derivatives, confirming to be a promising tool for subsequent virtual screening (VS) applications.

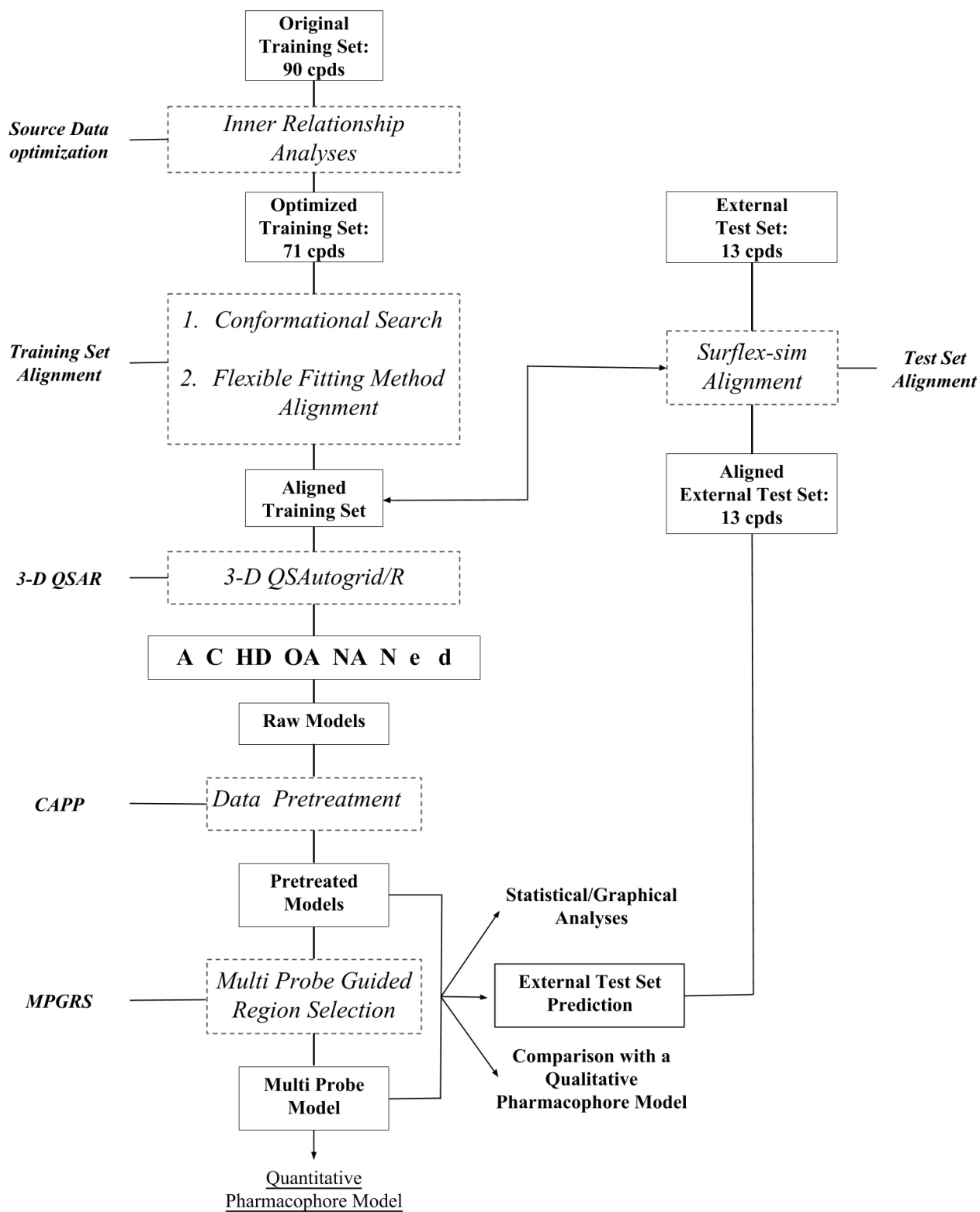


Figure 1. Computational procedure.

Introduction

Tuberculosis (TB) is a serious, as yet unsolved, public-health problem.⁴ There are several reasons that make difficult its eradication, such as long-term drug treatment, poor patient compliance and the development of drug resistant strains like: multidrug-resistant tuberculosis (MDR-TB), extensively drug-resistant tuberculosis (XDR-TB) and totally-drug resistant tuberculosis (TDR-TB).^{5, 6} Therefore, the development of new anti-TB compounds is mandatory in a period in which the spread of HIV has determined, inevitably, a worldwide increase in TB. The study characterizing this chapter, allowed, without any available crystallographic data, to derive, through the 3-D QSAutogrid/R procedure⁷ several 3-D QSAR monoprobe models, and a first three-dimensional pharmacophoric model (by the means of the MPGRS package)⁷ able to quantitatively inspect the areas and features fundamental for anti-TB activity. All the 3-D QSARs were compared to a previously qualitative pharmacophore model,^{8, 9} and tested for their predictive capabilities against an external test set composed of 13 newly synthesized R-4-amino-3-isoxazolidinone derivatives¹⁰ endowed with micromolar activity against *M. tuberculosis* (Figure 1 shows the whole procedure).

Methodology

Starting from a data set composed by 90 thiomorpholines and methylpiperazinyl compounds based on the Pyrrole ring as a scaffold,¹⁻³ for which activity data was originally determined as MIC ($\mu\text{g/mL}$) and used to derive a previously reported qualitative pharmacophore model,^{8, 9} a final training set of 71 compounds (Appendix A Table A1), was selected by application of the inner relationship analysis. It was decided to apply the inner relationship investigation in order to improve the robustness and the predictive capability of the 3-D QSAR models: this

method, generally, allows to discard, from the training set, those molecules recognized as potentially detrimental for the PLS application. Since no crystallographic data was available for these compounds, the same alignment protocol, used to derive the previously developed qualitative pharmacophore model for anti-TB,^{8,9} was adopted.¹⁰ By the application of the 3-D QSAutogrid/R procedure,⁷ 8 monoprobe 3-D QSAR PLS models (Table 1) were built, and optimized through the CAPP procedure,⁷ then, by means of MPGRS package,⁷ a final multiprobe (MP) 3-D QSAR PLS model (Table 2) was derived. Cross-validations (LOO and K5FCV) and Y-scrambling (YS) investigation, confirmed both their internal predictive capability and the absence of chance correlation (Tables 1 and 2). Either the best monoprobe (A, HD, and NA, Figure 2) and multiprobe models were then selected at the optimal principal component (PC)⁷ for further analyses.

Table 1. 3-D QSAutogrid/R PLS models statistical results (CAPP process was applied).

model	P	PC	r^2	q^2_{LOO}	q^2_{K5FCV}	r^2_{YS}	q^2_{YS}	V
1	A	3	0.92	0.86	0.85	0.36	-0.33	3758
2	C	3	0.92	0.86	0.85	0.37	-0.33	4492
3	HD	3	0.91	0.85	0.84	0.39	-0.31	1217
4	NA	3	0.91	0.86	0.85	0.31	-0.33	531
5	N	3	0.91	0.85	0.85	0.32	-0.30	477
6	OA	3	0.91	0.85	0.85	0.36	-0.33	658
7	e	4	0.88	0.78	0.76	0.40	-0.48	468
8	d	4	0.91	0.85	0.84	0.35	-0.44	4412

P: Autogrid Probe, PC: optimal number of principal components/latent variables, r^2 : conventional square-correlation coefficient; q^2_{LOO} : cross-validation correlation coefficient using the leave-one-out method; q^2_{K5FCV} : cross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations; r^2_{YS} : average square-correlation coefficient obtained after Y-scrambling process using 100 iterations; q^2_{YS} : average cross-validation correlation coefficient using the leave-one-out method obtained after Y-scrambling process using 100 iterations; V: number of active variables.

Table 2. MPGRS. Multi Probe model statistical results.

MPGRS 3-D QSAR							
$PC_{FL:SL}$	r^2	q^2_{LOO}	q^2_{K5FCV}	$SDEP_{LOO}$	$SDEP_{K5FCV}$	r^2_{YS}	q^2_{YS}
1:3	0.88	0.80	0.80	0.32	0.32	0.31	-0.31

$PC_{FL:SL}$: optimal number of principal first level (FL) and second level (SL) components for the MPGRS model; r^2 : conventional square-correlation coefficient; q^2_{LOO} : cross-validation correlation coefficient using the leave-one-out method; q^2_{K5FCV} : cross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations.

Single Probe 3-D QSAR models

Analyses of PLS-coefficients plots allowed to define, in complete agreement with the original qualitative pharmacophoric model,^{8, 9} the needed chemical characteristics for anti-TB activity of pyrrole compounds. As for example, PLS-coefficients plots (Figures 3A, 3B, 3C) defined four areas: over N1, C2, C3, and C5 substituents of the pyrrole ring, similarly to those obtained⁹ from the original model (Figure 3D) which defined 4 pharmacophoric features: a hydrogen bond acceptor feature (HBA), two aromatic ring features (RA1, RA2), and an hydrophobic feature (HY). All the selected monoprobe models suggested the positive effect (red colored positive PLS coefficients, Figure 3) of bulky groups as substituent at N1, C2 and C5, respectively overlapping the HY, RA1 and RA2 features; on the contrary at the C3 position (HBA feature) steric groups are less tolerated (blue colored, negative PLS coefficients, Figure 3A and 3C) whereas attractive interactions could increase the biological activity (blue colored, negative PLS coefficients, Figure 3B). Since activity contribution plots have the capability to show the recalculated 3-D activity profile for each molecule, it was possible to highlight, more specifically, how the quantitative models predicted the effects of each training set molecule three dimensionally.

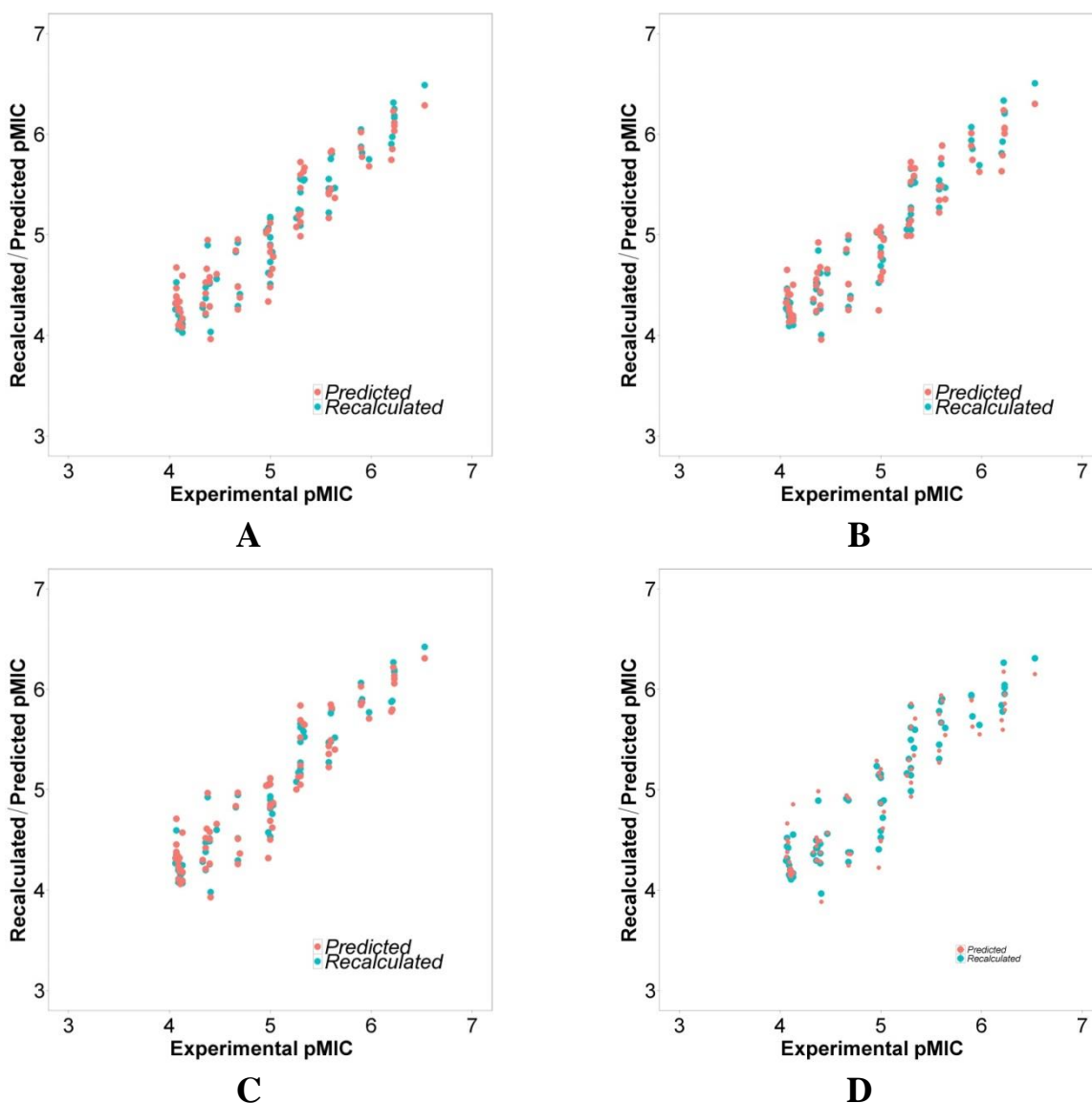


Figure 2. Fitting (r^2) and Cross-Validation (q^2 K-5-Fold) plots. A: from the A probe model at PC3; B: from HD probe model at PC3; C: from the NA probe model at PC3; D: from the multi probe(MP) model at PC_{1:3}.

Analysis of these maps confirmed, generally, the information addressed by the PLS-coefficients plots and allowed to highlight some dependencies between the anti-tubercular activity and the chemical structures. In particular, considering the HBA feature, the HD model plot (Figure 4) clearly showed how the sulfur atom in a thiomorpholine group (characterizing the most active compound **60**, Figure 4A)

increases the biological activity (satisfying the HBA feature), contrary to a situation in which, as for the case of methylpiperazinyl derivatives, repulsive interactions are mainly present in the same area (i.e. the least active molecule **21**, Figure 4B). Moreover, different useful informations were obtained from the simultaneous analysis of PLS-loadings and score plots' clusters, considering the A probe model at the first three principal components (Figure 5), allowing to understand how different molecular conformations and substituents' orientations could affect the biological response. Starting from the first principal component (PC1) was possible to determine that the region between RA2 and HBA features should be preferably occupied than those between RA1 and HBA; PC2 and PC3 highlighted, respectively, how substituents at C3 capable to fill the areas between RA2 and HBA and over HBA features determine a detrimental effect on the biological activity.

Multi-Probe Guided Region-Variable Selection

The MPGRS package⁷ allowed to derive the first quantitative pharmacophoric model able to correlate the structural features of pyrrole derivatives with their biological data. The optimal MP 3-D QSAR model (Table 2) was characterized by standard coefficients similar to those of the monoprobe models (Table 1), but the interpretation was greatly enhanced since all the different monoprobe suggestions were condensed together. By analyzing the different plots, similar conclusion to those obtained from the different mono-probe models were obtained, moreover new information was found. As reported extensively,¹⁰ by means of PLS-coefficients, PLS-loadings plots and score results (Figures 6 and 7) was possible to increase the resolution of the HBA region revealing, in addition to the electrostatic feature, that a limited steric repulsion is tolerated (Figures 6A and 6B, Appendix A Table A3). Moreover, figure 6B shows how the thiomorpholinometil moiety of compound **60** (the most active compound) satisfies both of these features, leading

to higher anti-tubercular activity, contrary to compound **21** (the least active compound).

External test set Prediction analysis

All the 3-D QSARs were externally validated using a set composed of 13 newly synthesized R-4-amino-3-isoxazolidinone derivatives (Table A2), previously tested against the *M. tuberculosis* (Appendix A Table A4). All the 3-D QSARs were able to predict these compounds with low SDEP values, confirming the robustness and predictability of the models.

Table 3. Test Set predictions.

P	PC	SDEP_{EXT}
A	3	0.88
C	3	0.88
HD	3	0.81
NA	3	0.82
N	3	0.83
OA	3	0.84
e	4	0.90
d	4	1.51

SDEP values considering the optimal PCs; P: AutoGrid Probe; PC: optimal number of principal components/latent variables; SDEP_{EXT}: standard deviation error of prediction (or root mean squared error of prediction, RMSEP) for the external test set.

Table 4. MPGRS. Multi Probe model Test Set predictions.

P	PC_{FL:SL}	SDEP_{EXT}
AutoGrid MP	1:3	0.89

SDEP values considering the optimal first level and second level PCs. P: AutoGrid Multi-Probe; PC_{FL:SL}: optimal first level and second level PC; SDEP_{EXT}: standard deviation error of prediction (or root mean squared error of prediction, RMSEP) for the external test set.

Perspectives

The obtained quantitative models were then integrated, together with QSAR applications, in an optimized Virtual Screening (VS) protocol, initially validated using the ChEMBL database and finally applied to NCI Diversity Set. A total of 120 molecules have been identified as potential anti-tuberculosis agents for which it will be determined the anti-mycobacterial activity

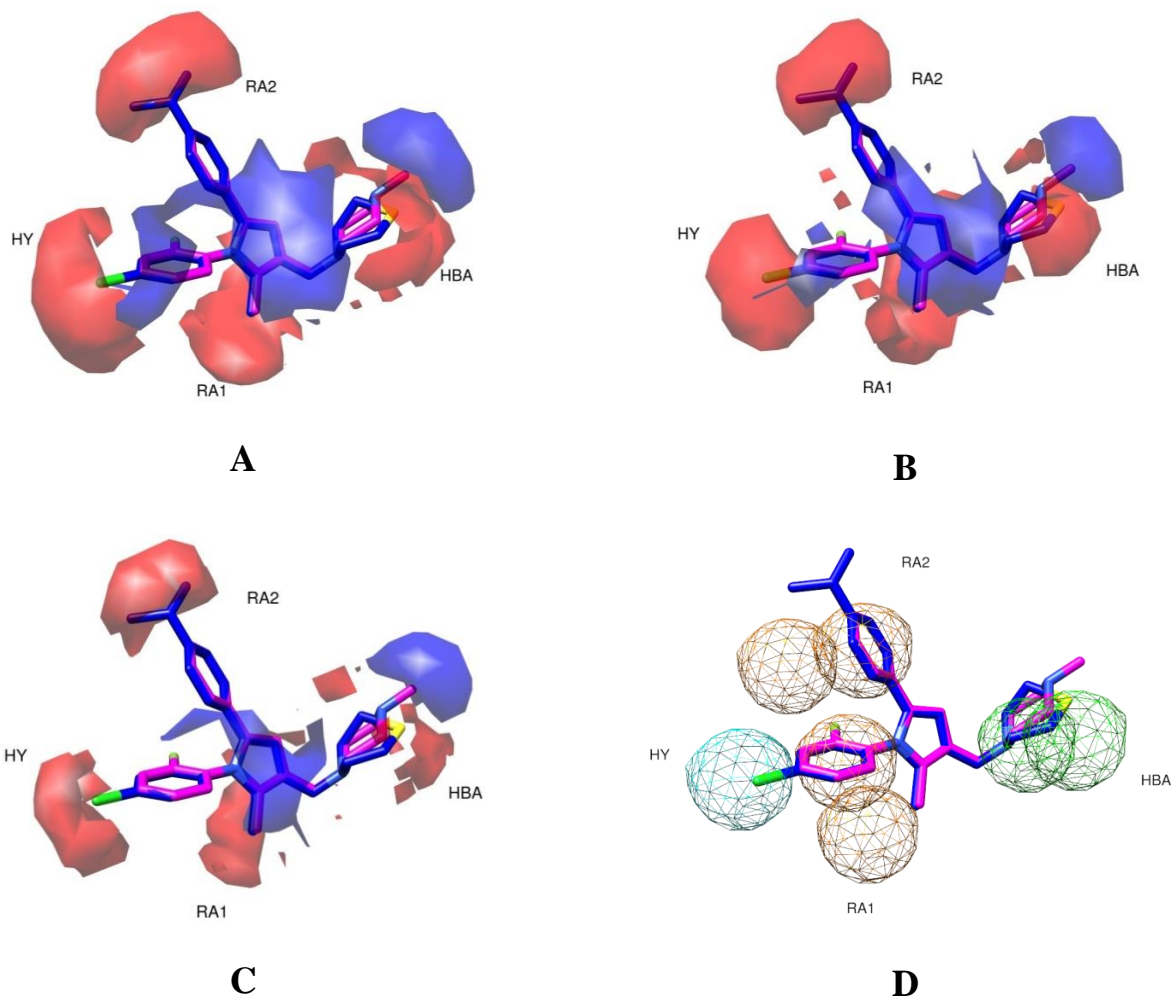


Figure 3. The most active (**60** in blue) and the least active (**21** in magenta) compounds are shown. A: PLS-coefficients contour maps derived from A probe analysis (contour levels: 80%; positive: red, negative: blue); B: PLS-coefficients contour maps derived from HD probe analysis (contour levels: 85%; positive: red, negative: blue); C: PLS-coefficients contour maps derived from NA probe analysis (contour levels: 75%; positive: red, negative: blue). D: pharmacophoric features derived from the original pharmacophoric model:⁹ HY (hydrophobic feature), RA (aromatic feature), HBA (hydrogen bond acceptor feature).

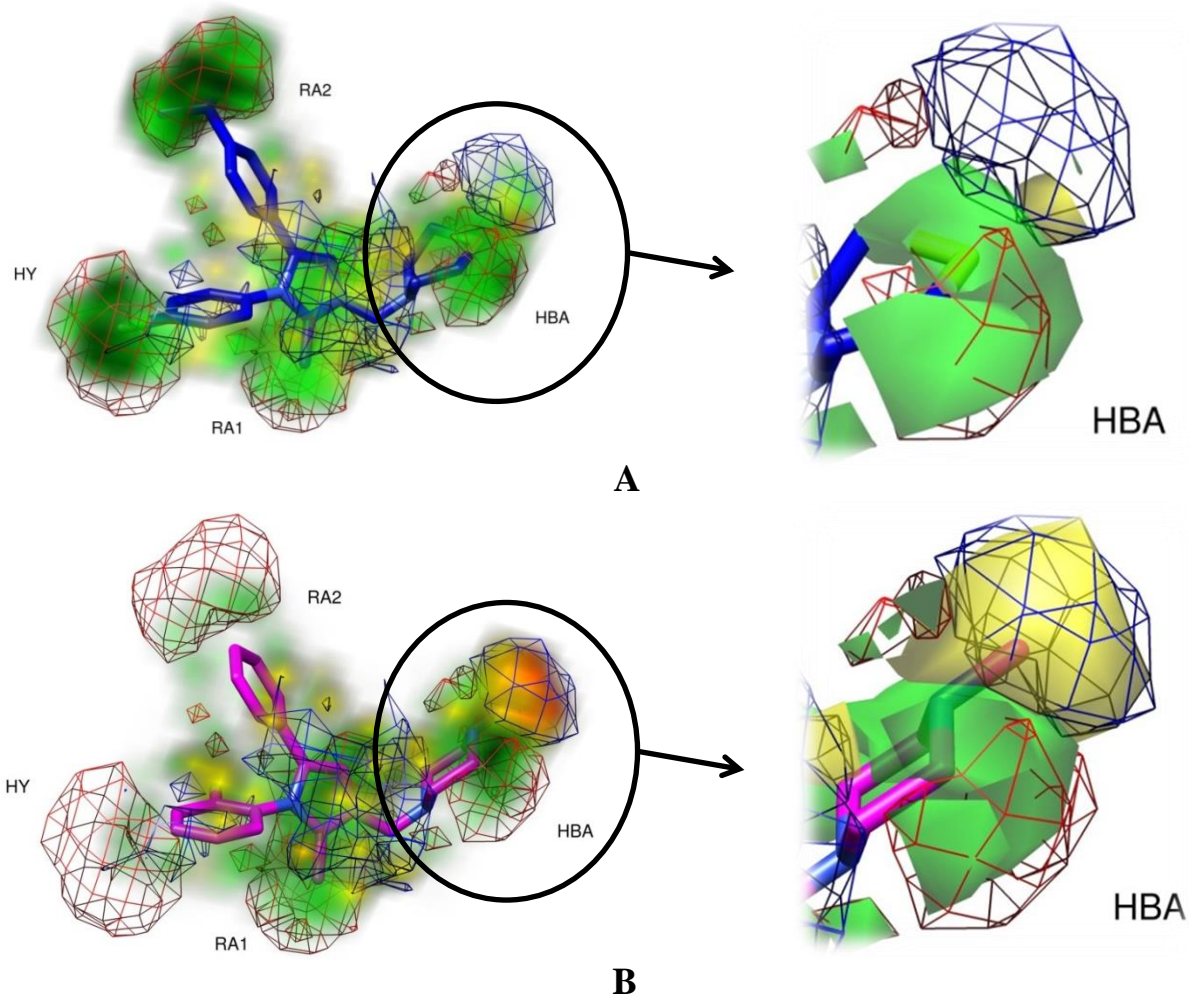


Figure 4. Probe HD. A: PLS-coefficients (mesh: 85%, positive: red, negative: blue) with activity contribution (solid: 90%, positive: green, negative: yellow) for compound **60** (blue); B: PLS-coefficients (mesh: 85%, positive: red, negative: blue) with activity contribution (solid: 90%, positive: green, negative: yellow) for compound **21**(magenta).Activity contributions in pictures on the left side are shown in color gradient: for both green and yellow polygons, the darker areas (the most important) are characterized by the highest numerical coefficients, the lighter areas (the less important) are characterized by the lowest numerical coefficients.

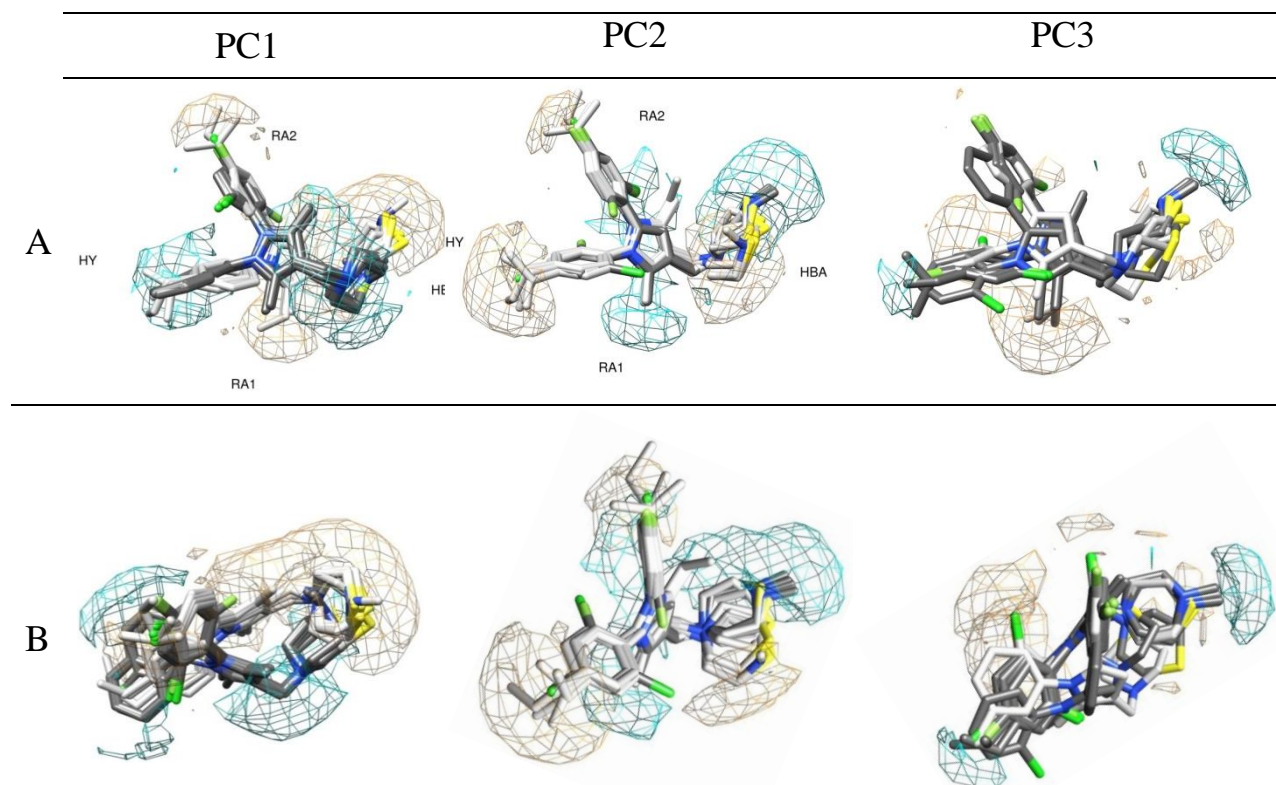


Figure 5. Probe A. PLS-loadings contour maps from the A probe analysis at PC1, PC2 and PC3 (contour levels: 60%; positive: orange, negative: cyan). The ten most important molecules for each cluster are plotted and color coded according to the cluster membership (molecules in the negative field cluster: dark grey, molecules in the positive field cluster: light grey). A: side view; B: top view.

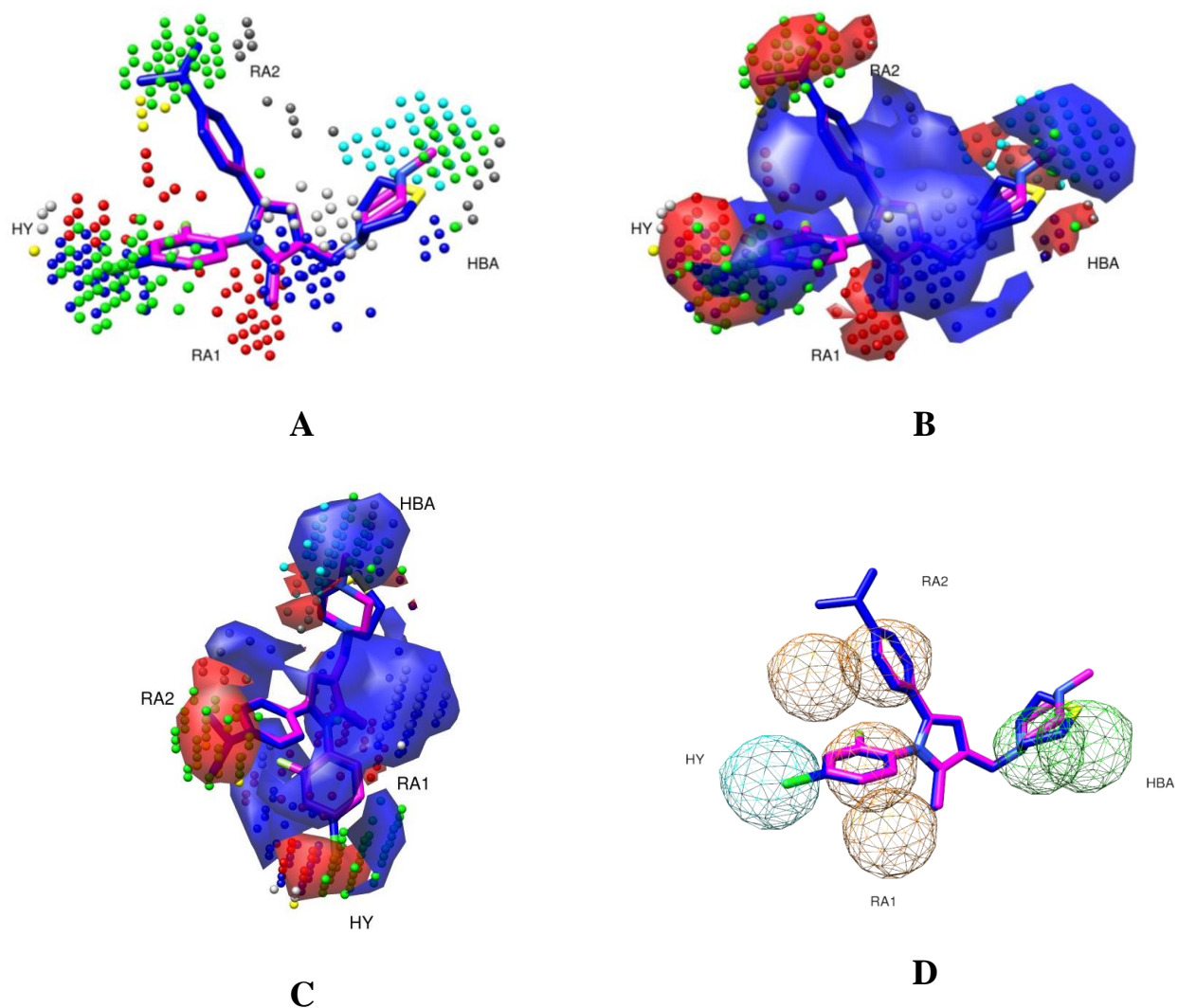


Figure 6. MPGRS. A: key points: the points are color coded according to that reported in Table A3; B: key points with PLS-coefficients contour maps (contour levels: positive 85%, red; negative 95%, blue); C: top view, key points with PLS-coefficients solid contour maps (contour levels: positive 85%, red; negative 95%, blue). The most active (**60** in blue) and the least active (**21** in magenta) compounds are shown. D: pharmacophoric features derived from the original pharmacophoric model:⁹ HY (hydrophobic feature); RA (aromatic feature), HBA (hydrogen bond acceptor feature).

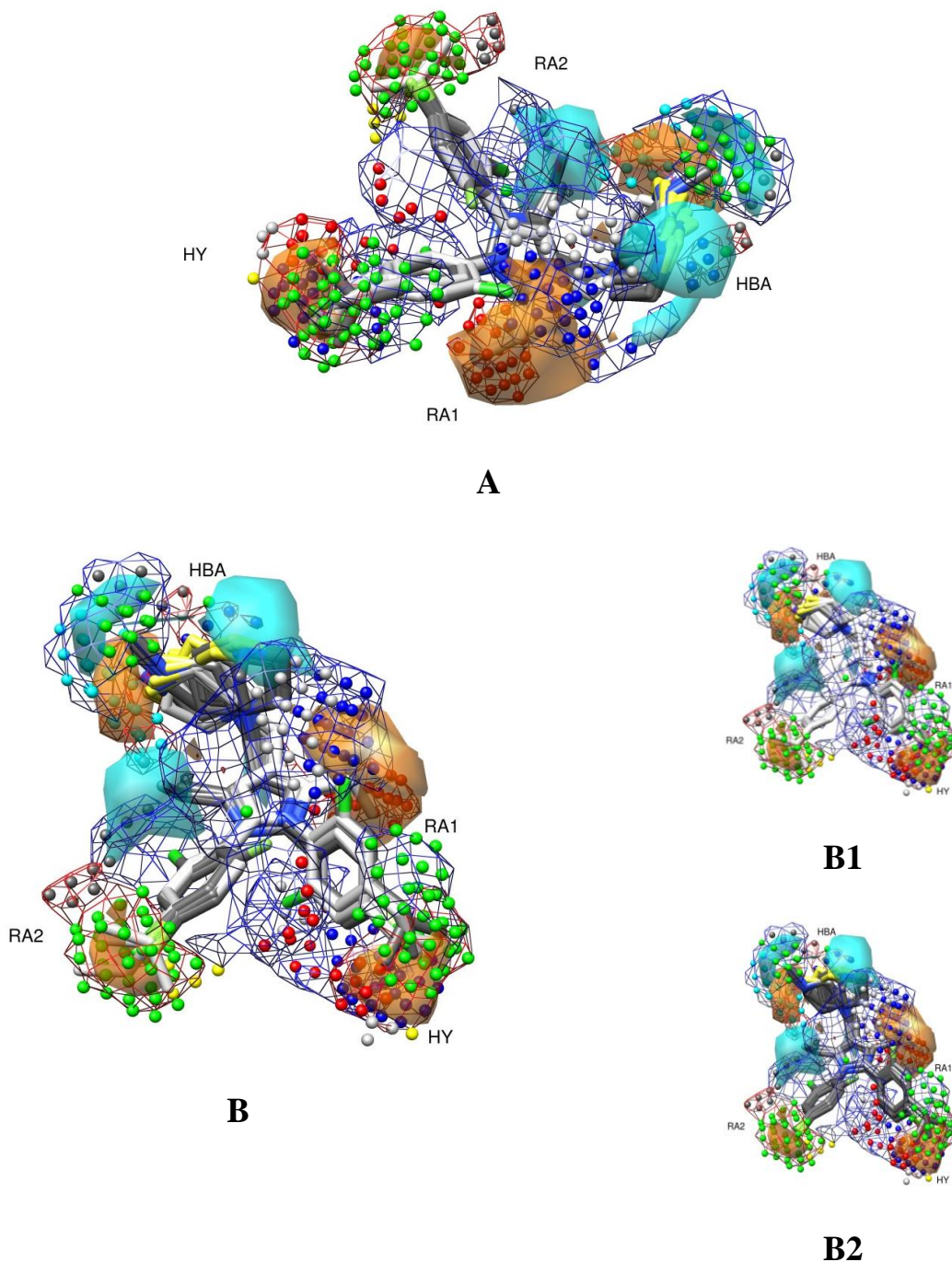
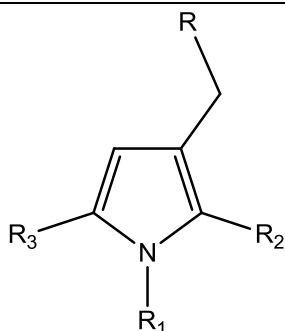


Figure 7. MPGRS. PLS-loadings contour maps at $PC_{1:3}$ (contour levels: 75%; positive: orange, negative: cyan) with PLS-coefficients (mesh levels: positive 85%, red; negative 95%, blue) and key points (see Table S5 for color coding). The ten most important molecules for each cluster are plotted and color coded (compounds in the positive loading field in light gray; compounds in the negative loading field in dark grey). A: side view; B: frontal view; B1: frontal view of only positive clustered molecules; B2: frontal view of only negative clustered molecules. HY (hydrophobic feature), RA (aromatic feature), HBA (hydrogen bond acceptor feature).

APPENDIX A1

Table A1. Structure and antimycobacterial activity against *M. tuberculosis* 103471 of the pyrrole derivatives used as training set for the generation of the 3-D QSAR models.



compd ^a	R ^b	R ₁	R ₂	R ₃	pMIC ^c
1	B	2-F-Ph	CH ₃	2-F-Ph	4.68
2	A	2-Cl-Ph	CH ₃	2-F-Ph	5
3	B	2-Cl-Ph	CH ₃	2-F-Ph	4.09
4	A	2-F-Ph	CH ₃	2-Cl-Ph	5
5	B	2-F-Ph	CH ₃	2-Cl-Ph	4.4
6	A	2-Cl-Ph	CH ₃	2-Cl-Ph	5.02
7	B	2-Cl-Ph	CH ₃	2-Cl-Ph	4.41
8	A	2-F-Ph	CH ₃	α-naphthyl	4.11
9	B	2-F-Ph	CH ₃	α-naphthyl	4.11
10	A	2-Cl-Ph	CH ₃	α-naphthyl	4.13
11	B	2-Cl-Ph	CH ₃	α-naphthyl	4.13
12	A	α-naphthyl	CH ₃	2-Cl-Ph	4.13
13	B	4-F-Ph	CH ₃	Ph	4.36
14	B	Ph	CH ₃	4-F-Ph	4.36
15	A	4-Cl-Ph	CH ₃	4-F-Ph	5.30
16	B	4-F-Ph	CH ₃	4-F-Ph	4.47
17	A	4-F-Ph	CH ₃	4-F-Ph	5.58
18	B	4-F-Ph	CH ₃	4-Cl-Ph	5.30
19	A	4-F-Ph	CH ₃	4-Cl-Ph	5.60
20	A	2-F-Ph	CH ₃	Ph	4.66
21	B	2-F-Ph	CH ₃	Ph	4.06
22	A	Ph	CH ₃	2-F-Ph	4.96
23	B	Ph	CH ₃	2-F-Ph	4.36
24	A	2-Cl-Ph	CH ₃	Ph	4.38
25	B	2-Cl-Ph	CH ₃	Ph	4.07
26	B	Ph	CH ₃	2-Cl-Ph	4.07
27	A	α-naphthyl	CH ₃	Ph	4.1
28	B	α-naphthyl	CH ₃	Ph	4.09
29	A	Ph	CH ₃	α-naphthyl	4.10
30	B	Ph	CH ₃	α-naphthyl	4.09
31	B	Ph	CH ₃	Ph	4.33
32	A	4-F-Ph	CH ₃	2-Cl-Ph	5.00
33	B	4-F-Ph	CH ₃	2-Cl-Ph	4.70
34	B	4-F-Ph	CH ₃	2-F-Ph	4.08
35	A	4-F-Ph	CH ₃	4-CH ₃ -Ph	5.98
36	B	4-F-Ph	CH ₃	3-CH ₃ -Ph	4.37
37	A	4-F-Ph	CH ₃	2-CH ₃ -Ph	4.98
38	B	4-F-Ph	CH ₃	2-CH ₃ -Ph	4.07

39	A	4-F-Ph	CH ₃	2,4-Cl ₂ -Ph	5.34
40	B	4-F-Ph	CH ₃	2,4-F ₂ -Ph	5.00
41	A	2-Cl-Ph	CH ₃	4-F-Ph	5.30
42	B	2-Cl-Ph	CH ₃	4-F-Ph	5.00
43	B	2-F-Ph	CH ₃	4-F-Ph	4.68
44	A	4-CH ₃ -Ph	CH ₃	4-F-Ph	5.58
45	A	3-CH ₃ -Ph	CH ₃	4-F-Ph	4.98
46	B	3-CH ₃ -Ph	CH ₃	4-F-Ph	4.40
47	A	2-CH ₃ -Ph	CH ₃	4-F-Ph	4.68
48	B	2-CH ₃ -Ph	CH ₃	4-F-Ph	4.10
49	A	2,4-Cl ₂ -Ph	CH ₃	4-F-Ph	5.64
50	B	2,4-Cl ₂ -Ph	CH ₃	4-F-Ph	5.03
51	A	2,4-F ₂ -Ph	CH ₃	4-F-Ph	5.30
52	B	2,4-F ₂ -Ph	CH ₃	4-F-Ph	4.40
53	A	4-F-Ph	CH ₃	4-C ₂ H ₅ -Ph	5.60
54	A	4-F-Ph	CH ₃	4- <i>i</i> -propyl-Ph	6.21
55	A	4-C ₂ H ₅ -Ph	CH ₃	4-F-Ph	5.30
56	A	4-C ₃ H ₇ -Ph	CH ₃	4-F-Ph	5.61
57	A	4-Cl-Ph	CH ₃	4-CH ₃ -Ph	5.90
58	A	4-Cl-Ph	CH ₃	4-C ₂ H ₅ -Ph	6.22
59	A	4-Cl-Ph	CH ₃	4-C ₃ H ₇ -Ph	6.23
60	A	4-Cl-Ph	CH ₃	4- <i>i</i> -propyl-Ph	6.53
61	A	4-CH ₃ -Ph	CH ₃	4-Cl-Ph	5.90
62	A	4-C ₂ H ₅ -Ph	CH ₃	4-Cl-Ph	5.91
63	A	4-C ₃ H ₇ -Ph	CH ₃	4-Cl-Ph	6.23
64	A	4- <i>i</i> -propyl-Ph	CH ₃	4-Cl-Ph	6.23
65	B	4-Cl-Ph	C ₂ H ₅	4-Cl-Ph	5.33
66	A	4-F-Ph	C ₂ H ₅	4-CH ₃ -Ph	6.20
67	A	Ph	C ₂ H ₅	Ph	5.26
68	A	Ph	C ₂ H ₅	4-F-Ph	5.58
69	A	4-F-Ph	C ₂ H ₅	Ph	5.28
70	A	2-F-Ph	C ₂ H ₅	4-F-Ph	5.30
71	A	2-F-Ph	C ₂ H ₅	2-F-Ph	5.00

^cCompound enumeration was assigned on the basis of the original increasing numbering from the oldest to the most recent reference. Table S1 shows the connections between the new and original enumerations.

^aA = thiomorpholin-4-yl and B = 4-methylpiperazin-1-yl

^bpMIC = -Log [MIC(μM) x 10⁻⁶]

Table A2. R-4-amino-3-isoxazolidinone derivatives: monocarbamates (**1a-e**), dicarbamates (**2a-f**) and amides (**3h,i**).

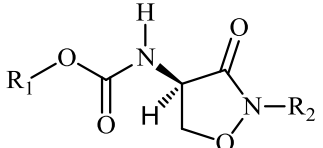
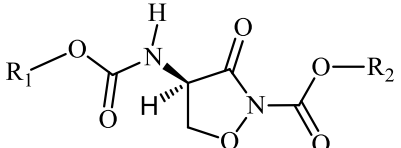




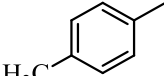
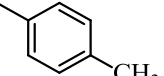
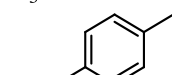
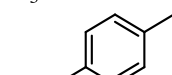
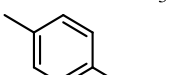
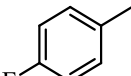
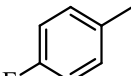
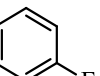
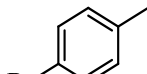
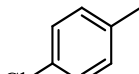
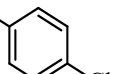
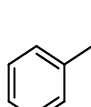
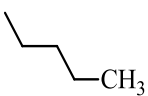
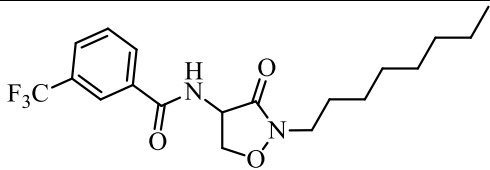
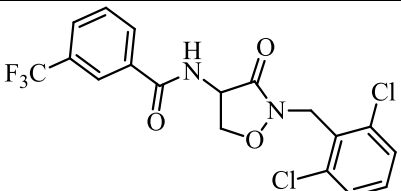
 Monocarbamates 1a-e			 Dicarbamates 2a-f		
#	R ₁	R ₂	#	R ₁	R ₂
1a		H	2a		
1b		H	2b		
1c		H	2c		
1d		H	2d		
1e		H	2e		
			2f		
 3h			 3i		
Amides 3h, 3i					

Table A3. List of the AutoGrid probes employed for MIF calculation and MPGRS subregion color coding.

Probe	Description	MPGRS Colour
A	Aromatic Carbon	Gray
C	Aliphatic (sp ³) Carbon	Dark Gray
HD	Hydrogen bonded to heteroatom	Green
NA	Hydrogen-bond-accepting amine nitrogen	Cyan
N	Amide nitrogen	Blue
OA	Hydrogen-bond-accepting oxygen	Red
e	Electrostatic	Orange
d	Desolvation	Yellow

Table A4. MIC data for D-4-amino-3-isoxazolidinone derivatives

compd	MIC($\mu\text{g/mL}$) ^a	pMIC ^b
1a	32	3.84
1b	32	3.87
1c	32	3.90
1d	3.1	4.89
1e	32	3.97
2a	32	4.03
2b	32	4.06
2c	32	4.10
2d	3.1	5.09
2e	32	4.11
2f	32	4.02
3h	64	3.78
3i	32	4.13

^a*M. tuberculosis* H37Rv (ATCC 27294) was used; MIC values represent the minimal concentrations of compounds completely inhibiting visible growth of mycobacteria.

^bpMIC = -Log [MIC(μM) x 10⁻⁶]

Bibliography

1. Biava, M.; Porretta, G. C.; Poce, G.; Supino, S.; Deidda, D.; Pompei, R.; Molicotti, P.; Manetti, F.; Botta, M., Antimycobacterial agents. Novel diarylpyrrole derivatives of BM212 endowed with high activity toward Mycobacterium tuberculosis and low cytotoxicity. *J Med Chem* **2006**, 49, 4946-52.
2. Biava, M.; Porretta, G. C.; Poce, G.; De Logu, A.; Saddi, M.; Meleddu, R.; Manetti, F.; De Rossi, E.; Botta, M., 1,5-Diphenylpyrrole derivatives as antimycobacterial agents. Probing the influence on antimycobacterial activity of lipophilic substituents at the phenyl rings. *J Med Chem* **2008**, 51, 3644-8.
3. Biava, M.; Porretta, G. C.; Poce, G.; De Logu, A.; Meleddu, R.; De Rossi, E.; Manetti, F.; Botta, M., 1,5-Diaryl-2-ethyl pyrrole derivatives as antimycobacterial agents: design, synthesis, and microbiological evaluation. *European journal of medicinal chemistry* **2009**, 44, 4734-8.
4. Eurosurveillance editorial, t., WHO publishes Global tuberculosis report 2013. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **2013**, 18.
5. Velayati, A. A.; Masjedi, M. R.; Farnia, P.; Tabarsi, P.; Ghanavi, J.; Ziazarifi, A. H.; Hoffner, S. E., Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran. *Chest* **2009**, 136, 420-5.
6. Udwardia, Z. F.; Amale, R. A.; Ajbani, K. K.; Rodrigues, C., Totally drug-resistant tuberculosis in India. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2012**, 54, 579-81.
7. Ballante, F.; Ragno, R., 3-D QSAutogrid/R: an alternative procedure to build 3-D QSAR models. Methodologies and applications. *Journal of chemical information and modeling* **2012**, 52, 1674-85.
8. Manetti, F.; Corelli, F.; Biava, M.; Fioravanti, R.; Porretta, G. C.; Botta, M., Building a pharmacophore model for a novel class of antitubercular compounds. *Farmaco* **2000**, 55, 484-91.
9. Biava; Mariangela; Fioravanti; Rossella; Porretta; Cesare, G.; Deidda; Delia; Lampis; Giorgio; Pompei; Raffaello; Tafi; Andrea; Manetti; Fabrizio, New derivatives of toluidine: Synthesis, antitubercular activity and pharmacophore hypothesis. *Med. Chem. Res* **2002**.
10. Friggeri, L.; Ballante, F.; Ragno, R.; Musmuca, I.; De Vita, D.; Manetti, F.; Biava, M.; Scipione, L.; Di Santo, R.; Costi, R.; Feroci, M.; Tortorella, S., Pharmacophore assessment through 3-D QSAR: evaluation of the predictive ability on new derivatives by the application on a series of antitubercular agents. *Journal of chemical information and modeling* **2013**, 53, 1463-74.

Chapter VII

Hsp90 Inhibitors (I). Definition of 3-D QSAutogrid/R Models as a Tool for Virtual Screening

Flavio Ballante, Antonia Caroli, Richard B. Wickersham III And Rino Ragno

Journal of Chemical Information and Modeling, (2013, submitted)

Preamble

Acknowledged to play a key role in the growth and survival of cancer cells, the multi-chaperone heat shock protein (Hsp) 90 represents a promising target in cancer therapy. This chapter describes a complete computational procedure, for building several structure based (SB) 3-D QSAR models used to derive a final multi-probe (MP) 3-D QSAR pharmacophoric model, able to recognize the most significant chemical features for HSP90 inhibition. All the 3-D QSARs, either mono- and multi-probe, built by means of the 3-DQSAutogrid/R protocol (Chapter II), were externally validated for robustness and predictiveness, and recognized suitable to be used as a predictive tool in a subsequent virtual screening (VS) protocol whose application and results are reported in Chapter VIII.

Introduction

Molecular chaperones are cellular machinery that assist the protein folding under physiological (playing a key role in protein maturation and stabilization) and stress conditions (preventing the formation of a misfolded or aggregated structure).¹ Among chaperones, the 90 kDa heat shock protein (Hsp90) gained much attention in recent years being one of the most extensively studied.^{2,3} Structurally, Hsp90 is a dimeric protein composed of three functional domains: 1) a highly conserved N-terminal ATP-binding domain, 2) a middle domain endowed with high affinity for client proteins and 3) a highly conserved C-terminal domain involved in the homodimerization of the protein.⁴

Hsp90 is present in cells in two isoforms, α (inducible, major form) and β (constitutive, minor form) found predominantly in the cytosol, and two paralogues, ER-resident Grp94 and mitochondrial tumor necrosis factor receptor-associated protein 1 (Trap1). Under basal, non-stressed conditions Hsp90 comprises approximately 1% of the cellular protein population whereas its expression increases significantly after exogenous injury:⁵ indeed, during some cellular stress conditions, such as heat, low pH, nutrient unavailability, hypoxia and malignancy, Hsp90 is overexpressed, promoting the maintenance of structural and functional integrity of client proteins involved in cell survival, proliferation and apoptosis. As reported,⁶ pharmacological inhibition of Hsp90 destabilizes proteins, leading to their degradation through the proteasome; moreover, many oncogenic proteins (i.e. p53 mutants, Raf1, Akt, Bcr-Abl, Her2, EGFR) which are essential for tumor growth are chaperoned by the Hsp90⁷ confirming its involvement in the tumor progression. Since the inhibition of Hsp90 ATPase activity leads to the degradation of client proteins, resulting in cell growth inhibition and apoptosis, Hsp90 emerged as a promising target for cancer therapy.⁸ Natural products such as geldanamycin⁹ and radicicol¹⁰ were among the first Hsp90 inhibitors discovered (Figure 1),

followed by derivatives with better pharmacological properties (17-AAG¹¹ and 17-DMAG,¹² Figure 1). Considering that blocking the ATPase activity is an effective approach, de novo drug design was accomplished and small molecules able to bind the N-terminal Hsp90 binding site were identified. Purine-scaffold inhibitors,¹³ dihydroxyphenylpyrazoles,¹⁴ isoxazole derivatives (NVP-AUY922),¹⁵ and carbazol-4-one benzamide derivatives (SNX-5422)¹⁶ were promising candidates for cancer therapy and advanced to clinical studies (Figure 1).

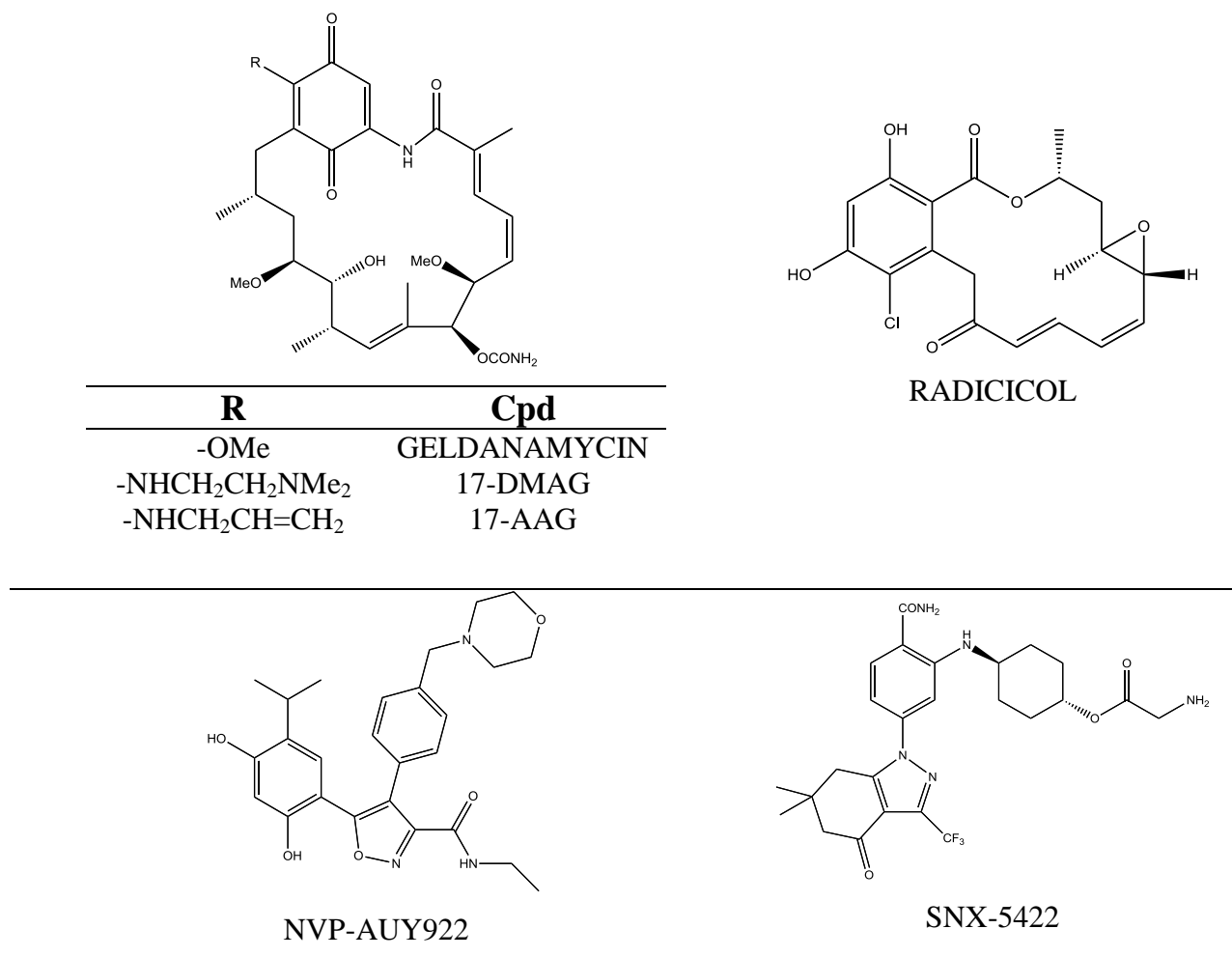


Figure 1. Hsp90 inhibitors: the ansamycin compound geldanamycin and its derivatives radicicol, isoxazole (NVP-AUY922) and carbazol-4-one benzamide (SNX-5422); all of these are in clinical trials.

Despite the huge amount of available data (until now 310 X-ray structures are reported in Protein Data Bank,¹⁷ 199 of which were from Homo Sapiens), only a limited number of 3-D QSAR applications were reported.^{4, 20-27}

In the present work, the 3-D QSAutogrid/R procedure¹⁸ (Chapter II) was applied to a dataset composed of 24 Hsp90/inhibitor co-crystals (Appendix A Table A1) to define three-dimensional quantitative structure-activity relationship (3-D QSAR) models. These were used to derive pharmacophoric quantitative models (as described in the anti-TB application Chapter VI)¹⁹ by the application of Multi Probe Guided Region Selection (MPGRS package),¹⁸ to correlate structural Hsp90 inhibitors features with their biological data. All 3-D QSAR models were externally validated for robustness and predictiveness as a tool in the subsequent virtual screening (VS) protocol (Chapter VIII). Figure 2 shows the overall procedure.

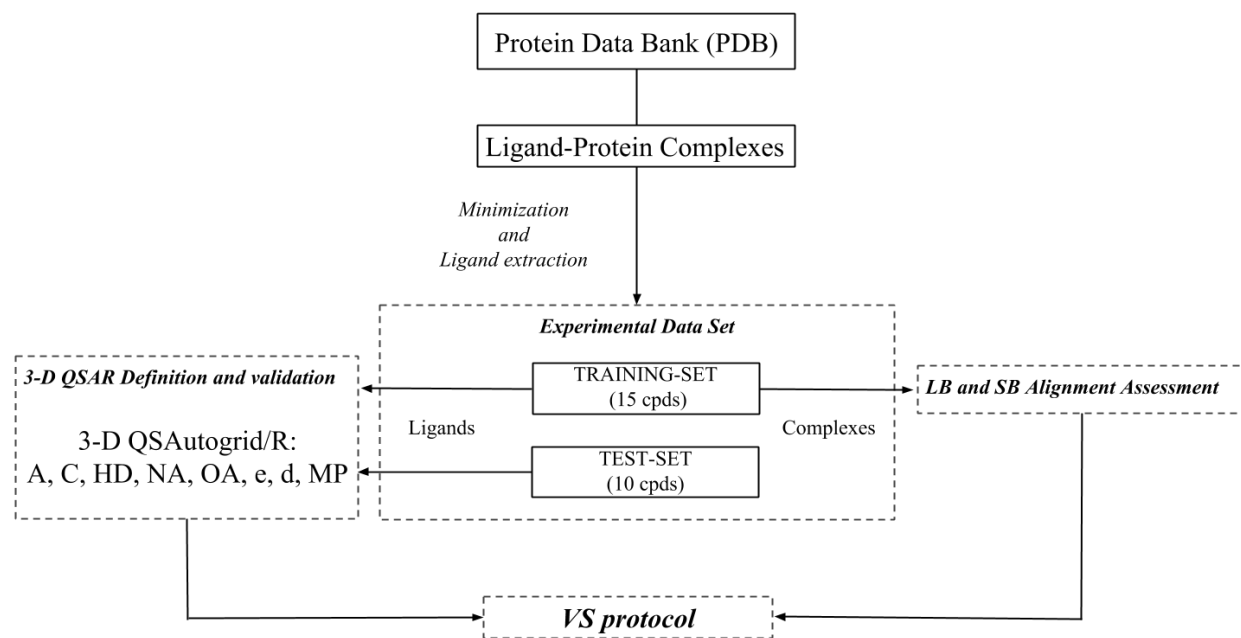


Figure 2. Computational procedure.

Computational methods

At the beginning of this study, all the co-crystallized ligand-protein complexes, available from the Protein Data Bank (PDB),¹⁷ were retrieved. From these, 24 were selected (Appendix A Table A1) as molecular dataset and submitted to a minimization process. Then, the ligands were extracted from the complexes and split into a training set and a test set composed of 15 and 9 compounds, respectively. By means of 3-D QSAutogrid/R procedure,¹⁸ a total of nine 3-D QSAR models were built: 8 monoprobe and a final multiprobe (MP). All the training set inhibitors were used in a ligand (LB) and structure-base (SB) protocols to validate the alignment procedure that was applied on new candidates (with unknown binding mode) during the subsequent virtual screening (VS) application, in which the 3-D QSARs were used as an external scoring function to predict their activity, as reported in Chapter VIII.

Data Set Selection

Initially, 39 co-crystallized ligand-protein complexes were available from the PDB,¹⁷ for which *in vitro* activity was determined mostly through two conventional biological assays: the measure of the ATPase activity (ATPase assay) and the measure of competitive inhibition using a fluorescent probe (Fluorescence Polarization). From these, 24 were tested with the same approach (ATPase assay, IC₅₀ values), resulting eligible to be used in a 3-D QSAR application. The data set compounds (Appendix A Table A1) are structurally related to purine,¹³ pyrazole,¹⁷,³⁸ 2-aminopyrimidine,⁷ triazine⁷ and N-aryl-benzimidazolone²⁰ scaffolds.

N-terminal Hsp90 Binding Site-Inhibitor Complex Structure Preparation

The 24 selected complexes were submitted to a previously reported molecular modeling protocol,²¹ then the minimized conformations of ligands, as extracted from the minimized complexes, were split in training and test sets (see Training Set and Test Set selection paragraph) to build and validate the structure-based statistical models (3-D QSAR), respectively.

Training Set and Test set selection

The data set was partitioned to maintain a similar range of activities between the training and test sets: in particular, the former was composed of 15 compounds (Appendix A, Table A2) with affinity values spanning about 3 orders of magnitude, from pIC₅₀ values of 3.70 to 6.26, and the latter was comprised of 9 compounds (Appendix A, Table A3) with affinity values ranging from 3.70 to 6.28.

Results and Discussion

By means of the 3-D QSAutogrid/R procedure,¹⁸ eight mono-probe 3-D QSAR PLS models were generated and optimized via the CAPP¹⁸ procedure (Table 1).

Table 1. CAPP settings adopted for the 3-D QSAR models.

Min Value	Parameter	Max Value	Step
0	PCO	10	1
0	Zeroing	0.1	0.01
0	MSDCO	1	0.025

PCO: Positive Cut Off, Zeroing: zeroing of very low data points ,MSDCO: Minimum SD Cut Off.

Then, by the application of the MPGRS procedure,¹⁸ a final multi probe (MP) model was derived to define the pharmacophoric features required for Hsp90 inhibitor activity and quantitatively correlate them with molecular structures (See below in the Multi-Probe Guided Region-Variable Selection paragraph). The obtained statistical results confirmed the predictive capabilities and robustness of the mono-probe models (Table 2).

Table 2. Autogrid/R PLS models statistical results (CAPP process was applied).

model	P	PC	r^2	q^2_{LOO}	q^2_{K5FCV}	r^2_{YS}	q^2_{YS}
1	A	2	0.93	0.61	0.59	0.69	-0.46
2	C	2	0.93	0.61	0.58	0.69	-0.50
3	HD	2	0.85	0.55	0.54	0.44	-0.52
4	NA	2	0.93	0.61	0.59	0.71	-0.42
5	N	2	0.93	0.62	0.58	0.68	-0.52
6	OA	2	0.93	0.61	0.59	0.69	-0.47
7	e	2	0.93	0.63	0.60	0.73	-0.50
8	d	1	0.72	0.61	0.60	0.12	-0.22

P:Autogrid Probe, PC: optimal number of principal components/latent variables, r^2 : conventional square-correlation coefficient; q^2_{LOO} : cross-validation correlation coefficient using the leave-one-out method; q^2_{K5FCV} : cross-validation correlation coefficient using the k-fold cross-validation with 5 random groups and 100 iterations; r^2_{YS} : average square correlation coefficient obtained after Y-scrambling process using 100 iterations; q^2_{YS} : average cross-validation correlation coefficient using the leave-one-out method obtained after Y-scrambling process using 100 iterations.

Three of them, obtained with A, N and OA probes (Figure 3), were selected, for further investigations.

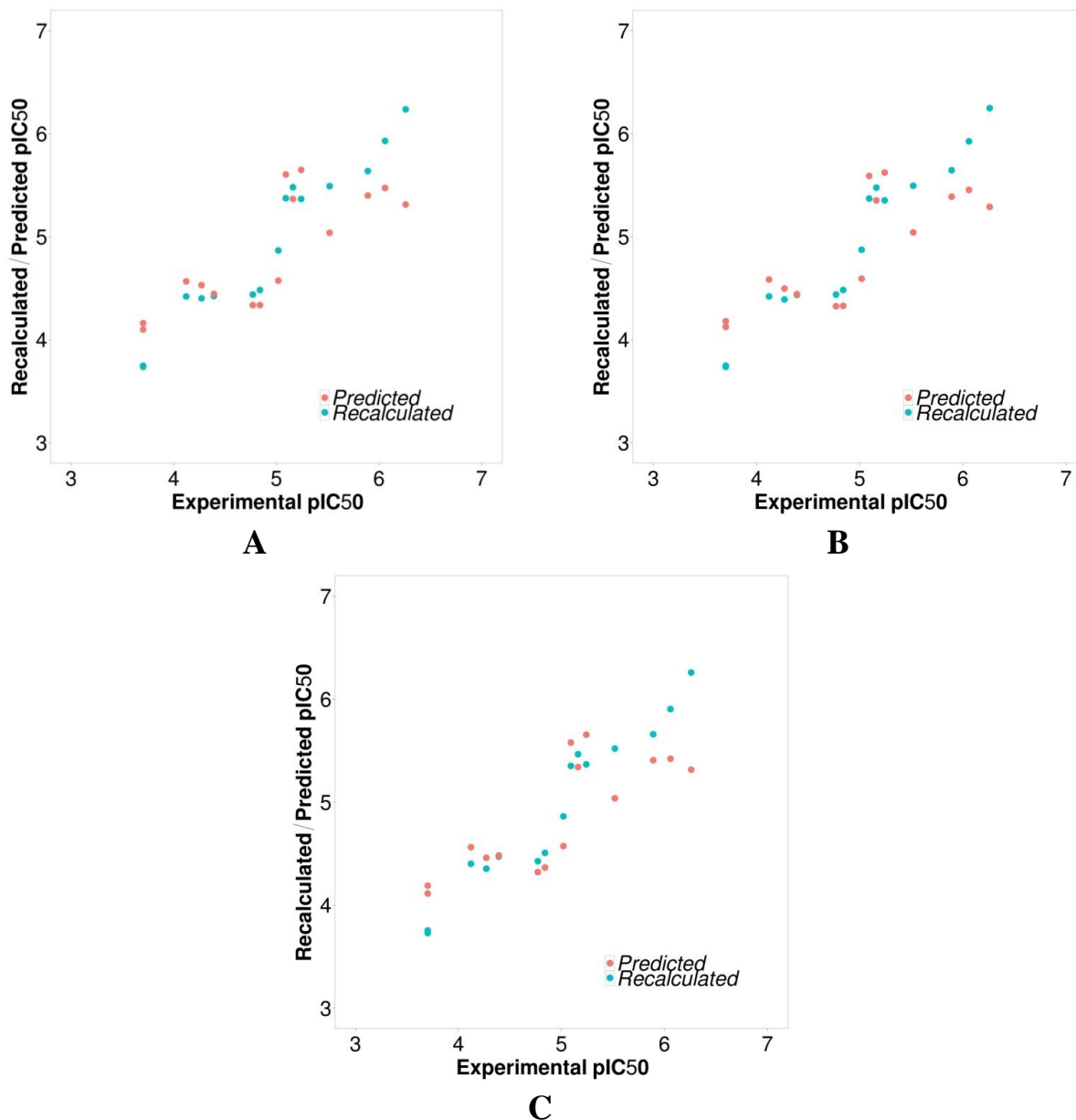


Figure 3. Fitting (r^2 , cyan points) and Cross-Validation (q^2 K-5-Fold, orange points) plots. A: from the A probe model at PC2; B: from N probe model at PC2; C: from the OA probe model at PC2.

Score and PLS-loadings plots were useful to detect the putative most important residues for each cluster of molecules detected by the score plots. Since similar results were obtained from all the three analyzed monoprobe models, only those

obtained from the aromatic (A) are reported. Starting from PC1 (Figure 4A) a clear discrimination between low ($pIC_{50} \leq 5$) and high active compounds ($pIC_{50} > 5$) was observed: indeed molecules characterized by the lowest activities such as all the purine based inhibitors¹³ (1UY8, 1UYC, 1UYD, 1UYE, 1UYG, 1UYH, 1UYK, red bars in Figures 4A) and 3B25⁷ (light green bar) are clustered in the negative field, whereas most of the more active compounds like: pyrazole^{14, 22} and *N*-Aryl-benzimidazolone²⁰ derivatives (2BT0, 2CCS, 2CCU, 3OWB, 3OWD, olive drab and blue bars respectively) are grouped in the positive PC1 score field. 2-aminopyrimidine and triazine derivatives⁷ (3B26 and 3B28, light green and dark green bars respectively) were characterized by low absolute PC1 score values. Considering the association between scores and PLS-loadings, was possible by superimposing the latters (LB extracted data) with the residues active site (SB info) to detect the putative most interacting residues for each cluster (Figures 5 and 6) on the basis of the first principal component (PC1) information. As shown in Figure 5A two “posing” areas were revealed as the main discriminating aspect between these two molecular series. By PLS-loadings analysis (Figure 6A) a series composed of GLY97, ILE96, ASP54, ALA55, LEU48, LYS58 and VAL186 was recognized to interact mainly with the positive clustered molecules (2BT0, 2CCS, 2CCU, 3OWB, 3OWD), whereas a second series composed of LEU107, TYR139 and PHE138 was recognized to interact mainly with the negatives (1UY8, 1UYC, 1UYD, 1UYE, 1UYG, 1UYH, 1UYK and 3B25). This fact suggested the importance in establishing interactions with the first series residues (LB/SB convergence point 1, Table 3). From PC2 (Figure 4B, 5B), this kind of analysis, suggested that a steric group overlapping the corresponding benzyl para-methoxy group of 1UYD (Figure 6B) is detrimental for the activity, indeed from the superimposed crystals, the presence of a conformationally conserved TYR139 side-chain in this area suggests a consequent steric limitation for bulkier ligands

(LB/SB convergence point 2, Table 3). Analyses of PLS-coefficients (ligand-based information) overlapped to the binding pockets (structure-based information) allowed a quali/quantitative definition of the different ligand/receptor interactions, extending the evidences provided by the PLS-loadings/PLS-scores interpretation. A good agreement was observed between the quantitative derivation (PLS-coefficients) and the experimental information (binding pocket residues, Figure 7): i.e. considering the aromatic atom probe A at PC2, the biggest positive coefficients area was mainly located in the sub-pocket delimited by LYS58, ILE96 and, GLY97 residues (Figure 7), suggesting that bioactivity improvement could be obtained by filling this hydrophobic area (LB/SB convergence point 3, Table 3) in agreement with previous reports.^{23, 24}

Negative PLS-coefficients were mainly placed in a space surrounded by LEU107, TYR139 and PHE138, confirming, as anticipated above (see PLS-loadings interpretation), a steric hindrance limitation exerted by TYR139 and PHE138 side chains (LB/SB convergence point 4, Table 3, Figure 7). As a consequence molecules such as the ligand in 1UYD, which overlaps this area, are less active.

A further smaller negative PLS-coefficients area runs among ALA55, SER52 and ASP93 (small blue area in Figure 7) and, considering also both results from OA PLS-coefficients (also negative), these suggest the presence of both steric and electrostatic features in this zone. In fact, these features represent a further LB/SB convergence point (LB/SB convergence point 5, Table 3) by the presence of electrostatic interactions with the three above listed residues (i.e H bonds), in agreement with previous reports.²⁵

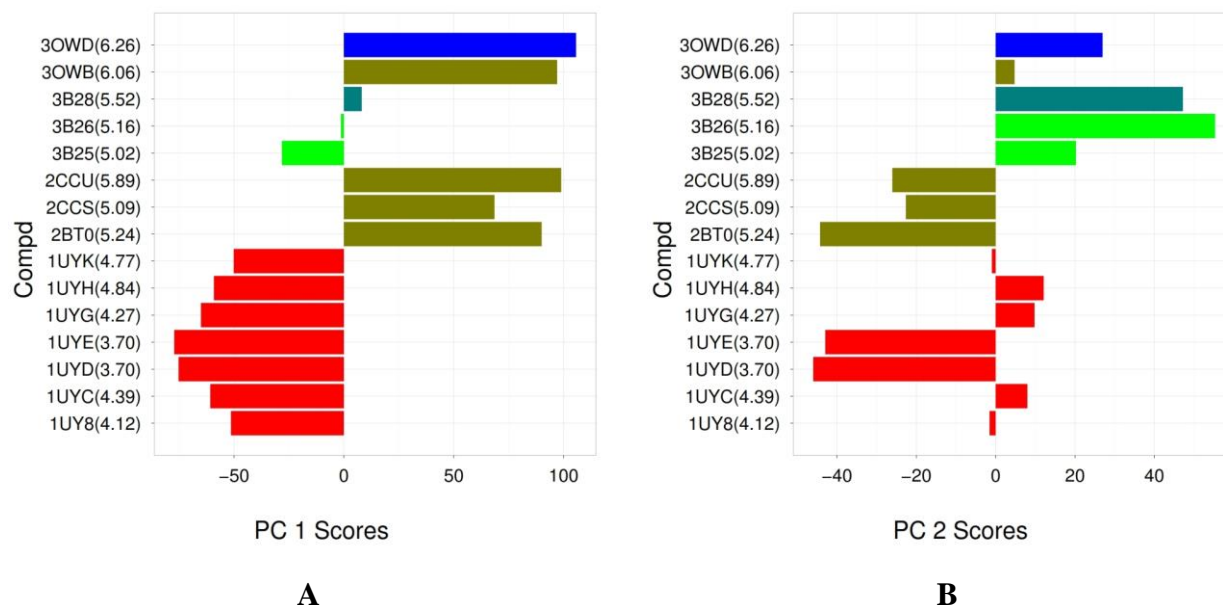


Figure 4. Aromatic atom (A) probe scores plots. Compounds are color-coded according to their scaffold: purine based inhibitors (entries: 1,2,3,4,5,6,7): red bars, 2-aminopyrimidine inhibitors (entries: 17, 18): light green bars, pyrazole inhibitors (entries: 13, 14, 16, 21): olive drab bars, *N*-Aryl-benzimidazolone inhibitors (entry: 22): blue bars, triazine inhibitors (entry: 20): dark green.

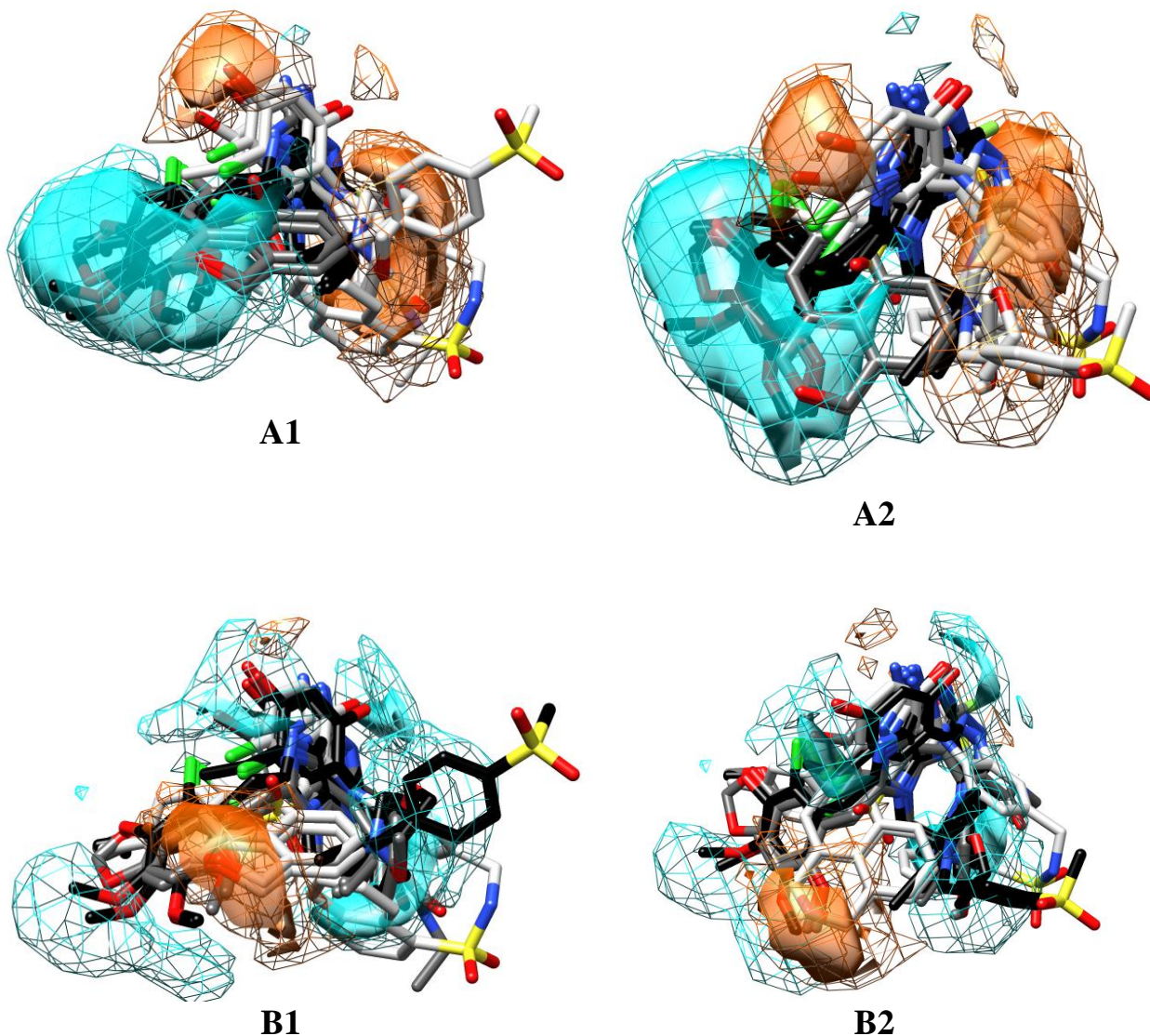
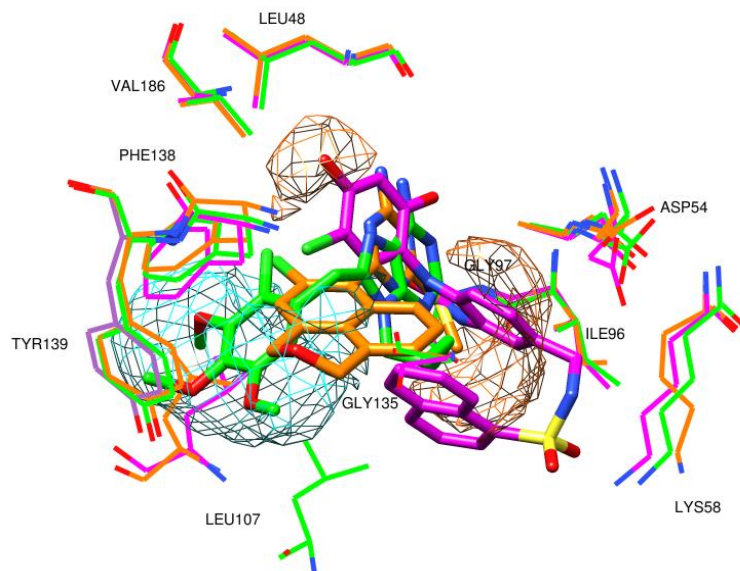
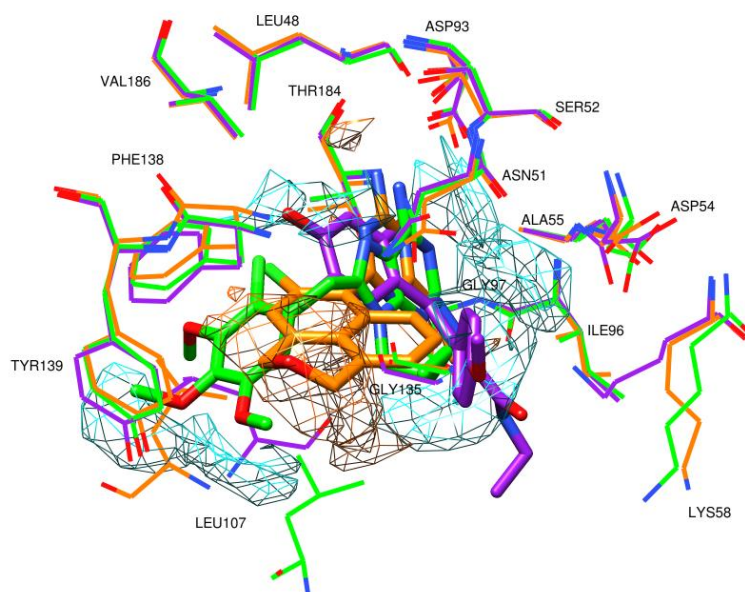


Figure 5. Probe A. A: PLS-loadings contour maps from the A probe analysis at PC1 (contour levels: solid 30%, mesh 60%; positive: orange, negative: cyan). All the molecules for each cluster are plotted and color coded according to the cluster membership (molecules in the negative field cluster: black, molecules in the positive field cluster: light grey, molecules in the center field cluster: dark grey). A1: front view; A2: top view. B: PLS-loadings contour maps from the A probe analysis at PC2 (contour levels: solid 30%, mesh 60%; positive: orange, negative: cyan). All the molecules for each cluster are plotted and color coded according to the cluster membership (molecules in the negative field cluster: black, molecules in the positive field cluster: light grey, molecules in the center field cluster: dark grey). B1: front view; B2: top view. Double level percentage is shown to with the purpose to point out together the most influencing loadings.



A



B

Figure 6. Probe A. A: PLS-loadings contour maps from the A probe analysis at PC1 (contour levels: mesh 30%; positive: orange, negative: cyan). In magenta 3OWD (positive cluster), in green 1UYD (negative cluster), in orange 3B28 (neutral cluster); B: PLS-loadings contour maps from the A probe analysis at PC2 (contour levels: mesh 50%; positive: orange, negative: cyan). In orange 3B28 (positive cluster), in green 1UYD (negative cluster), in purple 3OWB (neutral cluster).

Table 3. LB/SB convergence points.

N.	Reference Probe	Reference PC	Information	LB/SB convergence
1	A, N, OA/ MP	1/1:1	PLS Scores PLS Loadings X-ray	Interactions with GLY97, ILE96, ASP54, ALA55, LEU48, LYS58, VAL 186 are desirable
2	A, N, OA/ MP	2/1:2	PLS Scores PLS Loadings X-ray	Steric hindrance limitation near TYR139, TRP162
3	A, N, OA/ MP	2/1:2	PLS Coefficients X-ray	The hydrophobic sub-pockets composed of LYS58, ILE96 and GLY97, and VAL186 and LEU48, respectively, should be fulfilled
4	A, N, OA	2	PLS Coefficients X-ray	Steric hindrance limitation exerted by TYR139 and PHE138
5	A, N, OA/ MP	2/1:2	PLS Coefficients X-ray	Electrostatic interactions with ALA55, SER52, ASP93, THR184 and ASN51 should be preferable
6	MP	1:2	PLS Coefficients X-ray	Electrostatic interactions with the <i>para</i> -hydroxylic group of TYR139 could be advantageous; weak attractive interactions with the indole nitrogen of TRP162 could be established

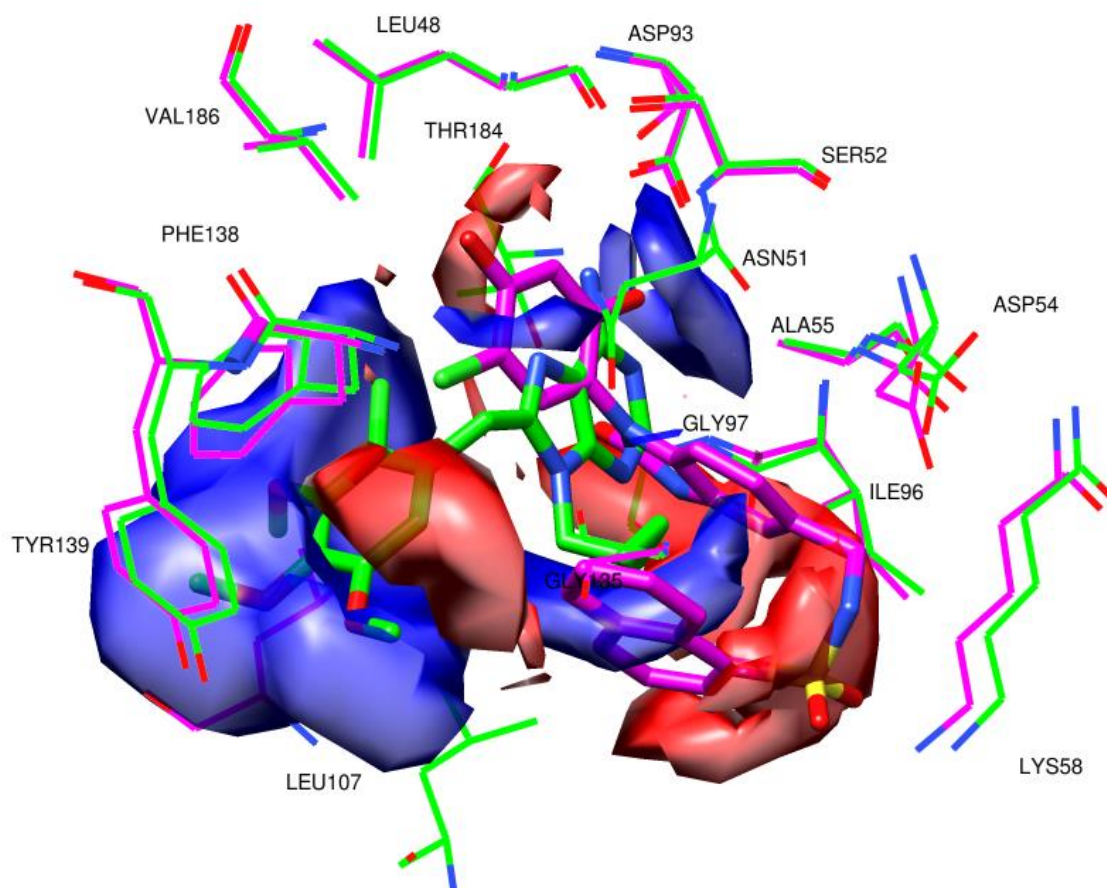


Figure 7. PLS-coefficients contour maps considering 2 PCs. Only the highest active (3OWD in magenta) and the lowest active (1UYD in green) compounds are shown. AutoGrid/R PLS coefficients contour maps derived from A probe analysis (Contour levels: 65%, positive red, negative blue).

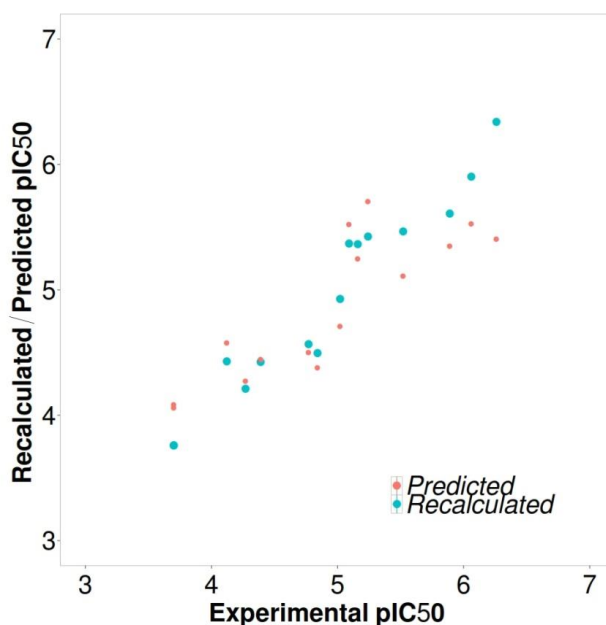
Application of Multi-Probe Guided Region-Variable Selection

A multi-probe (MP) 3-D QSAR model was derived by the application of the Multi Probe Guided Region Selection (MPGRS package).¹⁸ As reported, the MP 3-D QSAR model is obtained at the optimal second level PC of the selected first level PC subregions; in this case PCFL:SL=PC1:2, and its associated statistical coefficients were slightly improved with respect to those of the mono-probe models (Table 4, Figure 8).

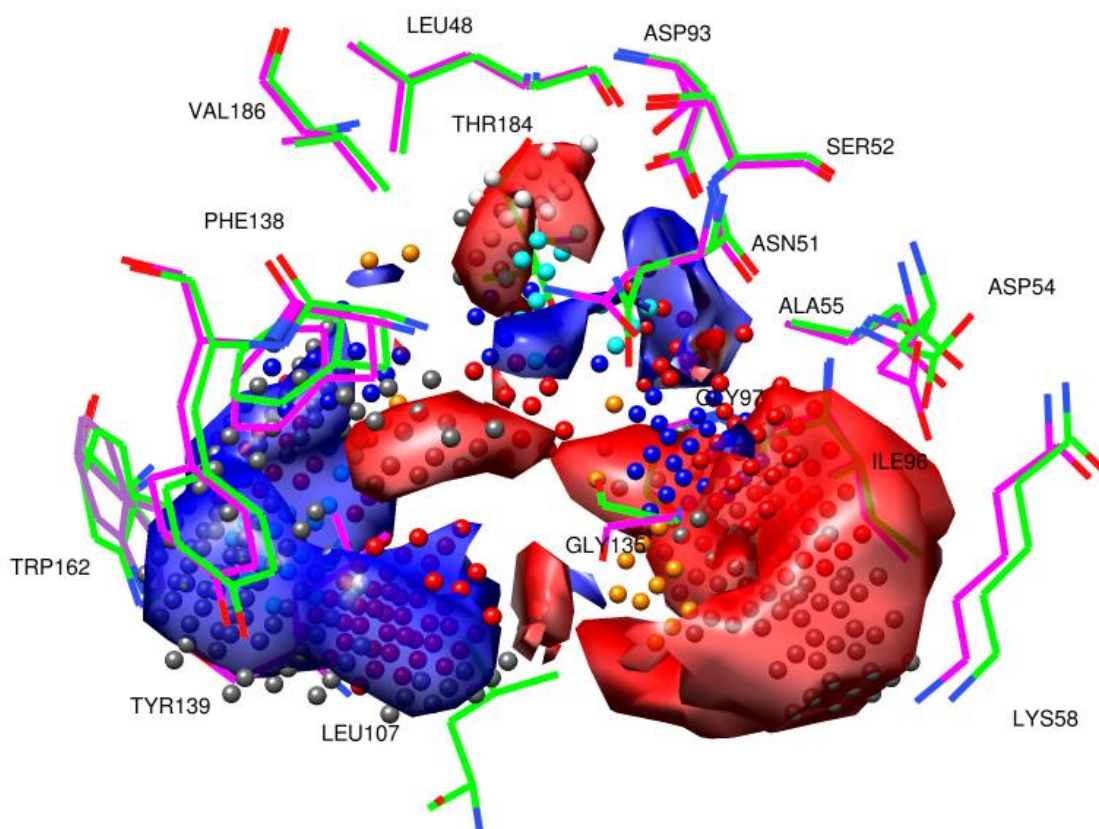
Table 4. Statistical Results Obtained from MPGRS Analysis

MPGRS 3-D QSAR							
$PC_{FL:SL}$	r^2	q^2_{LOO}	q^2_{K5FCV}	$SDEP_{LOO}$	$SDEP_{K5FCV}$	r^2_{YS}	q^2_{YS}
1:2	0.94	0.69	0.68	0.43	0.44	0.67	-0.50

$PC_{FL:SL}$: optimal number of principal first level (FL) and second level (SL) components for the MPGRS model; r^2 : conventional square-correlation coefficient; q^2_{LOO} : cross-validation correlation coefficient using the leave-one-out method; q^2_{K5FCV} : cross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations

**Figure 8.** MPGRS. Fitting (r^2 , cyan points) and Cross-Validation (q^2 K-5-Fold, orange points) plot: from the multi probe(MP) model at $PC_{1:2}$.

Analyses of multi probes scores and loadings led to similar conclusions as the mono-probe models (LB/SB convergence points 1 and 2, Table 3), however, a higher level interpretation allowed identifying both the areas of major interest and the more profitable associated features (Figure 8).



A

Figure 8. MPGRS. A: key points with PLS-coefficients contour maps (contour levels 75%: positive: red; negative: blue). 3OWD in magenta and 1UYD in green. The points are color coded according to that reported in the methodology reference.

Of particular interest are areas simultaneously characterized by negative MP PLS coefficients associated to atom probes able to derive negative fields such as OA, NA and HD. In particular, two distinct areas of negative MP PLS coefficients derived from OA probe are among THR184, ASP93, SER52 and ASN51 (Figure 9A) and TRP162, TYR139 and LEU107 (Figure 9B), suggesting that the establishment of attractive interactions (i.e. H-bond) should be advantageous in the proximity of the first residues' series as for the *para*-hydroxylic group of TYR139 (LB/SB convergence points 5 and 6, Table 3). Furthermore, between TYR139 and

TRP162, a negative PLS-coefficient area (Figure 9B), characterized by NA and C probes, was observed, which suggests, besides the steric limitation, that some weak attractive interaction with the indole nitrogen of TRP162 could be established (LB/SB convergence points 2 and 6, Table 3). Finally, two positive coefficient polyhedra, characterized both by steric/hydrophobic and electrostatic probes (A, C, OA, N and NA), were recognized in the proximity of LYS58 and ILE96, and VAL186 and LEU48 (Figures 9C and 9D, respectively) confirming the importance of placing steric groups in these areas (LB/SB convergence point 3, Table 3).

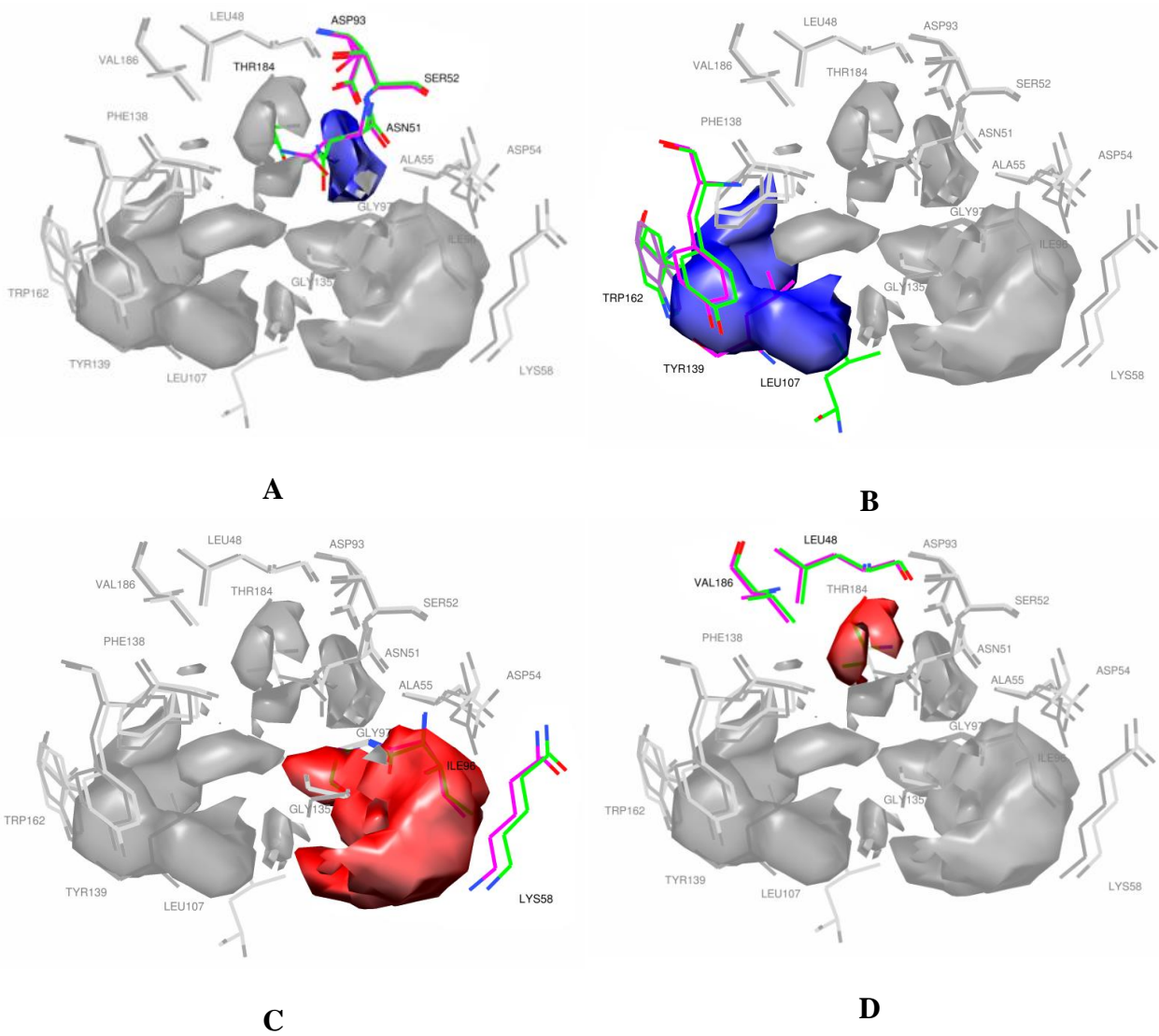


Figure 9. MPGRS. A-D: PLS-coefficients contour maps (contour levels 75%: positive: red; negative: blue), each area is highlighted maintaining its own color.

As a further assessment the multi probe information was compared with a previously reported pharmacophoric model,^{22, 43} showing a high degree of agreement

External Test Set Prediction Analysis.

The external test set composed of 9 Hsp90 inhibitors (Appendix A Table A3) was applied to each of the eight 3-D QSAutoGrid/R mono probe models to assess their predictive capability. In general, low errors of prediction were observed for all models (Table 5), and the correct activity trend was reproduced (Figure10).

Table 5. Test Set predictions: SDEP values considering the optimal PCs; P: AutoGrid Probe; PC: optimal number of principal components/latent variables; SDEP_{EXT}: standard deviation error of prediction (or root mean squared error of prediction, RMSEP) for the external test set.

P	PC	SDEP _{EXT}
A	2	0.79
C	2	0.79
HD	2	0.78
NA	2	0.80
N	2	0.79
OA	2	0.81
e	2	0.79
d	1	0.86

As for the monoprobe models, the MP is endowed with good predictive capabilities (Table 6), reproducing with a good approximation the test set activity trend.

Table 6. MPGRS. Multi Probe model Test Set predictions.

P	PC _{FL:SL}	SDEP _{EXT}
AutoGrid MP	1:2	0.81

SDEP values considering the optimal first level and second level PCs. P: AutoGrid Multi-Probe; PC_{FL:SL}: optimal first level and second level PC; SDEP_{EXT}: standard deviation error of prediction (or root mean squared error of prediction, RMSEP) for the external test set.

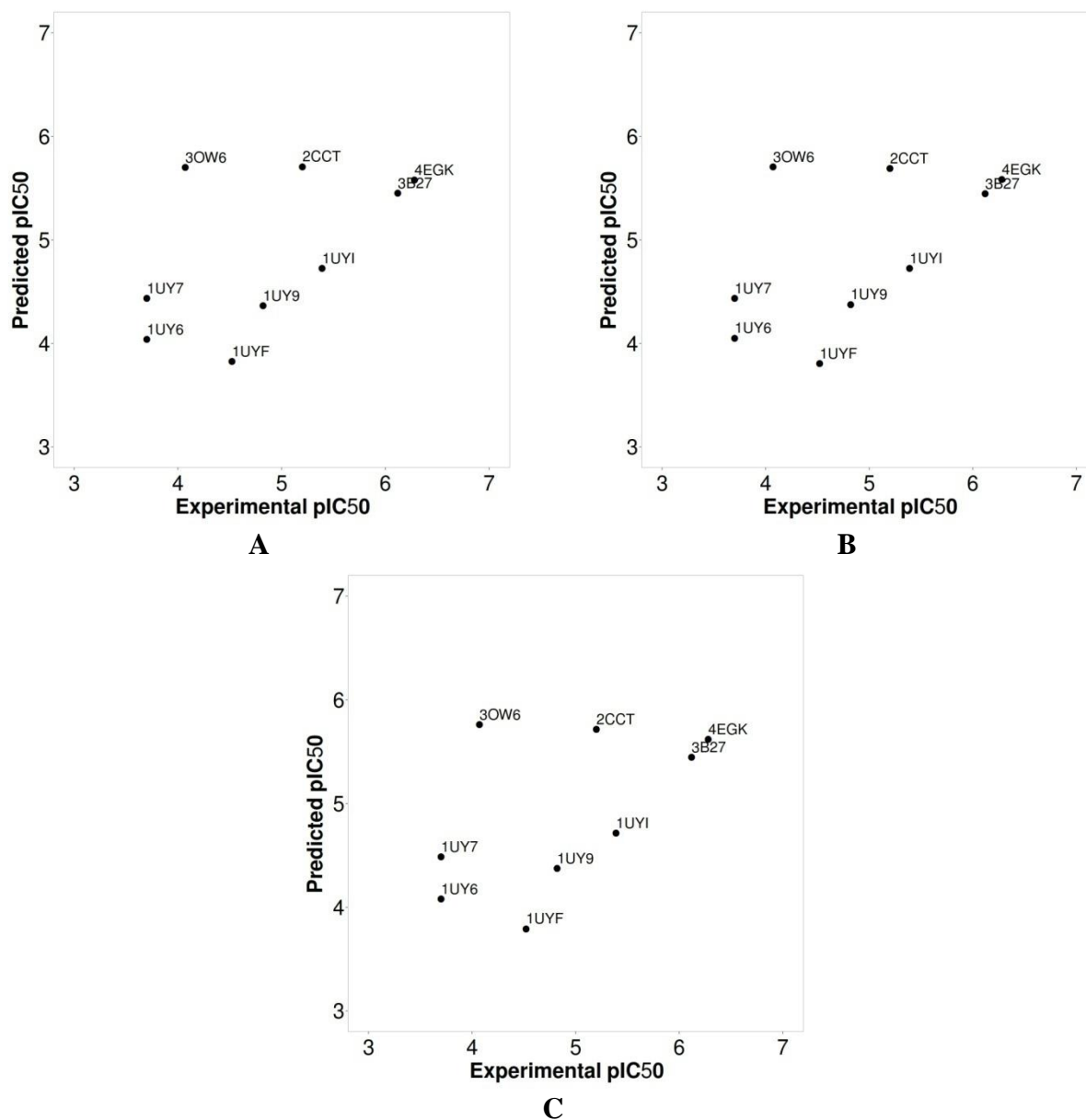
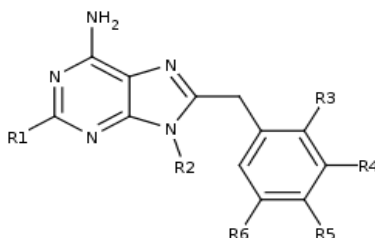


Figure 10. Experimental vs Predicted Test Set plots. A: from the A probe model at PC2; B: from the N probe model at PC2; C: from the OA probe model at PC2.

APPENDIX A

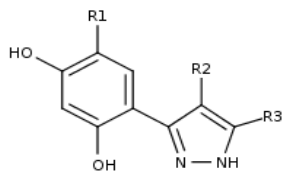
Table A1. Hsp90 inhibitors data set: (15 compounds used as training set, see also Table A2; and 9 compounds as test set, see also Table A3) shown are the purine (**A**) and pyrazole (**B**) scaffold derivatives for which ATPase activity values are known.

A) Purine¹³ scaffold derivatives



entry	PDB code	R1	R2	R3	R4	R5	R6	IC ₅₀ ATPase(μM)
1 ^a	1UY8	H	n-butyl	H	OMe	H	H	75
2 ^a	1UYC	H	n-butyl	OMe	H	H	OMe	41
3 ^a	1UYD	H	n-butyl	Cl	OMe	OMe	OMe	>200
4 ^a	1UYE	H	1-pentynyl	Cl	OMe	OMe	OMe	>200
5 ^a	1UYG	F	H	OMe	H	H	OMe	53.5
6 ^a	1UYH	F	n-butyl	OMe	H	H	OMe	14.3
7 ^a	1UYK	F	n-butyl	H	OCH ₂ O bridge		H	17.1
8 ^b	1UY6	H	n-butyl	H	OMe	OMe	OMe	>200
9 ^b	1UY7	H	n-butyl	H	H	OMe	H	>200
10 ^b	1UY9	H	n-butyl	H	OCH ₂ O bridge		H	15.3
11 ^b	1UYF							
12 ^b	1UYI	F	1-pentynyl	OMe	H	H	OMe	4.1

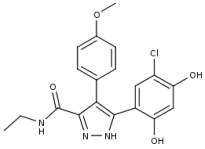
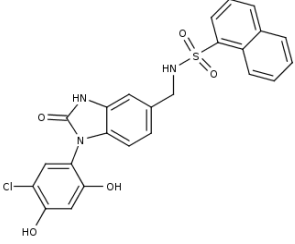
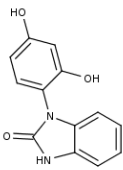
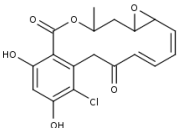
B) Pyrazole^{14, 22} scaffold derivatives



entry	PDB code	R1	R2	R3	IC ₅₀ ATPase(μM)
13 ^a	2BT0	Et		Me	5.7
14 ^a	2CCS	Cl	1-piperazine	H	8.2
15 ^b	2CCT	Cl	1-piperazine	CONHEt	6.3
16 ^a	2CCU	Cl		H	1.3

C) 2-aminopyrimidine,⁷ triazine, pyrazole and N-aryl-benzimidazolone²⁰

entry	PDB code		IC ₅₀ ATPase(μM)
17 ^a	3B25		9.6
18 ^a	3B26		6.9
19 ^b	3B27		0.75
20 ^a	3B28		3

21^a	3OWB		0.88
22^a	3OWD		0.55
23^b	3OW6		85
24^b	4EGK		0.2

^aCompounds used as training set

^bCompounds used as test set

Table A2. Training Set.

entry	PDB Code	IC₅₀ ATPase(μM)	pIC₅₀^a
1	1UY8	75	4.12
2	1UYC	41	4.39
3	1UYD	200	3.70
4	1UYE	200	3.70
5	1UYG	53.5	4.27
6	1UYH	14.3	4.84
7	1UYK	17.1	4.77
13	2BT0	5.7	5.24
14	2CCS	8.2	5.09
16	2CCU	1.3	5.89
17	3B25	9.6	5.02
18	3B26	6.9	5.16
20	3B28	3	5.52
21	3OWB	0.88	6.06
22	3OWD	0.55	6.26

$$^a\text{pIC}_{50} = -\text{Log} [\text{IC}_{50}\text{ATPase}(\mu\text{M}) \times 10^{-6}]$$

Table A3. Test Set

entry	PDB Code	IC₅₀ ATPase(μM)	pIC₅₀^a
8	1UY6	200	3.70
9	1UY7	200	3.70
10	1UY9	15.3	4.82
11	1UYF	30	4.52
12	1UYI	41	5.39
15	2CCT	6.3	5.20
19	3B27	0.75	6.12
23	3OW6	85	4.07
24	4EGK	0.2	6.28

$$^b\text{pIC}_{50} = -\text{Log} [\text{IC}_{50}\text{ATPase}(\mu\text{M}) \times 10^{-6}]$$

Bibliography

1. Buchstaller, H. P.; Eggenweiler, H. M.; Sirrenberg, C.; Gradler, U.; Musil, D.; Hoppe, E.; Zimmermann, A.; Schwartz, H.; Marz, J.; Bomke, J.; Wegener, A.; Wolf, M., Fragment-based discovery of hydroxy-indazole-carboxamides as novel small molecule inhibitors of Hsp90. *Bioorganic & medicinal chemistry letters* **2012**, *22*, 4396-403.
2. Whitesell, L.; Lindquist, S. L., HSP90 and the chaperoning of cancer. *Nature reviews. Cancer* **2005**, *5*, 761-72.
3. Trepel, J.; Mollapour, M.; Giaccone, G.; Neckers, L., Targeting the dynamic HSP90 complex in cancer. *Nature reviews. Cancer* **2010**, *10*, 537-49.
4. Liu, J. L.; Wang, F. F.; Ma, Z.; Wang, X.; Wang, Y. H., Structural Determination of Three Different Series of Compounds as Hsp90 Inhibitors Using 3D-QSAR Modeling, Molecular Docking and Molecular Dynamics Methods. *Int J Mol Sci* **2011**, *12*, 946-970.
5. Chaudhury, S.; Welch, T. R.; Blagg, B. S. J., Hsp90 as a target for drug development. *ChemMedChem* **2006**, *1*, 1331-+.
6. Den, R. B.; Lu, B., Heat shock protein 90 inhibition: rationale and clinical potential. *Ther Adv Med Oncol* *4*, 211-8.
7. Miura, T.; Fukami, T. A.; Hasegawa, K.; Ono, N.; Suda, A.; Shindo, H.; Yoon, D. O.; Kim, S. J.; Na, Y. J.; Aoki, Y.; Shimma, N.; Tsukuda, T.; Shiratori, Y., Lead generation of heat shock protein 90 inhibitors by a combination of fragment-based approach, virtual screening, and structure-based drug design. *Bioorg Med Chem Lett* *21*, 5778-83.
8. Ge, J.; Normant, E.; Porter, J. R.; Ali, J. A.; Dembski, M. S.; Gao, Y.; Georges, A. T.; Grenier, L.; Pak, R. H.; Patterson, J.; Sydor, J. R.; Tibbitts, T. T.; Tong, J. K.; Adams, J.; Palombella, V. J., Design, synthesis, and biological evaluation of hydroquinone derivatives of 17-amino-17-demethoxygeldanamycin as potent, water-soluble inhibitors of Hsp90. *J Med Chem* **2006**, *49*, 4606-15.
9. Stebbins, C. E.; Russo, A. A.; Schneider, C.; Rosen, N.; Hartl, F. U.; Pavletich, N. P., Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent. *Cell* **1997**, *89*, 239-50.
10. Schulte, T. W.; Akinaga, S.; Soga, S.; Sullivan, W.; Stensgard, B.; Toft, D.; Neckers, L. M., Antibiotic radicicol binds to the N-terminal domain of Hsp90 and shares important biologic activities with geldanamycin. *Cell stress & chaperones* **1998**, *3*, 100-8.
11. Schulte, T. W.; Neckers, L. M., The benzoquinone ansamycin 17-allylamino-17-demethoxygeldanamycin binds to HSP90 and shares important biologic activities with geldanamycin. *Cancer chemotherapy and pharmacology* **1998**, *42*, 273-9.

12. Ramanathan, R. K.; Egorin, M. J.; Erlichman, C.; Remick, S. C.; Ramalingam, S. S.; Naret, C.; Holleran, J. L.; TenEyck, C. J.; Ivy, S. P.; Belani, C. P., Phase I pharmacokinetic and pharmacodynamic study of 17-dimethylaminoethylamino-17-demethoxygeldanamycin, an inhibitor of heat-shock protein 90, in patients with advanced solid tumors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **2010**, *28*, 1520-6.
13. Wright, L.; Barril, X.; Dymock, B.; Sheridan, L.; Surgenor, A.; Beswick, M.; Drysdale, M.; Collier, A.; Massey, A.; Davies, N.; Fink, A.; Fromont, C.; Aherne, W.; Boxall, K.; Sharp, S.; Workman, P.; Hubbard, R. E., Structure-activity relationships in purine-based inhibitor binding to HSP90 isoforms. *Chem Biol* **2004**, *11*, 775-85.
14. Dymock, B. W.; Barril, X.; Brough, P. A.; Cansfield, J. E.; Massey, A.; McDonald, E.; Hubbard, R. E.; Surgenor, A.; Roughley, S. D.; Webb, P.; Workman, P.; Wright, L.; Drysdale, M. J., Novel, potent small-molecule inhibitors of the molecular chaperone Hsp90 discovered through structure-based design. *J Med Chem* **2005**, *48*, 4212-5.
15. Brough, P. A.; Aherne, W.; Barril, X.; Borgognoni, J.; Boxall, K.; Cansfield, J. E.; Cheung, K. M.; Collins, I.; Davies, N. G.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Finch, H.; Fink, A.; Hayes, A.; Howes, R.; Hubbard, R. E.; James, K.; Jordan, A. M.; Lockie, A.; Martins, V.; Massey, A.; Matthews, T. P.; McDonald, E.; Northfield, C. J.; Pearl, L. H.; Prodromou, C.; Ray, S.; Raynaud, F. I.; Roughley, S. D.; Sharp, S. Y.; Surgenor, A.; Walmsley, D. L.; Webb, P.; Wood, M.; Workman, P.; Wright, L., 4,5-diarylisoaxazole Hsp90 chaperone inhibitors: potential therapeutic agents for the treatment of cancer. *J Med Chem* **2008**, *51*, 196-218.
16. Rajan, A.; Kelly, R. J.; Trepel, J. B.; Kim, Y. S.; Alarcon, S. V.; Kummar, S.; Gutierrez, M.; Crandon, S.; Zein, W. M.; Jain, L.; Mannargudi, B.; Figg, W. D.; Houk, B. E.; Shnaidman, M.; Brega, N.; Giaccone, G., A phase I study of PF-04929113 (SNX-5422), an orally bioavailable heat shock protein 90 inhibitor, in patients with refractory solid tumor malignancies and lymphomas. *Clinical cancer research : an official journal of the American Association for Cancer Research* **17**, 6831-9.
17. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000**, *28*, 235-42.
18. Ballante, F.; Ragno, R., 3-D QSAutogrid/R: an alternative procedure to build 3-D QSAR models. Methodologies and applications. *Journal of chemical information and modeling* **2012**, *52*, 1674-85.
19. Friggeri, L.; Ballante, F.; Ragno, R.; Musmuca, I.; De Vita, D.; Manetti, F.; Biava, M.; Scipione, L.; Di Santo, R.; Costi, R.; Feroci, M.; Tortorella, S., Pharmacophore assessment through 3-D QSAR: evaluation of the predictive ability

- on new derivatives by the application on a series of antitubercular agents. *Journal of chemical information and modeling* **2013**, 53, 1463-74.
20. Bruncko, M.; Tahir, S. K.; Song, X.; Chen, J.; Ding, H.; Huth, J. R.; Jin, S.; Judge, R. A.; Madar, D. J.; Park, C. H.; Park, C. M.; Petros, A. M.; Tse, C.; Rosenberg, S. H.; Elmore, S. W., N-aryl-benzimidazolones as novel small molecule HSP90 inhibitors. *Bioorg Med Chem Lett* **20**, 7503-6.
21. Musmuca, I.; Caroli, A.; Mai, A.; Kaushik-Basu, N.; Arora, P.; Ragno, R., Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors. *J Chem Inf Model* **50**, 662-76.
22. Barril, X.; Beswick, M. C.; Collier, A.; Drysdale, M. J.; Dymock, B. W.; Fink, A.; Grant, K.; Howes, R.; Jordan, A. M.; Massey, A.; Surgenor, A.; Wayne, J.; Workman, P.; Wright, L., 4-Amino derivatives of the Hsp90 inhibitor CCT018159. *Bioorg Med Chem Lett* **2006**, 16, 2543-8.
23. Miura, T.; Fukami, T. A.; Hasegawa, K.; Ono, N.; Suda, A.; Shindo, H.; Yoon, D. O.; Kim, S. J.; Na, Y. J.; Aoki, Y.; Shimma, N.; Tsukuda, T.; Shiratori, Y., Lead generation of heat shock protein 90 inhibitors by a combination of fragment-based approach, virtual screening, and structure-based drug design. *Bioorganic & medicinal chemistry letters* **2011**, 21, 5778-83.
24. Sakkiah, S.; Thangapandian, S.; John, S.; Kwon, Y. J.; Lee, K. W., 3D QSAR pharmacophore based virtual screening and molecular docking for identification of potential HSP90 inhibitors. *European journal of medicinal chemistry* **2010**, 45, 2132-40.
25. Gopalsamy, A.; Shi, M.; Golas, J.; Vogán, E.; Jacob, J.; Johnson, M.; Lee, F.; Nilakantan, R.; Petersen, R.; Svenson, K.; Chopra, R.; Tam, M. S.; Wen, Y.; Ellingboe, J.; Arndt, K.; Boschelli, F., Discovery of benzisoxazoles as potent inhibitors of chaperone heat shock protein 90. *J Med Chem* **2008**, 51, 373-5.

Chapter VIII

Hsp90 Inhibitors (II). Combining ligand-based and structure-based approaches for Virtual Screening application

Antonia Caroli, Flavio Ballante, Richard B. Wickersham III, Federico Corelli And Rino Ragno

Journal of Chemical Information and Modeling, (2013, submitted)

Preamble

The present Chapter summarizes efforts for the discovery of hit compounds with molecular scaffolds previously untested as Hsp90 inhibitors. To select potential new Hsp90 inhibitors, three-dimensional quantitative structure-activity relationships (3-D QSAR), ligand-based (LB) and structure-based (SB) alignments methods, and a LB-SB virtual screening (LB-SB-VS) protocol were applied. Then, the NCI Diversity Set was virtually screened employing the LB-SB-VS strategy, and 80 molecules were selected for enzyme-based biological assays. Among the tested molecules, four derivatives showed IC_{50} values ranging between 18-63 μ M.

Introduction

Computer-aided Virtual (*in silico*) Screening (VS) represent a powerful technique in identifying new bioactive compounds from large chemical databases. Currently, different VS strategies can be employed on the basis of the source data: when 3-D structures are unavailable, ligand based (LB) methods, such as QSAR, 3-D QSAR and pharmacophore-based procedures are commonly used; otherwise when 3-D information is available, structure based (SB) protocols, like ligand-protein docking procedures, are mainly applied. Anyway, it should be considered that molecular docking has a significant limitation represented by the interdependence between sampling and scoring as generally implemented. Indeed, this interdependence limits the rate of success for the identification of near-native poses in virtual screening. To enhance the VS performance, another strategy is to use either SB and LB methods, as for the VS application characterizing the study herein presented, in which a combination of Autodock (SB) and Surflex-Sim (LB) scoring functions were used jointly with the predictive ability of previously built 3-D QSAR models (Chapter VII).

Methodology overview

As reported in Chapter VII, several three-dimensional quantitative-structure activity relationship (3-D QSAR) models were built and externally validated using a data set composed of 24 Hsp90 inhibitors, 15 of which were used as training set and the other 9 as test set. These models were implemented, as a predictive tool, in a virtual screening procedure (Figure 1) used to predict the inhibitory ability of 1785 compounds (NCI Diversity Set) suggesting which of these should be chosen to be biologically assayed. Considering the unavailability of 3-D informations, two different alignment procedures, a LB methodology using Surflex-sim¹ and a SB

methodology using AutoDock4,² previously validated, were adopted to predict their binding poses (see Alignment rules paragraph), in order to be subsequently tested by the 3-D QSAR models. Therefore two different sets of predicted binding conformations were used as external test sets for the 3-D QSAR models, yielding two sets of predicted pIC₅₀ values. By means of a consensus scoring procedure in which either the predicted pIC₅₀ values and the predicted free binding energy from the Autodock4 docking were used, 80 molecules were then selected for enzyme-based biological assays. Among the tested molecules, four derivatives showed IC₅₀ values in the micromolar range.

Alignment Rules:

To obtain the pose of molecules with unknown binding modes (NCI Diversity Set), either ligand based (LB) and structure based (SB) alignment procedures were applied, using Surflex-sim and AutoDock4 respectively, applying a protocol completely described in a previous work.³

Virtual Screening (VS).

By the application of the alignment protocol on the NCI Diversity Set, two different binding poses, derived from Surflex-Sim and AutoDock4, were obtained for each compound, therefore two different sets were composed and predicted by the selected 3-D QSAR models: A, N, OA and MP (as specified in Chapter VII). A rank by rank strategy⁴ was then applied, considering the predicted pIC₅₀ values together with the corresponding predicted free binding energy released by AutoDock4 to select the 80 top ranked compounds which were selected for biological investigations. From these, four molecules (NCI23128, NCI23128,

NCI17285 and NCI170578) showed IC_{50} values between 18 and 63 μM (Table 1), confirming the consistence of the applied strategy towards untested molecules.

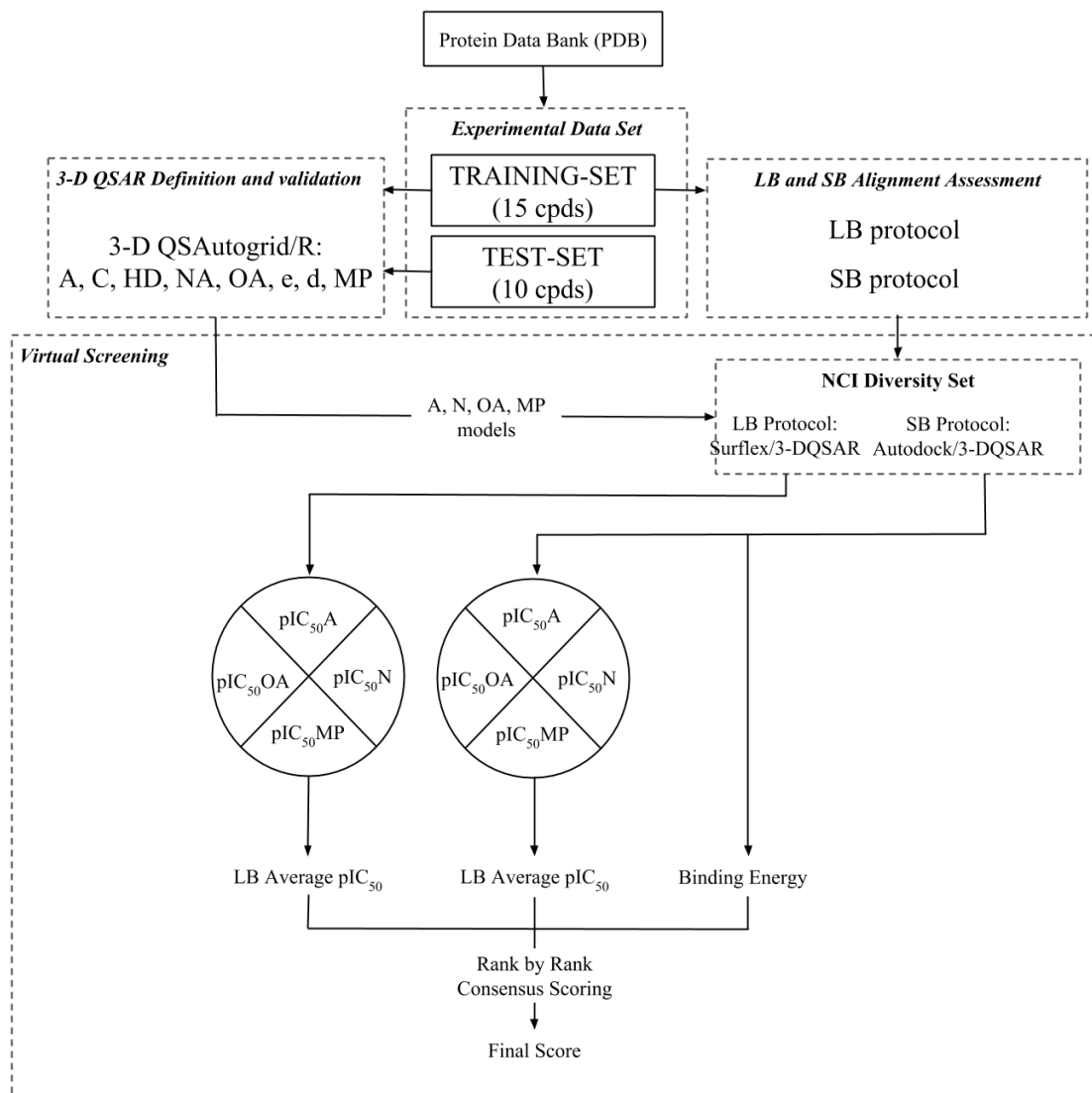
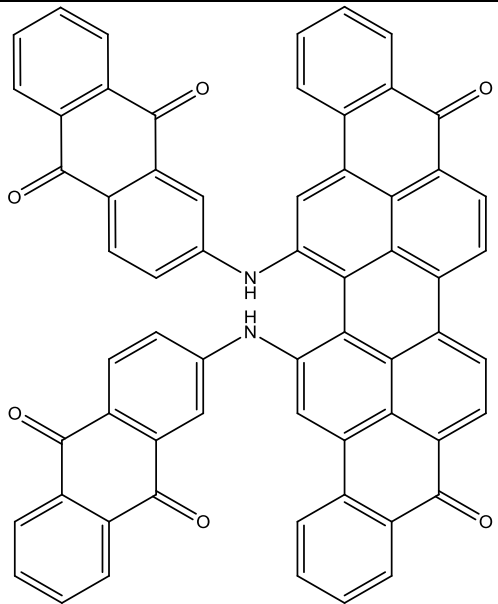
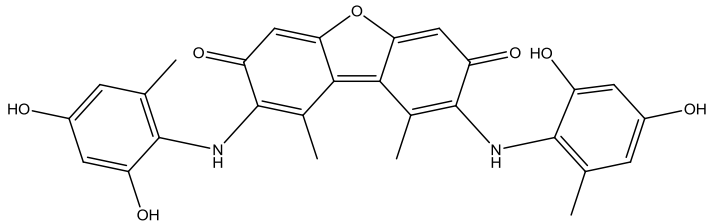
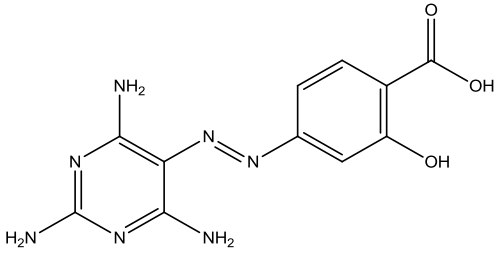
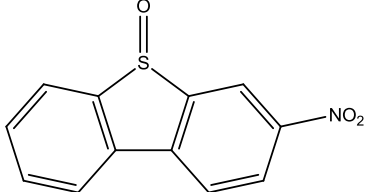


Figure 1. Computational procedure.

Table 1. Molecular structure and biological activity of the most active compounds selected by the VS protocol.

	ID	IC ₅₀ [μM]
	NCI23128	18
	NCI610930	32
	NCI117285	49
	NCI170578	63

Binding mode analysis of new HSP90 inhibitors.

Among the most active screened compounds, attention was focused on **NCI610930**, **NCI170578** and **NCI117285** endowed with interesting scaffolds for new Hsp90 inhibitors. Investigating the activity contribution plots obtained from the selected compounds, it was possible to highlight how the quantitative models (A, N, OA) predicted the interactions of these test set molecules helping to define the crucial interaction during the binding mode analysis. As shown in Figure 2, starting from the most active compound to the least active (**NCI610930** → **NCI117285** → **NCI170578**), a positive predicted activity contribution area (green surface, Figure 2), in the proximity of LYS58, ILE96, and GLY97, decreases in magnitude jointly with the biological response; while considering **NCI170578** and **NCI117285** (the least active compounds) another positive predicted activity contribution area appeared near LEU48, VAL186, THR 184, ASP93, SER 52, and LEU48. These results were a further confirmation of the importance of these two residues' series, as previously predicted by the 3-D QSAR models (Chapter VII). Moreover, **NCI610930** and **NCI170578** are respectively a dibenzofurandione and a dibenzothiophene derivatives, that could be ascribed to the tricyclic series of Hsp90 inhibitors, recently identified⁵ as a new interesting scaffold for Hsp90 inhibitors, confirming the predictive capability of the applied VS protocol.

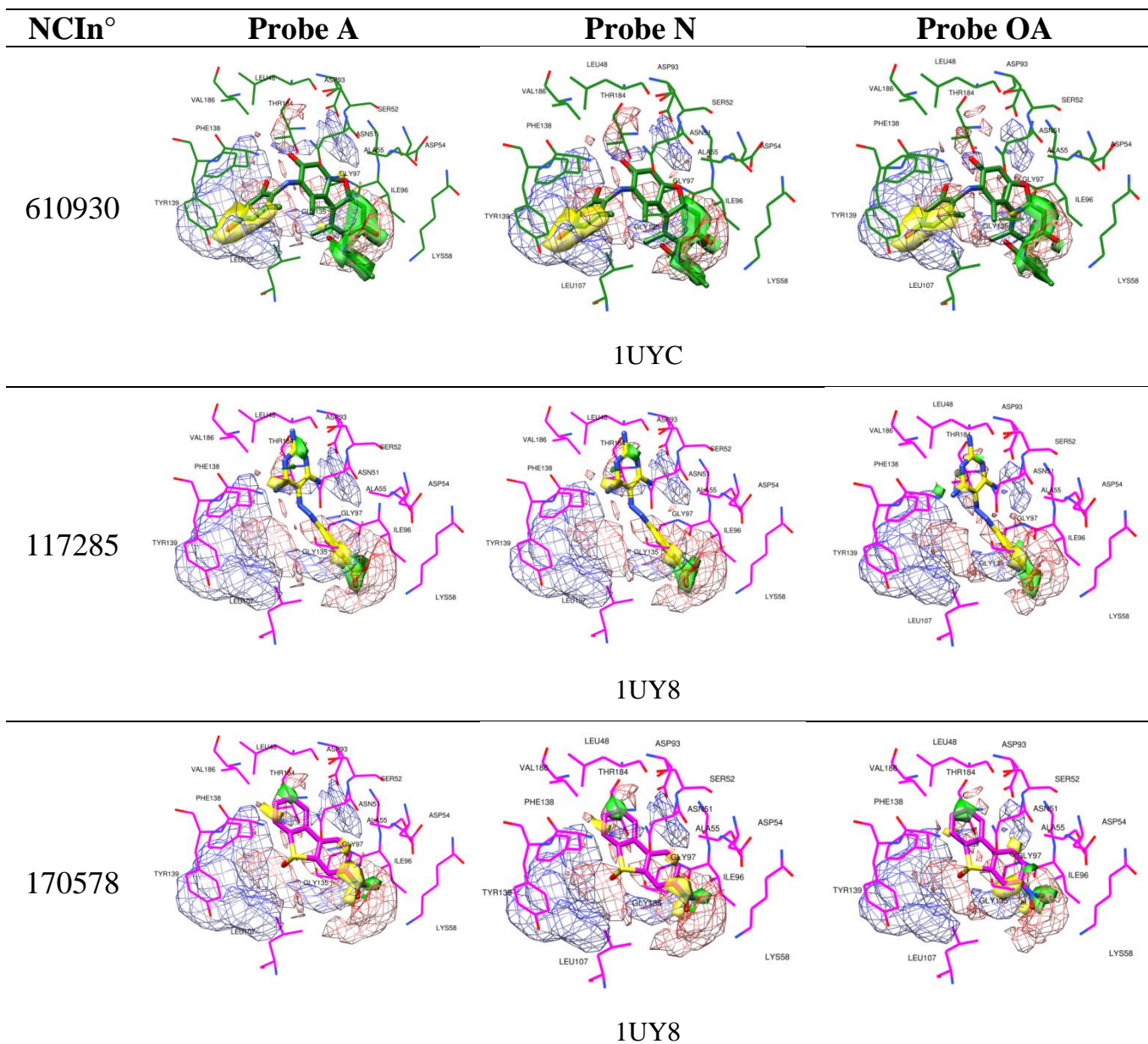


Figure 2. Predicted activity contribution plots (solid: 75%, positive: green, negative: yellow), overlapped with PLS coefficients plots (mesh: 65%, positive: red, negative: blue) obtained from the used 3-D QSAR models at the selected PC,⁶ for the most active screened compounds in their BC system (protein and pose): **NCI610930** in 1UYC, **NCI117285** and **NCI170578** in 1UY8.

Bibliography

1. Jain, A. N., Ligand-based structural hypotheses for virtual screening. *J Med Chem* **2004**, 47, 947-61.
2. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **2009**, 30, 2785-91.
3. Musmuca, I.; Caroli, A.; Mai, A.; Kaushik-Basu, N.; Arora, P.; Ragno, R., Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors. *J Chem Inf Model* **50**, 662-76.
4. Markovic, D.; Darvas, O., [Inborn errors of pentose and hexose metabolism and current methods of their discovery]. *Med Arh* **1975**, 29, 31-3.
5. Vallee, F.; Carrez, C.; Pilorge, F.; Dupuy, A.; Parent, A.; Bertin, L.; Thompson, F.; Ferrari, P.; Fassy, F.; Lambertson, A.; Thomas, A.; Arrebola, R.; Guerif, S.; Rohaut, A.; Certal, V.; Ruxer, J. M.; Gouyon, T.; Delorme, C.; Jouanen, A.; Dumas, J.; Grepin, C.; Combeau, C.; Goulaouic, H.; Dereu, N.; Mikol, V.; Mailliet, P.; Minoux, H., Tricyclic series of heat shock protein 90 (Hsp90) inhibitors part I: discovery of tricyclic imidazo[4,5-c]pyridines as potent inhibitors of the Hsp90 molecular chaperone. *J Med Chem* **2011**, 54, 7206-19.
6. Ballante, F.; Caroli, A.; Wickersham III, R. B.; Ragno, R., Hsp90 Inhibitors (I). Definition of 3-D QSAutogrid/R Models as a Tool for Virtual Screening *Journal of Chemical Information and Modeling* **2013**, Submitted.

Chapter IX

Research Period Abroad I



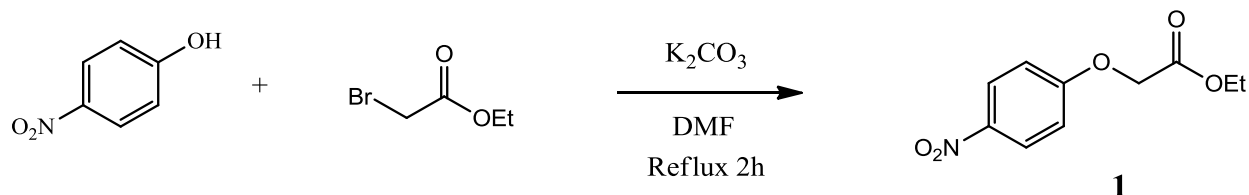
Supervisor: Prof. Gilbert Kirsch

Introduction

My first research abroad was conducted for six months (beginning: September 2012, end: March 2013) at the Laboratoire d'Ingénierie et Moléculaire Pharmacologique Biochimie (LIMBP) of the Université de Lorraine Metz (France), directed by Prof. Gilbert Kirsch, and characterized by the application of organic synthesis to obtain new thienopyrimidinone derivatives as potential inhibitors of vascular endothelial growth factor receptor-2 (VEGFR-2). These new compounds were designed as a consequence of the informations obtained from the analyses characterizing the Chapter V. In particular, among the several applied synthesis, here will be reported those concerning the 3 compounds (7a, 7b, 7c) which were finally obtained and biologically tested for their inhibitory activity against EGFR and VEGFR-2.

Synthesis of compounds 7a, 7b, 7c as new potential VEGFR-2 inhibitors

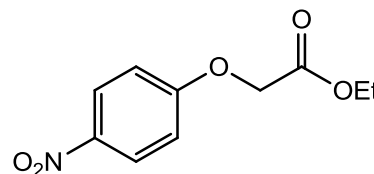
Ethyl(4-nitrophenoxy)acetate (1)



A solution of **4-nitrophenol** (150 mmol) in DMF (300 ml) was firstly obtained, then potassium carbonate (225 mmol) and **ethylbromoacetate** (165 mmol) were added and the mixture was stirred at reflux (80°C) for 2h. Then, the reaction mixture was cooled, poured over ice-water, filtered at a reduced pressure using a Buchner flask and dried *in vacuo*.

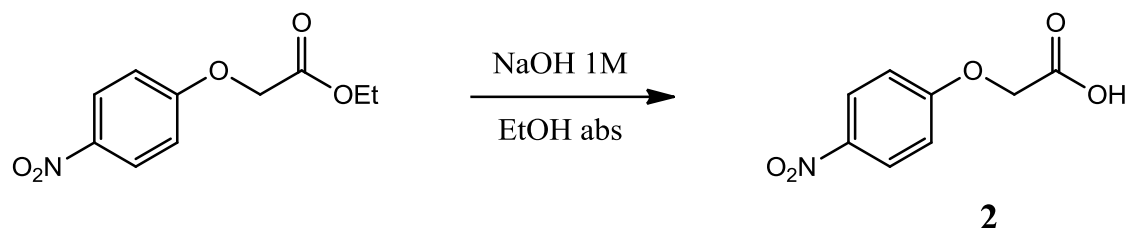
Ethyl(4-nitrophenoxy)

Molecular Weight	225.198
Aspect	Pale yellow solid
Yield	95%



¹H NMR (250MHz, DMSO): δ_H 8.18 (d, 2H), 7.17 (d, 2H), 4.98 (s, 2H, CH₂), 4.17 (q, 2H, CH₂), 1.20 (t, 3H, CH₃)

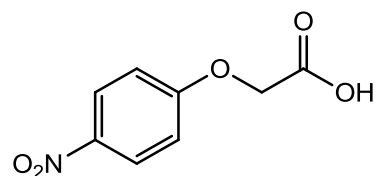
(4-Nitrophenoxy) acetic acid (2)



A solution of **Ethyl(4-nitrophenoxy)acetate** (136 mmol) in absolute ethanol (255 ml) was firstly composed, then sodium hydroxide 1M (255 mmol) was added and the mixture was stirred at reflux (85°C) for 2h. Subsequently, the reaction mixture was stirred at room temperature for 1h, poured over ice-water and acidificated with HCl 2N to adjust the pH value to 6. Finally the mixture was filtered at a reduced pressure using a Buchner flask and the solid phase was washed with water and dried on vacuum at 60°C.

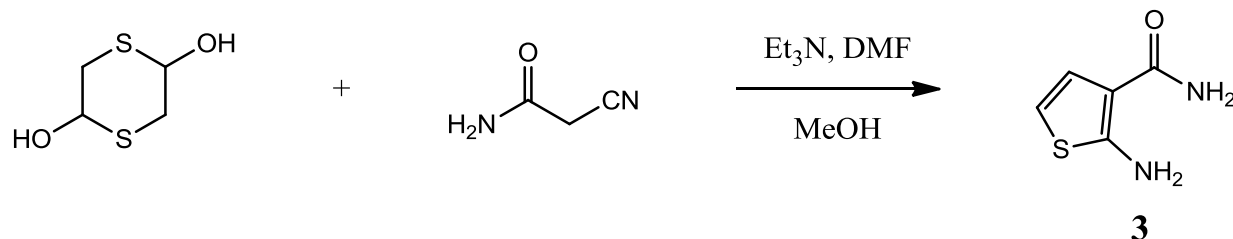
(4-Nitrophenoxy) acetic acid

Molecular Weight	197
Aspect	Yellow solid
Yield	93%



¹H NMR (250MHz, DMSO): δ_H 13.23 (s, 1H), 8.20 (d, 2H), 7.13 (d, 2H), 4.87 (s, 2H)

2-Aminothiophene-3-carboxamide (**3**)



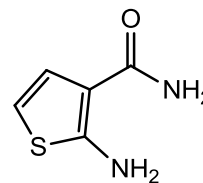
Two different procedures, characterized by different heating methods, were applied for the synthesis of 2-Aminothiophene-3-carboxamide. The first was conducted heating under reflux, the second was microwave assisted.

Procedure 1): A stirred mixture of **2-cyanoacetamide** (200 mmol), **1,4-dithiane-2,5-diol**(100 mmol) in Methanol(80 ml) and Triethylamine (10 ml) was heated at reflux (65 °C) for 3h. After cooling to room temperature, the mixture was concentrated under reduced pressure, poured over ice-water, filtered and dried *in vacuo*.

Procedure 2): A mixture of **2-cyanoacetamide** (200mmol), **1,4-dithiane-2,5-diol** (100 mmol) in Methanol(80 ml) and Triethylamine (5 ml) was heated under microwave irradiation (50 °C, 100W x 10 min). After cooling to room temperature, the mixture was concentrated under reduced pressure, poured over ice-water, filtered and dried *in vacuo*.

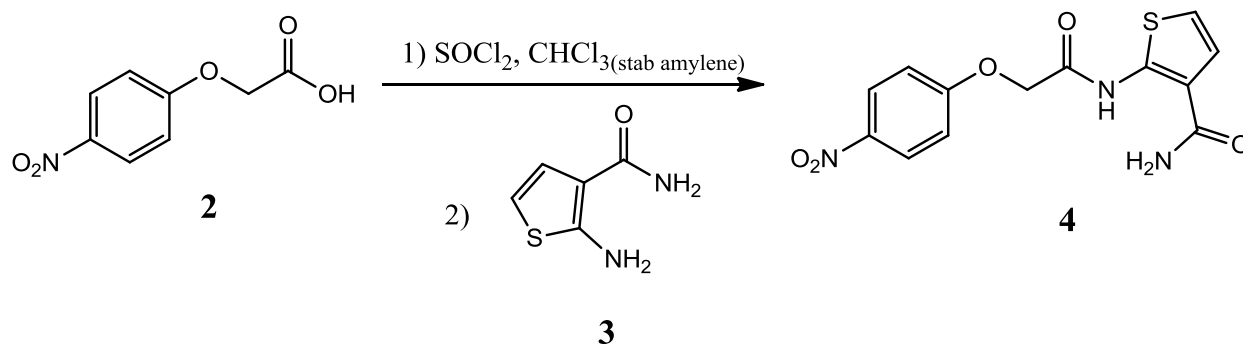
2-Aminothiophene-3-carboxamide

Molecular Weight	142
Aspect	Brown solid
Yield	Reflux: 74%
	Microwave: 79%



¹H NMR (250MHz, DMSO): δ_{H} 7.19 (s, 3H), 7.02 (d, 1H), 6.72 (bs, 1H), 6.2 (d, 1H)

2-[[4-Nitrophenoxy)acetyl]amino}-3-thiophenecarboxamide (4)



Reaction mixture 1 :

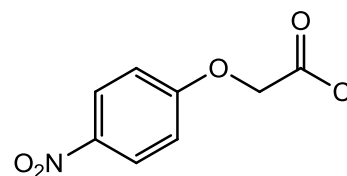
A solution of **2** (100 mmol) and amylene stabilized chloroform (added until solubilization of the acid) was firstly obtained. Then thionyl chloride 99.5+% (73 ml) was added and the mixture was stirred at reflux (60°C) for 2h. Subsequently, the reaction mixture was concentrated under reduced pressure, to obtain **(4-nitrophenoxy)acetyl chloride** (97 mmol), a red colored solid.

Reaction mixture 2:

Another solution of **3** (90 mmol) and amylene stabilized chloroform (added until solubilization of the acid) was obtained. Then Triethylamine (31 ml) was added, then mixture was stirred and cooled to 10 °C. At this point **(4-nitrophenoxy)acetyl chloride** was resolubilized in amylene stabilized chloroform and slowly added to the reaction mixture 2 at 10 °C. Following the addition, the mixture was stirred at room temperature for 1h, then poured in ice/water and filtered to obtain the final compound **4**.

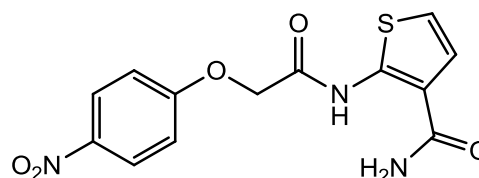
(4-nitrophenoxy)acetyl chloride

Molecular Weight	215.591
Aspect	Red solid
Yield	97%



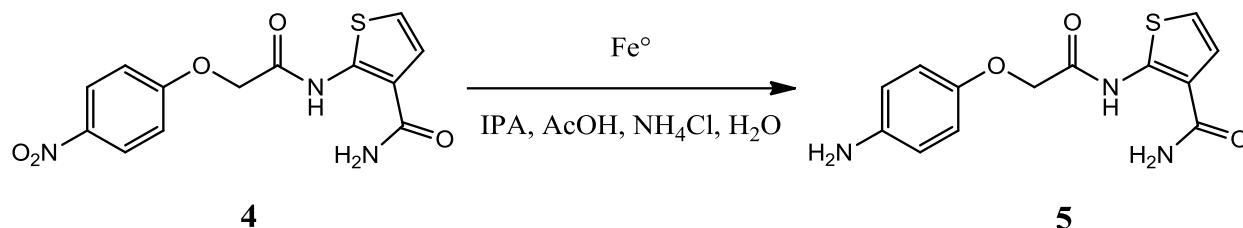
2-[[4-Nitrophenoxy)acetyl]amino]-3-thiophenecarboxamide

Molecular Weight	321.309
Aspect	Goldenrod solid
Yield	76%



¹H NMR (250MHz, DMSO): δ_H 13.00(s, 1H), 8.28(d, 2H), 7.96(s, 1H), 7.61(s,1H)
7.45(d, 1H), 7.27(d, 2H), 7.04(d, 1H), 5.06(s, 2H)

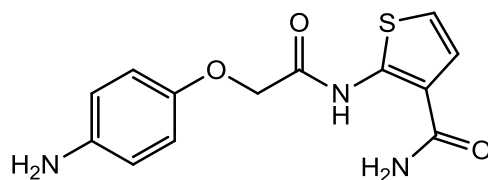
2-[[[4-aminophenyl)acetyl]amino]-3-thiophenecarboxamide (5)



A 500 ml, three-necked flask, equipped with a mechanical stirrer, thermometer, and a reflux condenser, connected to an argon line was charged with **4** (67 mmol), ammonium chloride (1.79 g), isopropanol (134 ml), water (27 ml), and acetic acid (2.2 ml). The mixture was stirred and warmed to 60°C. At this point the heat was turned off and iron powder (22.44 g) was added in small portions (temperature held steady at 60-65°C during the addition). Following the addition the mixture was refluxed for 30 min and stirred at room temperature overnight. The hot reaction mixture was diluted with 223 ml of warm ethyl acetate and filtered through a pad of Super Cel. The solids were washed with 2 x 110 ml of warm ethyl acetate. The filtrate and the washes were then extracted with 2 x 110 ml of water, dried over magnesium sulfate and concentrated.

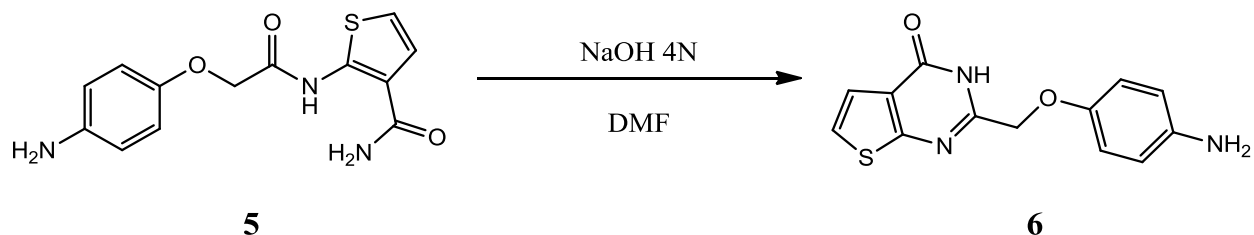
2-[[[4-aminophenyl)acetyl]amino]-3-thiophenecarboxamide

Molecular Weight	291.326
Aspect	dark grey solid
Yield	51%



¹H NMR (250MHz, DMSO): δ_{H} 12.88(s, 1H), 7.91(s, 1H), 7.53(s, 1H), 7.43(d, 2H), 7.01(d, 1H), 6.83-6.79(m, 2H), 6.55-6.51(m, 2H), 4.71(bs, 2H), 4.66(s, 2H),

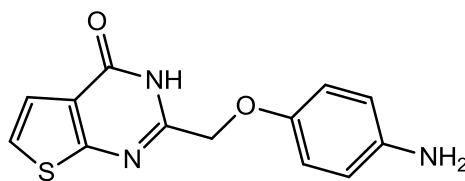
2-(4-aminophenoxyethyl)-3H-thieno[2,3-d]pyrimidin-4-one (**6**)



A stirred mixture of **5** (34 mmol), sodium hydroxide 4N (113.5 ml), Dimethylformamide (113.5 ml) was heated at reflux (100 °C) for 3h, then cooled at room temperature and diluted with water (680 ml). At this point the reaction mixture was acidificated with HCl 2N to adjust the pH value to 6 and then filtered to obtain 4.86 g (17 mmol) of filtrate. The filtrate was then extracted with ethyl acetate and the organic phase was washed with a saturated sodium chloride solution, dried with magnesium sulfate and finally filtered to obtain 0.3g (1 mmol) of **6**.

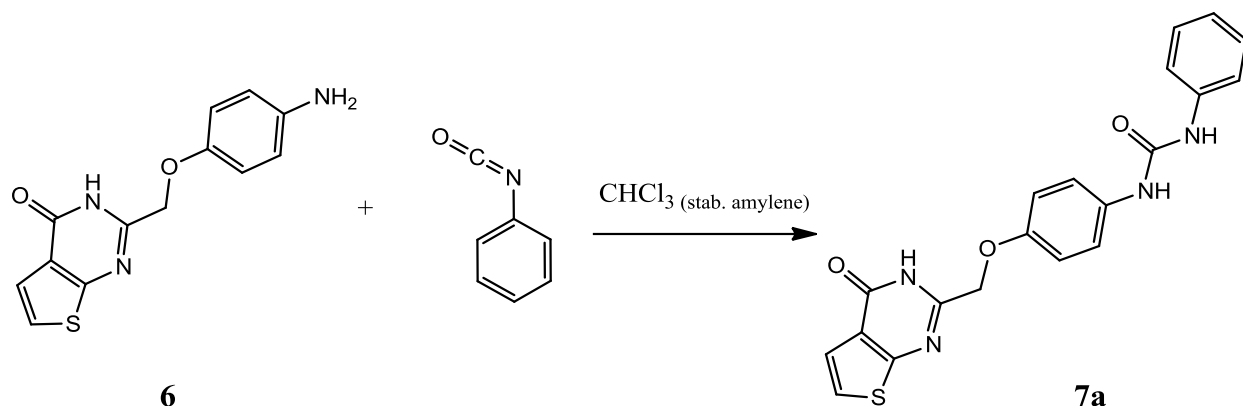
2-(4-aminophenoxyethyl)-3H-thieno[2,3-d]pyrimidin-4-one

Molecular Weight	273.310
Aspect	olive drab solid
Yield	53%



¹H NMR (400MHz, DMSO): δ_H 12.65(bs, 1H), 7.61-7.58(m, 1H), 7.40-7.39(m, 1H), 6.78-6.74(m, 2H), 6.53-6.49(m, 2H), 4.87(s,2H), 4.71(bs, 2H)

3-[4-({4-oxo-3H-thieno[2,3-d]pyrimidin-2-yl}methoxy)phenyl]-1-phenylurea (7a)



Reaction mixture 1:

2 mmol of **6** were dissolved in amylene stabilized chloroform (8 ml) to compose the first reaction mixture.

Reaction mixture 2:

2.4 mmol of phenylisocyanate were dissolved in amylene stabilized chloroform (4ml) to form the second reaction mixture

Reaction mixture 2 was added drop-wise to reaction mixture 1, the resulting mixture was stirred at room temperature overnight. Since the TLC showed an uncompleted reaction, more phenylisocyanate (1.2 mmol) was added and the mixture was stirred for 12 h, filtered and then purified by crystallization with dioxane obtaining 0.40 g of **7a**.

3-[4-({4-oxo-3H-thieno[2,3-d]pyrimidin-2-yl}methoxy)phenyl]-1-phenylurea

Molecular Weight

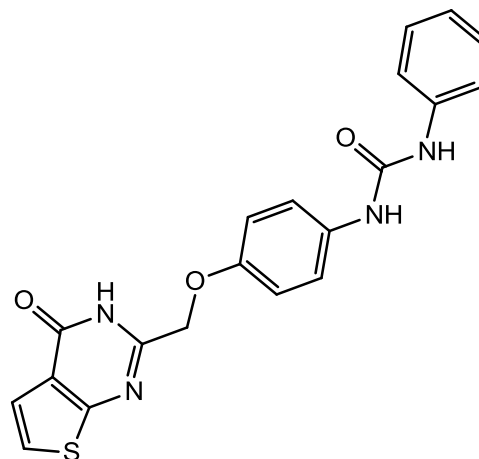
392.431

Aspect

Light grey solid

Yield

50%



$^1\text{H NMR}$ (400MHz, DMSO): δ_{H} 12.71(s, 1H), 8.61(s, 1H), 8.52(s, 1H), 7.6(d, 1H), 7.45-7.36(m, 5H), 7.27(t,2H), 6.99-6.95(m, 3H), 4.99(s, 2H)

MS (ESI, $[\text{M}+\text{Na}]^+$)
 m/z for $\text{C}_{20}\text{H}_{16}\text{N}_4\text{NaO}_3\text{S}$

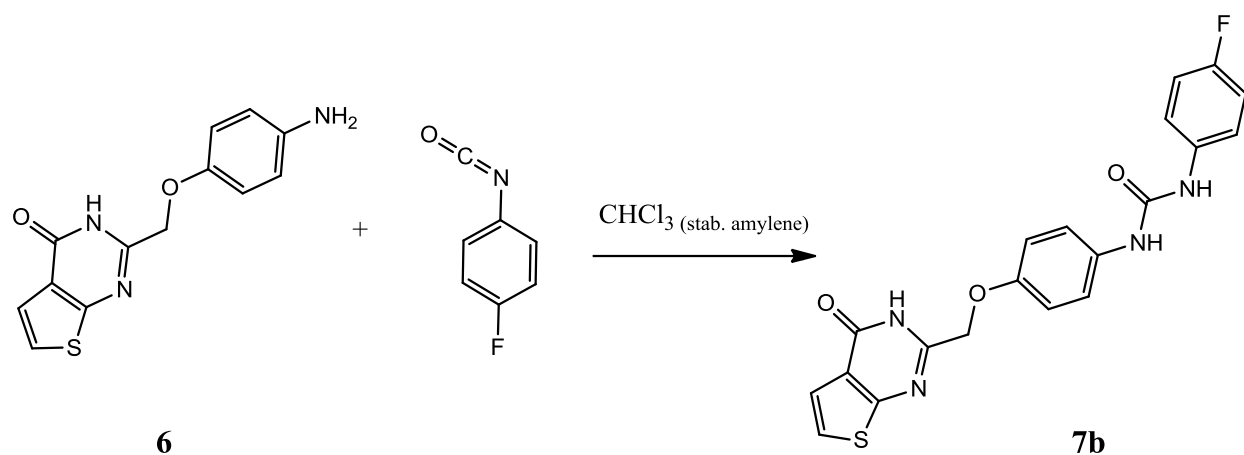
theoretic: 415.0840

obtained: 415.0834

Microanalysis:

3% of H_2O	C	H	N	S
theoretic:	58,76	4,39	13,70	7,84
obtained:	58,94	3,54	13,29	7,01

1-(4-fluorophenyl)-3-[4-({4-oxo-3H-thieno[2,3-d]pyrimidin-2-yl}methoxy)phenyl] urea (7b)

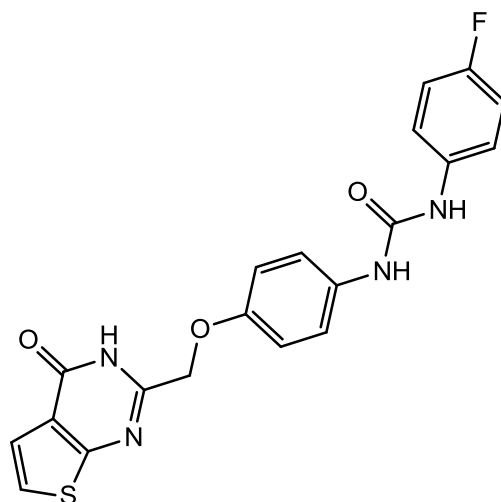


The same procedure as **7a** was applied, but in this case was not necessary to add more 4-fluoro-phenylisocyanate. After filtration, the solid was purified by crystallization in MeOH obtaining 0.33g of **7b**.

1-(4-fluorophenyl)-3-[4-({4-oxo-3H-thieno[2,3-d]pyrimidin-2-yl}methoxy)phenyl]urea

Molecular Weight 410.422
Aspect Light grey solid

Yield 40%

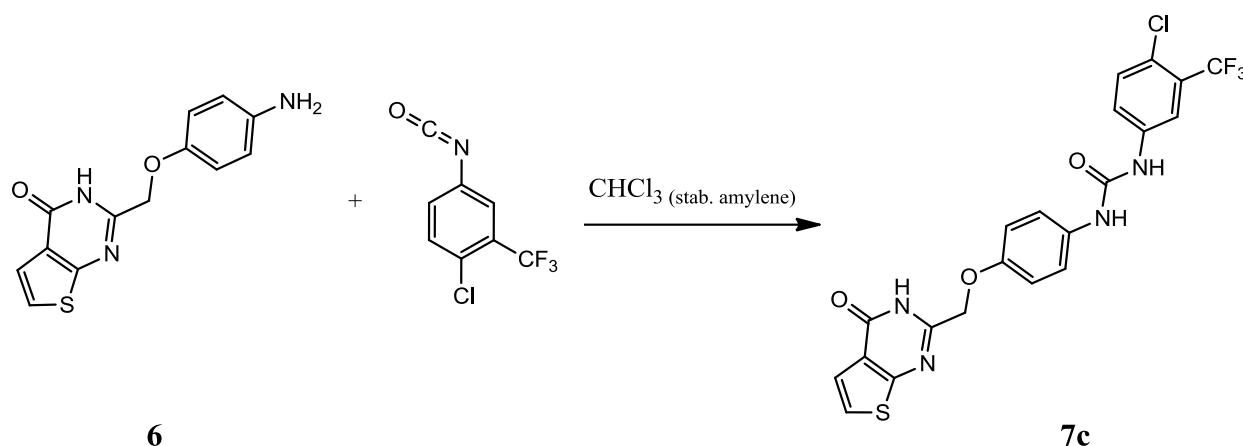


¹H NMR (400MHz, DMSO): δ_H 12.66(s, 1H), 8.64(s, 1H), 8.51(s, 1H), 7.6(d, 2H), 7.46-7.36(m, 4H), 7.1-6.97(m, 4H), 4.99(s, 2H)

MS (ESI, [M+Na]⁺) theoretic: 433.0746
m/z for
C₂₀H₁₅FN₄NaO₃S obtained: 433.0737

	3% of H ₂ O	C	H	N	S
Microanalysis:	theoretic:	56,77	3,90	13,23	8,00
	obtained:	56,97	3,59	13,39	7,62

1-[4-chloro-3-(trifluoromethyl)phenyl]-3-[4-({4-oxo-3H-thieno[2,3-d]pyrimidin-2-yl}methoxy)phenyl]urea (7c)



The same procedure as **7a** was applied, but quantities were different:

Reaction mixture 1:

0.5 mmol of **6** were dissolved in amylene stabilized chloroform (2 ml) to compose the first reaction mixture

Reaction mixture 2:

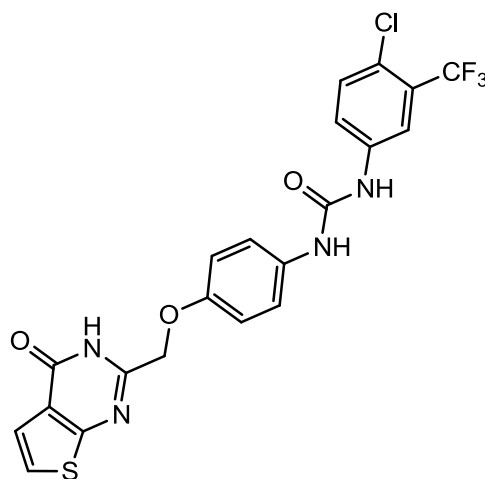
0.6 mmol of 4-Chloro-3-trifluoromethyl-phenylisocyanate were dissolved in amylene stabilized chloroform (1ml) to form the second reaction mixture

Reaction mixture 2 was added drop-wise to reaction mixture 1, the resulting mixture was stirred at room temperature overnight. Since the TLC showed an uncompleted reaction, was added triethylamine (0.245ml), but there wasn't any improvement. The mixture was then concentrated under reduced pressure, dissolved with diethylether, filtrate and the obtain solid was washed with methanol to obtain 0.13g of **7c**.

1-[4-chloro-3-(trifluoromethyl)phenyl]-3-[4-({4-oxo-3H-thieno[2,3-d]pyrimidin-2-yl}methoxy)phenyl]urea

Molecular Weight 494.874
Aspect Light grey solid

Yield 52%



^1H NMR (400MHz, DMSO): δ_{H} 12.57(bs, 1H), 9.07(s, 1H), 8.65(s, 1H), 8.09(s, 1H), 7.64-7.58(m, 3H), 7.41-7.38(m, 3H), 7.01-6.99(m, 2H), 5.00(s, 2H)

^{19}F NMR (376 MHz, DMSO): δ_{F} -61.44(s, 3F)

MS (ESI, $[\text{M}+\text{Na}]^+$) theoretical: 517.0325
 m/z for
C₂₁H₁₄ClF₃N₄NaO₃S obtained: 517.0323

		C	H	N	S
Microanalysis:	theoretic:	50.97	2.85	11.32	6.48
	obtained:	51.08	2.77	11.56	6.41

Biological evaluation for compounds 7a, 7b, 7c

Compounds **7a**, **7b** and **7c** were investigated for their inhibitor activity against either EGFR and VEGFR-2, at the Institut de Chimie Pharmaceutique Albert Lespagnol, Université Lille 2, by the “Groupe de Recherche Interdisciplinaire Innovation et Optimisation Thérapeutique EA4481” under the supervision of Pr. Patrick Depreux.

Inhibition activities were derived against EGFR (purified from A431 cell membranes) and VEGFR-2 (recombinant human protein) by quantifying the enzymatic tyrosine kinase activity, from the incorporation of radiolabeled ATP ([γ 32P]ATP) on a peptide substrate containing tyrosine residues [Poly(Glu, Tyr) 4:1] (Figure 1). Due to the possibility that other endogenous proteins may be phosphorylated, a further test was performed without PolyGluTyr and the result was subtracted from the value obtained in the presence of the screened compound. The reactions took place in a “Multiscreen® Durapore®” (Millipore™) 96-well plate. The wells were pre-wetted with 100 μ L of water, and the receptors (20 ng of EGFR and 10ng of VEGFR-2) were pre-incubated in the presence or absence of the test compound (1-dose mode: 10 μ M), previously dissolved in DMSO, for 5 minutes at 37 °C. The final DMSO concentration was 0.1%. Two different solutions (Table 1) were prepared for each enzyme and added (50 μ L) to start the reactions. After 1 hour (at 28 °C) the reactions were stopped by adding 20 μ L of trichloroacetic acid (TCA) 100% which caused precipitation lasting 30 minutes followed by the radioactivity count procedure. As mentioned above tests were accomplished with and without the substrate, in order to eliminate the contribution of endogenous proteins to the radioactivity. For each compound, inhibitory activity (%) against EGFR and VEGFR were determined (Table 2).

Table 1. Used reagents

EGFR	HEPES 50 mM pH 7,5, BSA 0,1 mg.mL ⁻¹ , MnCl ₂ 5 mM, Na ₃ VO ₄ 100 μM, DTT 0,5 mM, poly(Glu4/Tyr) 250 μg.mL ⁻¹ , ATP 5 μM, [³² P] ATP 0,5 μCi
VEGFR-2	Tris 50 mM pH 7,5, BSA 25 μg.mL ⁻¹ , MnCl ₂ 1,5 mM, MgCl ₂ 10 mM, Na ₃ VO ₄ 100 μM, DTT 2,5 mM, poly(Glu4/Tyr) 250 μg.mL ⁻¹ , ATP 5 μM, [³² P] ATP 0,5 μCi, β-glycérophosphate 5 mM.

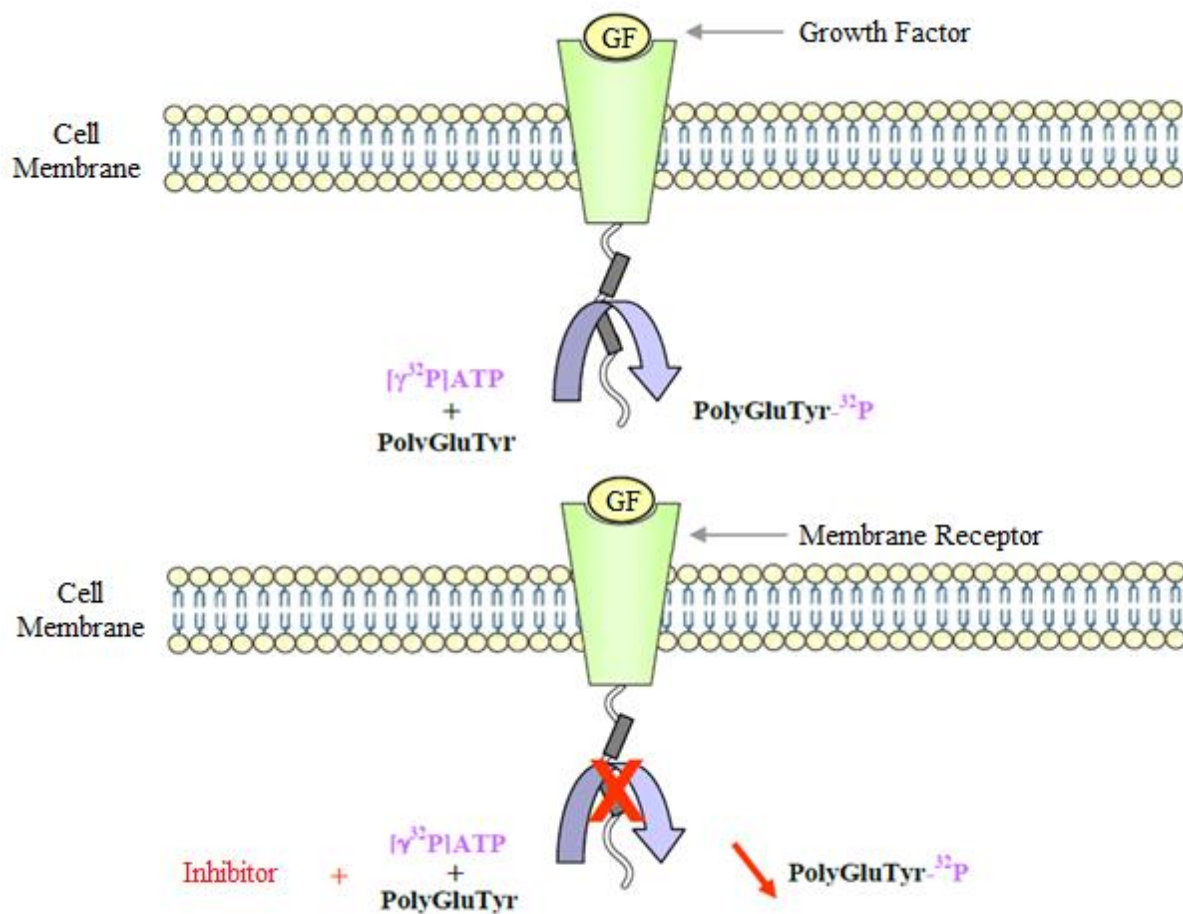
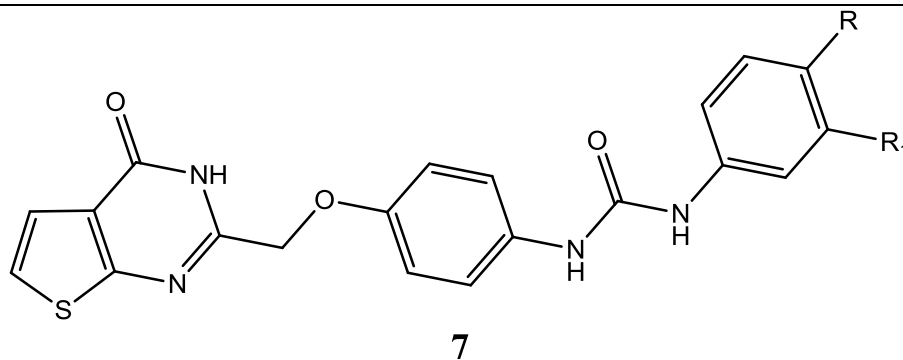


Figure 1. Assay of protein kinases using radiolabeled ATP.

Table 2. Structures and either EGFR, VEGFR-2 inhibitory activity of thieno[2,3-d]pyrimidinones **7a-7c**.



#	R	R ₁	% Inhibition @ 10 μM	
			EGFR	VEGFR-2
7a	H	H	n.a. ^a	9%
7b	F	H	n.a. ^a	5%
7c	Cl	CF ₃	n.a. ^a	13%

^a Not active. No inhibition was observed up to 10 μM of the tested compound.

Chapter X

Research Period Abroad II



Supervisor Prof. Garland R. Marshall

Introduction

My second research activity abroad, was conducted for three months (beginning: August 2013, end: October 2013) at the Marshall Lab., directed by Prof. Garland R. Marshall, in the Department of Biochemistry and Molecular Biophysics of Washington University School of Medicine in St. Louis (MO, USA) to investigate the activity profile of new Histone Deacetylases (HDACs) inhibitors.

HDACs activity investigation

In vitro determinations were conducted applying the Electrophoretic mobility shift assay (EMSA)¹ by using the LabChip[®] EZReaderII² instrument (Figure 1, Caliper-Perkin Elmer[®]).



Figure 1. The LabChip[®] EZReaderII.

Fundamentally, three main objectives were accomplished:

- a) Training on the EZReaderII instrument
- b) Define the Enzymatic assays protocol for HDACs
- c) Start the inhibitor titrations HDAC3 and HDAC6 isoforms

LabChip[®] EZReaderII:

The LabChip[®] EZReaderII combines the basic principles of capillary electrophoresis in a microfluidic environment to analyze enzymatic assays with or without the addition of a termination or quenching reagent. The core of the instrument is the so-called “Chip” (Figure 2), which is characterized by:

- 6 Upstream and 2 Downstream Electrode Wells (Figure 2A);
- 12 Sippers (Figure 2B);
- a detection window (Figure 2C)
- a network of miniaturized channels (like those characterizing the detection window, Figure 2C) through which fluids and chemicals are moved to perform experiments;

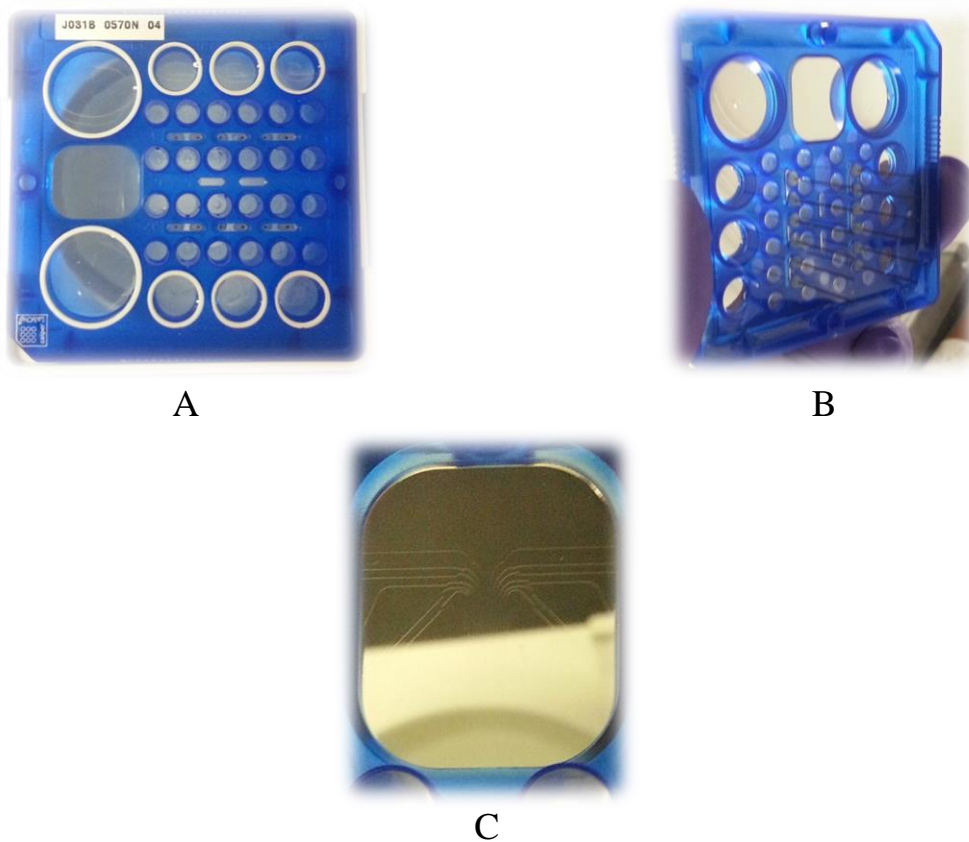


Figure 2. The LabChip[®] EZReaderII Chip.

As shown in Figure 3, the whole analysis process can be summarized in the following steps:

- 1) using vacuum pressure, reactions are sipped, from the sample wells through fused silica sippers located in the bottom of the chip;
- 2) both electrokinetics and pressure, are exerted on the chip to generate fluid motion through the microchannels. By applying an electric-potential difference across the separation channel, fluorescently labeled substrates and products are separated by electrophoresis and detected by LED-induced fluorescence;
- 3) both the substrate and the formed product are detected and measured for each sample. The amount of product is determined as $\frac{\text{product peak heights}}{\text{product} + \text{substrate peaks heights}}$.

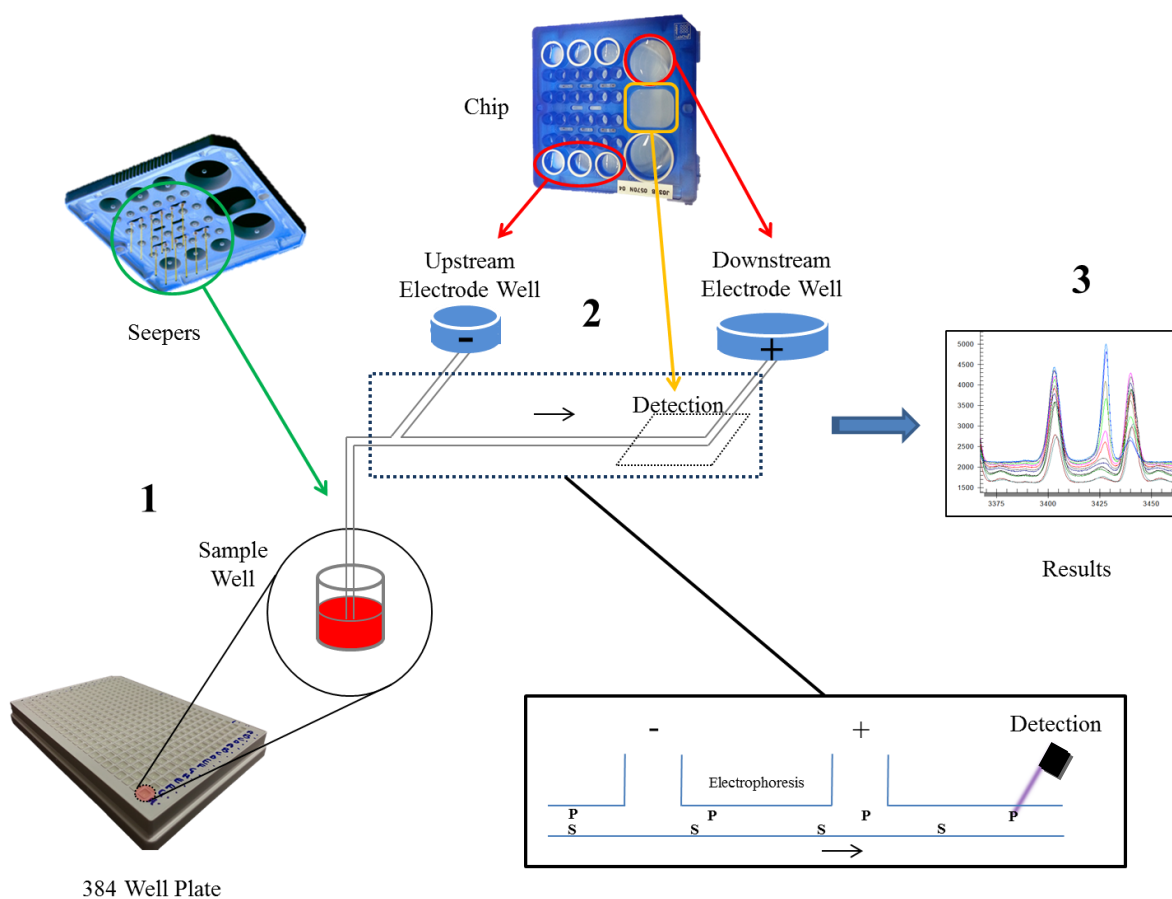


Figure 3. Overview of the Mobility Shift Assay performed by the LabChip[®] EZReaderII.

Materials and Methods

Screened Compounds:

Several enzymatic assays were performed to investigate the HDAC 3 and 6 inhibitor activities (IC₅₀ values), of a new molecular set (Table 1) characterized by chemo-physical properties ascribable to anti-HDAC activity, as follows :

- ✓ new largazole derivatives, synthesized in the Marshall lab.,
- ✓ new compounds synthesized in the laboratory supervised by Prof. Antonello Mai, (Department Of Chemistry and Drug Technologies, Faculty of Pharmacy and Medicine – Sapienza Università di Roma).

Moreover, three stock compounds: SAHA, Tubastatin A and Entinostat, from Sellekchem[®] tubes, were used as standard compounds.

Table 1. List of screened compounds.

#	Cpd	Stock Solution mM in 100% DMSO	#	Cpd	Stock Solution mM in 100% DMSO
1	MC1716	10	16	MC2776	10
2	MC1723	10	17	MC2780	10
3	MC1739	10	18	MC2984	10
4	MC1742	10	19	MC3004	20
5	MC1746	10	20	MC3031	10
6	MC1862	10	21	MC3050	10
7	MC2122	20	22	MC3079	10
8	MC2126	10	23	SD-L-148	20
9	MC2129	10	24	SD-L-256	10
10	MC2195	20	25	SDM141	20
11	MC2427	10	26	SDM146	20
12	MC2625	10	27	ENTINOSTAT ^a	50
13	MC2664	10	28	SAHA ^a	100
14	MC2726	10	29	TUBASTATIN A ^a	100
15	MC2727	10			

^aStock compounds: from Sellekchem[®] Tubes

The synthesized compounds (originally as powder) were solubilized in 100% DMSO to get the final micromolar concentrations (mM, Table 1), moreover different solutions, required for the enzymatic assays, were prepared:

1. Separation Buffer
2. Reaction Buffer
3. Dye Marker
4. Substrate Mix
5. Enzyme Mix

Overall, the “relationship” between the instrument and the necessary components can be represented like the chart in Figure 4:

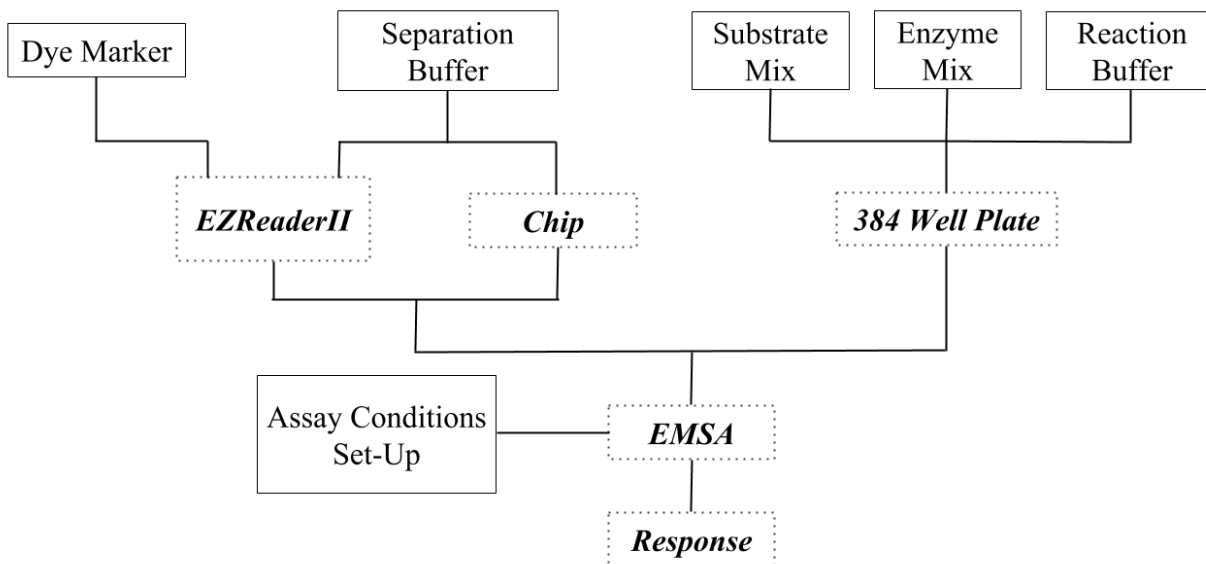


Figure 4. The hardware and chemical protagonists characterizing the assays.

Separation Buffer:

The separation buffer is the solution that runs through the machine and the chip. It is used mainly to preserve the separation conditions and prevent peptide sticking in the microfluidic chip; since it is stable at room temperature for seven days, a fresh sterilized solution (at least 200 mL, Figure 5) was primed into the EZReaderII every week.

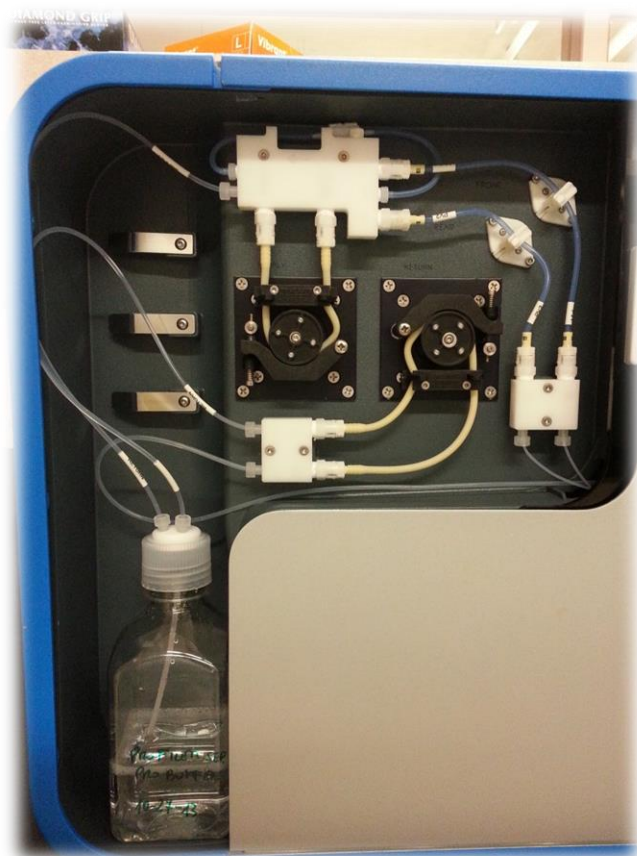


Figure 5. Recirculation system for the separation buffer.

Differently, the chip required much more attention, needing to be “prepared” (firstly washed with milliQ water and separation buffer, then re-charged with the latter) every week or after 40 hours of run time.

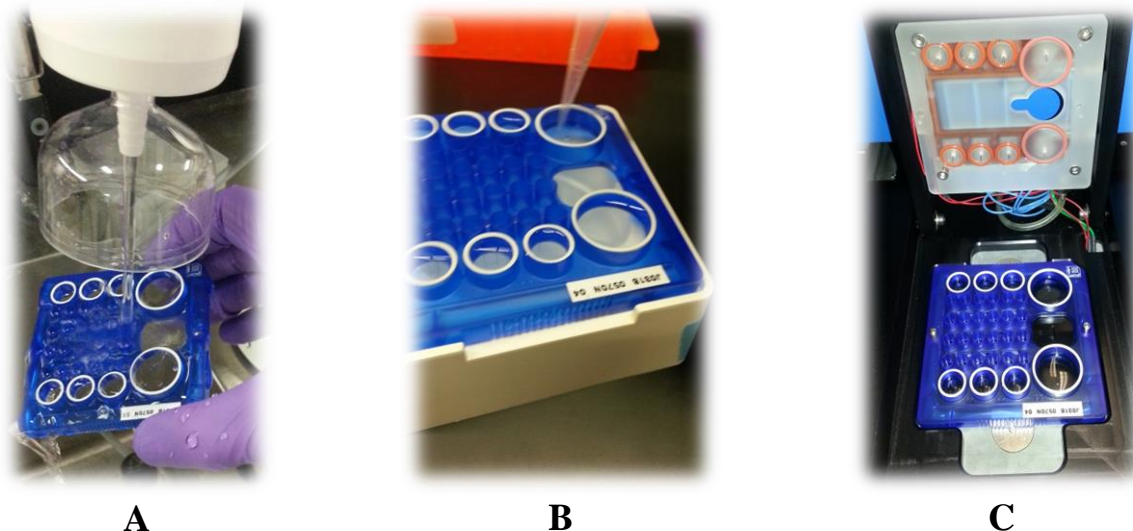


Figure 6. Chip preparation: every week or after 40 hours of run time it needs to be completely washed with milliQ water (A) and separation buffer (B), then dried, charged with new separation buffer and finally installed in the EZReaderII instrument (C).

The separation buffer (10mM EDTA) was obtained following the company guidelines, by merging a pre-separation buffer (1mM EDTA) with a solution of EDTA 0.5M (pH 8), whose compositions are described in Table 2:

Table 2. Separation Buffer. The solution was finally sterilized using a 0.2 μm pore filter membrane.

Final solution	Intermediate solution	Composition
Separation Buffer (10mM EDTA)	Pre-separation buffer	<ul style="list-style-type: none"> • 100mM HEPES, pH 7.3 • 0.015% Brij-35 • 1mM EDTA • 0.1% CR-3 • 5% DMSO
	EDTA 0.5M	<ul style="list-style-type: none"> • milliQ water • EDTA (powder) • pH adjusted to 8.0 using NaOH (pellets)

Reaction Buffer (RxnB):

The reaction buffer (RxnB) represents the solution to be added into the plate wells together with the reagents. It was prepared following the company guidelines (Table 3).

Table 3. Reaction buffer. The solution was finally sterilized using a 0.2 μm pore filter membrane.

Final solution	Intermediate solution	Composition
Reaction Buffer 0.01 % w/v BSA ^a	Base buffer	<ul style="list-style-type: none">• 25mM Tris-HCl pH 8.0• 137mM NaCl• 2.7 mM KCl• 1 mM MgCl₂
	BSA ^a 1% w/v	<ul style="list-style-type: none">• milliQ water• BSA^a powder

^aBSA: Bovine Serum Albumine

Dye Marker:

The Dye marker is the 1X peptide solution used to create the plate row markers during the assay: specifically the H219 peptide (Table 4) was used to accomplish this role.

Substrate Mix:

The substrate mix is the solution containing the peptide to be deacetylated. For the assays, H218 and H219 peptides (Table 4) were used in the case of HDAC3 and HDAC6 respectively. As reported above, H219 was utilized also as the dye marker.

Table 4. The used substrates: H218 and H219 peptides.

H218	H219
5-FITC-AHA-TSPQPKK(Ac)-CONH ₂ <ul style="list-style-type: none">• Derived from p53• 1.5 mM in 100% DMSO	5-FITC-AHA-LGKGGAK(Ac) -CONH ₂ <ul style="list-style-type: none">• Derived FROM Histone 4• 1.5 mM 100% DMSO

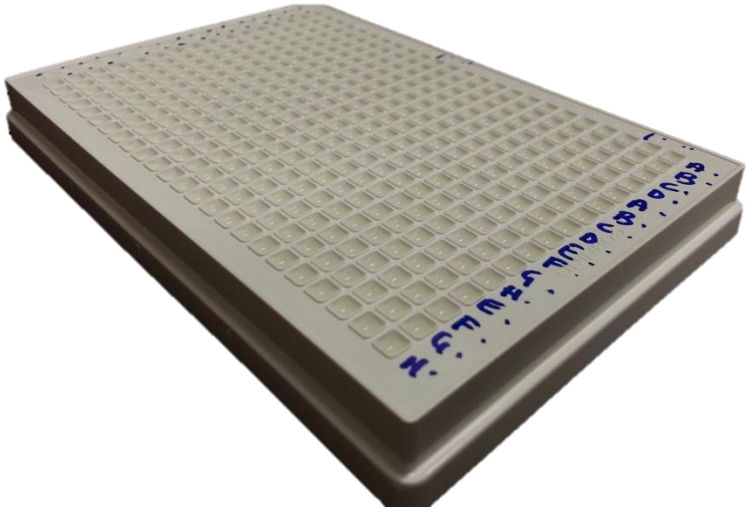
Enzyme Mix:

This is the solution containing the enzyme (HDAC) and the RxnB.

Assay plate:

For all the assays, Corning® 384 assay plates (Figure 7) were utilized.

Figure 7. Used assay plates.

Corning® (#3673)	Specifics
	<ul style="list-style-type: none">• 384 assay plate• Non-binding surface• Round bottom• White polystyrene• Working volume: 5-40 µL

Enzyme assays workflow:

The applied workflow involved, firstly, the enzyme titration step, followed by the enzyme inhibitor titration.

Enzyme titration:

Enzyme titration (Figure 8) is necessary to:

- ✓ determine the enzyme concentration, for a 30% substrate conversion, to be used during the inhibitory titration;
- ✓ establish the assay parameters as: Pressure, Downstream Voltage, Upstream Voltage, Buffer Sip, Sample/Marker Sip Time, Final Delay, Peak Order, Cycles

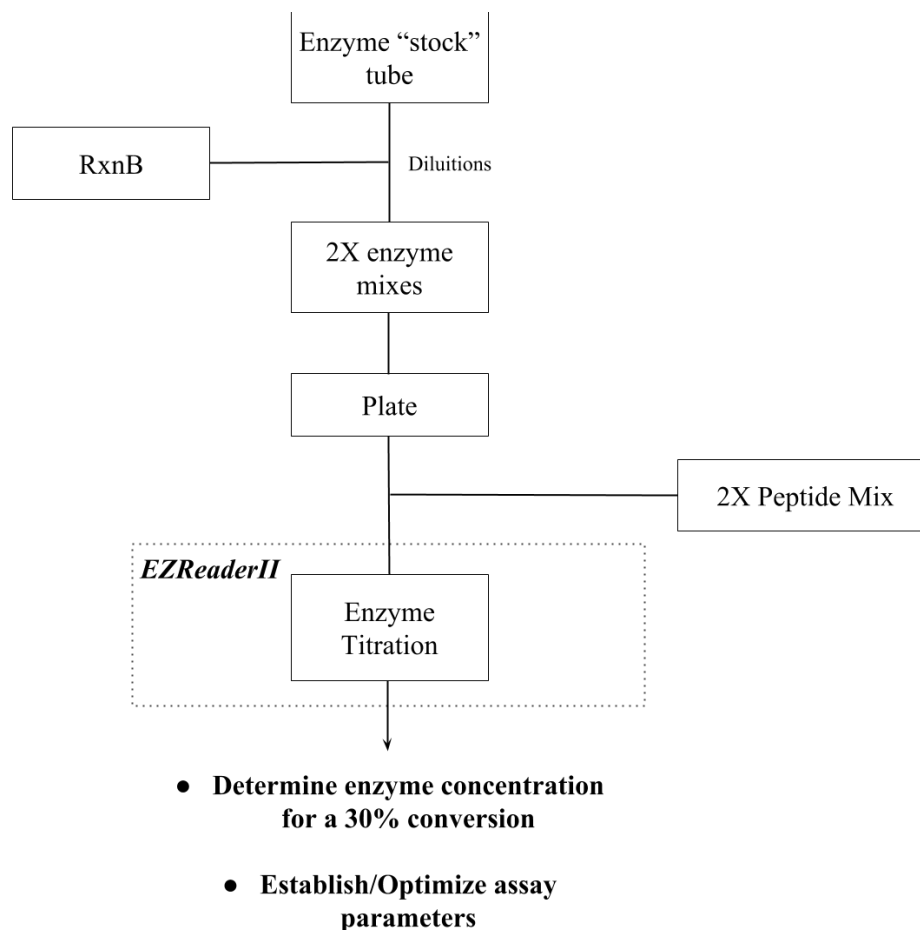


Figure 8. Enzyme titration workflow.

Enzyme titration was performed in 6-dose mode with 3-fold serial dilution, using 2-replicates; since HDAC 3 and HDAC 6 are characterized by different activities, different ranges were used (Figure 9). Initially, different 2X enzyme mix solutions were obtained by diluting the stock enzyme solution with RxnB and placed in plate wells, then an equal volume of 2X peptide mix was added to each well to reach the final (1X) desired concentration and start the reaction (deacetylation).

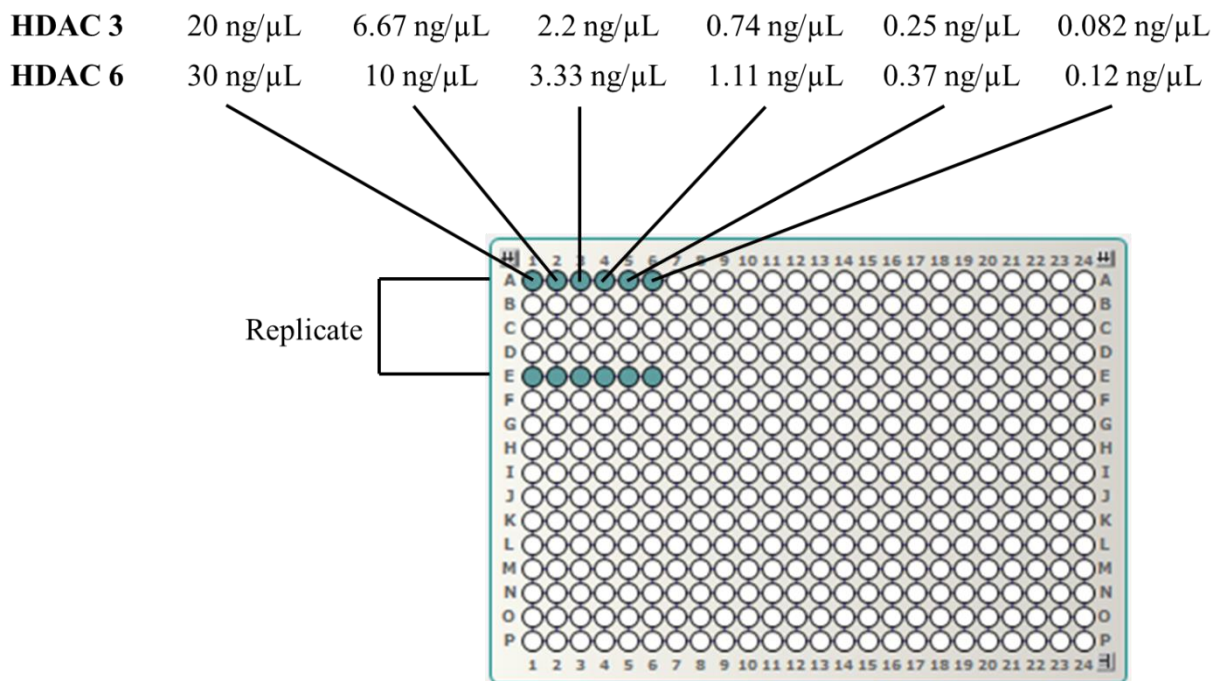


Figure 9. Enzyme titration was conducted in 6-dose mode with 3-fold serial dilution, starting from 20 ng/μL and 30 ng/μL for HDAC 3 and HDAC 6 respectively.

Then, the plate was placed in the LabChip[®] EZReaderII instrument, previously “charged” with the dye marker, to start a 45 cycles detection (\approx 1 hour), following the reaction in a real-time mode (example in Figure 10) and finally obtain conversion rates (example in Figure 11). The whole procedure was optimized in order to use for each well a total volume equal to 20 μL (10 μL of 2X enzyme mix + 10 μL of 2X peptide mix).

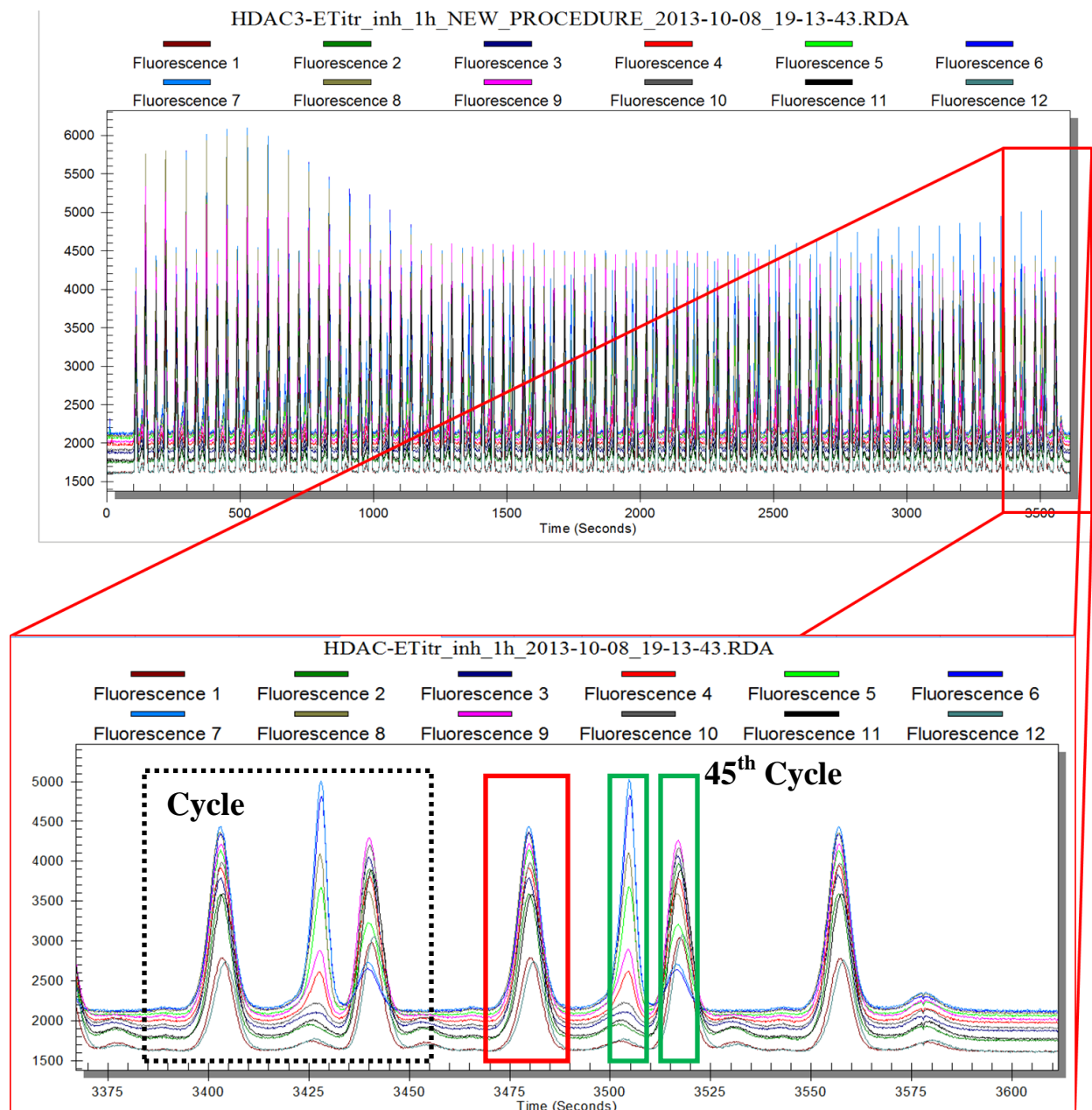


Figure 10. Enzyme titration outputs. The reaction is monitored, as it progresses, by sequentially “sipping” samples onto the chip at various time intervals (cycles): removal of acetyl groups from peptide substrates produces distinct mobility-shifts. Three different peak types are detectable for each cycle: the marker peaks (red contoured), the product peaks (green contoured) and the substrate peaks (yellow contoured). Since the chip is characterized by 12 sippers, 12 different wells (6 for each row) were monitored at the same time: each color represents data from a different reaction well.

1	1	2	3	4	5	6
▶ A	82.9	56.4	27.1	10.9	8.47	8.49
B						
C						
D						
▶ E	83.4	57.9	25.2	8.49	7.05	7.41
F						
G						

Figure 11. Enzyme titration results expressed as substrate conversion rates. As for example are shown the results obtained from one HDAC3 titration (Enzyme from Sigma-Aldrich™, Product N.SRP0104, Lot N.5000320524, specific activity ≥ 3000 pmol/min/ μ g). In this specific case the characterizing enzyme concentration of well n. 3 (2.2 ng/ μ L, compare with Figure 9) was used for the inhibitor titration, since it was the closest to determine a conversion rate equal to 30%.

Once determined the enzymatic concentration capable to convert (deacetylate) a substrate quantity close to 30%, it was possible to proceed for the inhibitor titration.

Inhibitor titration:

A total of 32 titrations for HDAC 3 and 24 titrations for HDAC 6 were performed, following the screening scheme specified in Table 5. Compounds were tested in 10-dose IC₅₀ mode with 3-fold serial dilution starting from 30 μ M. For each replicate, well plates n. 11 (23) and 12 (24) were used as 0% inhibition and 100% inhibition controls, respectively (Figure 12). Each well was charged with a total of 15 μ L of reactions following the sequence shown in Figure 13. After 1 hour of incubation reactions were stopped by adding, to each well, 15 μ L of termination buffer (reaction buffer + EDTA 10 mM), then the plate was introduced in the instrument to start the titration and derive conversion and inhibition rates (as for example Figures 14 and 15 show both the outputs for the 3rd HDAC 3 screening).

Conversion rates were calculated considering the $\frac{\text{product peak heights}}{\text{product} + \text{substrate peaks heights}}$ ratio, whereas inhibition rates were obtained considering the % of controls (averaging the results across sippers).

Table 5. Screening scheme.

CPD	ID	Plate Row	Screening N.	Used Enzyme
MC2984	A	A, E	1	HDAC3
SDM 141	B	B, F		
SDM 146	C	C, G		
SD-L-256	D	D, H		
SD-L-148 (Largazole)	E	I, M		HDAC6
SAHA ^a (Vorinostat)	F	J, N		
TUBASTATIN A ^a	G	K, O		
ENTINOSTAT ^a	H	L, P		
MC2727	A	A, E	2	HDAC3
MC2726	B	B, F		
MC2625	C	C, G		
MC2664	D	D, H		
MC2780	E	I, M		HDAC6
MC2776	F	J, N		
MC3079	G	K, O		
SAHA (Vorinostat) ^a	H	L, P		
MC3031	A	A, E	3	HDAC3
MC3004	B	B, F		
MC3050	C	C, G		
MC1742	D	D, H		
MC1862	E	I, M		HDAC6
MC2126	F	J, N		
MC2129	G	K, O		
SAHA (Vorinostat) ^a	H	L, P		
MC1716	A	A, E	4	HDAC3
MC1723	B	B, F		
MC1746	C	C, G		
MC1739	D	D, H		
MC2122	E	I, M		
MC2427	F	J, N		
MC2195	G	K, O		
SAHA (Vorinostat) ^a	H	L, P		

^aSelleckchem Tubes

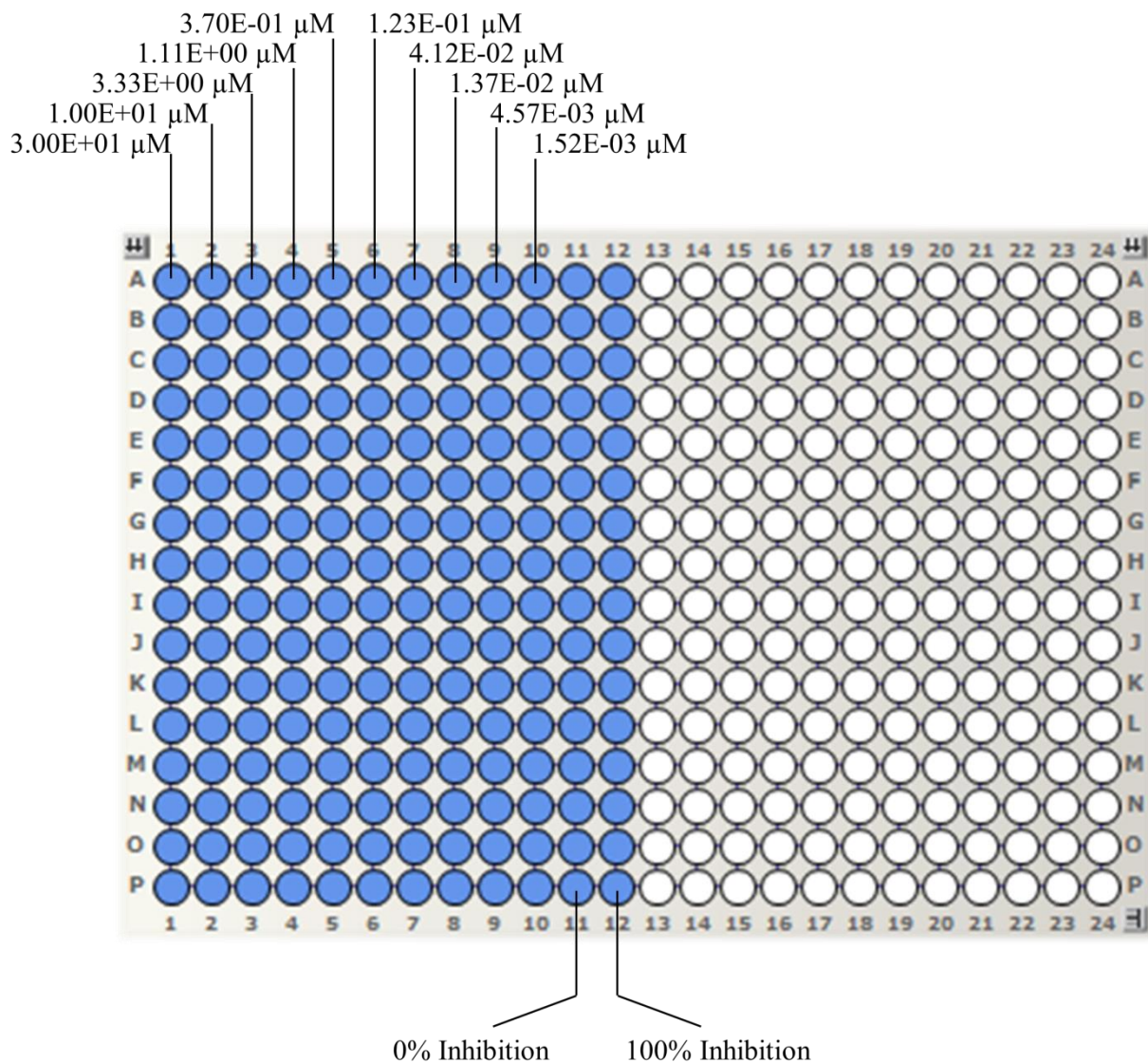


Figure 12. Plate scheme applied during the inhibitor titration.

Inhibition rates allowed to derive, for each screened compound, the characteristic inhibitory dose-response curve (Figure 16 and 19), the characteristic IC_{50} value (Table 6 and 7) and the relative inhibitory potency (Figures 17,18,20,21).

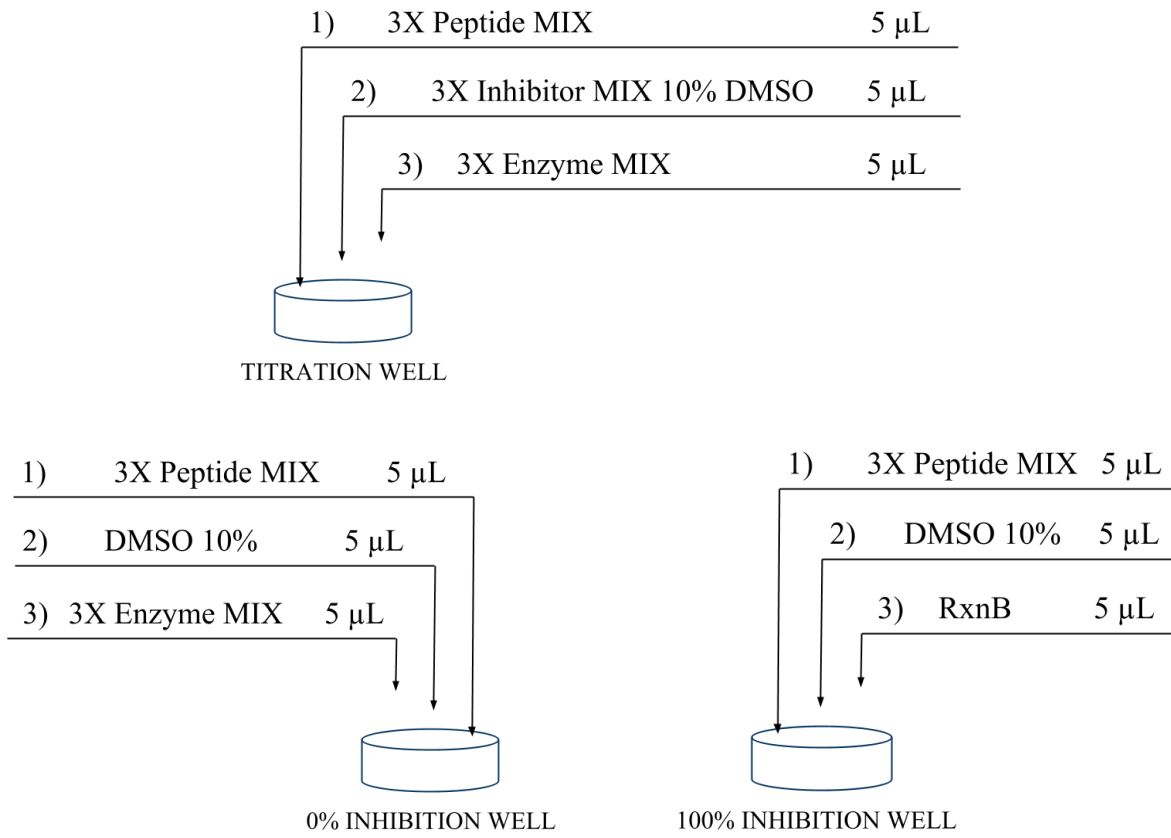


Figure 13. Protocol adopted to fill the wells.

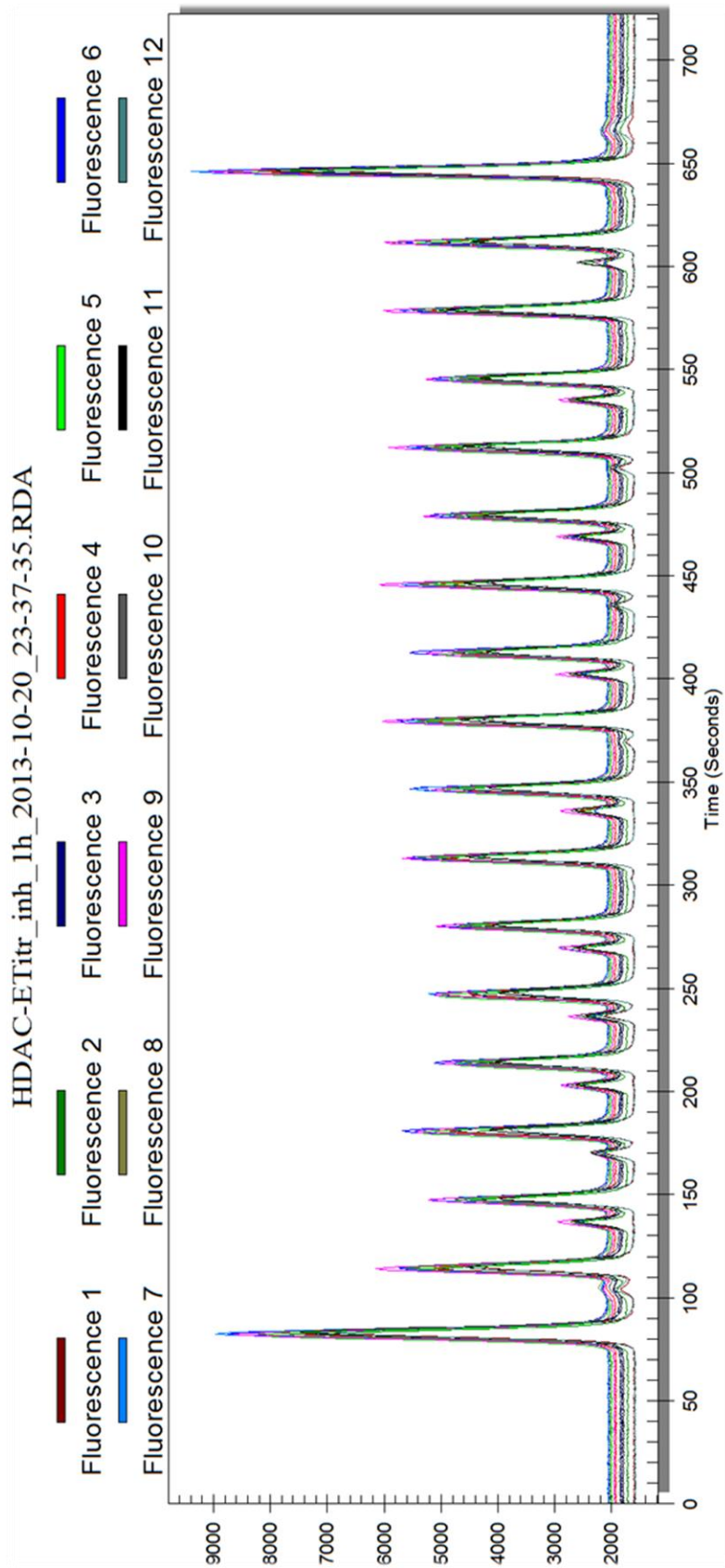


Figure 14. HDAC 3rd screening. Data Tracings (CCD2 signals).

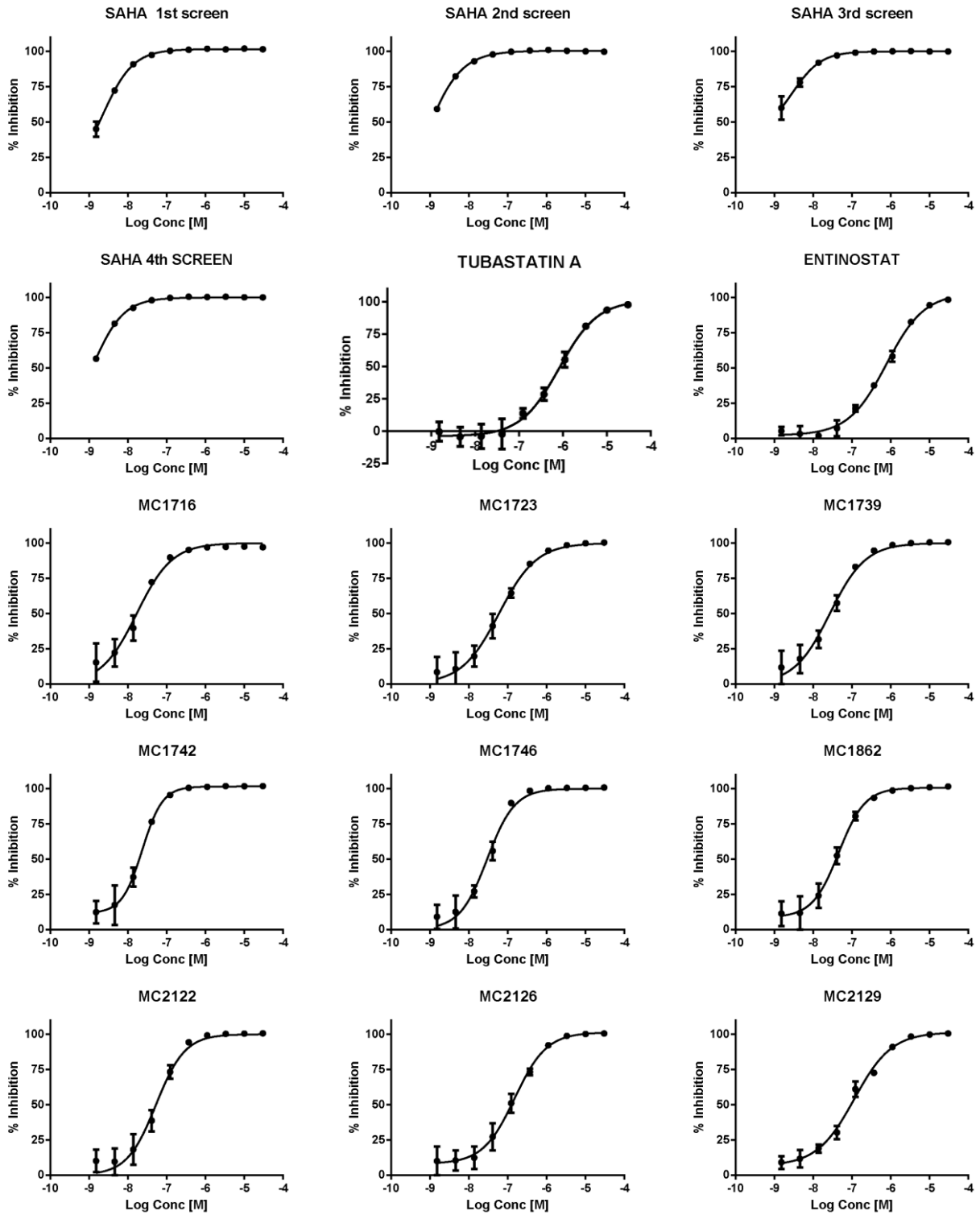


A



B

Figure 15. HDAC 3rd screening: Conversion rates (A) and inhibition rates (B) outputs. Rows and columns refers to those of the plate (compare with Figure 12 and Table 5).



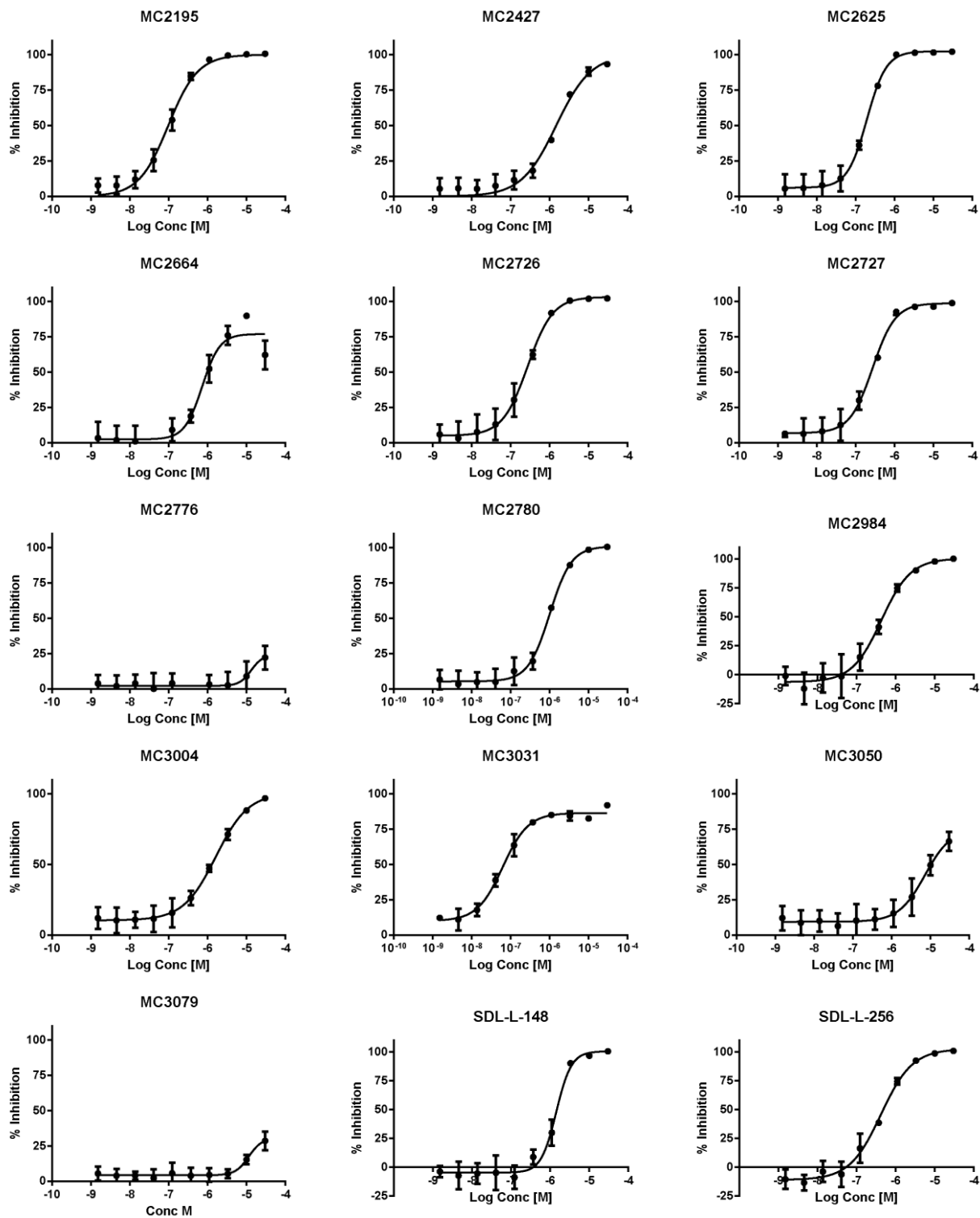


Figure 16. HDAC 3 inhibitory dose-response curves.

Table 6. HDAC 3: assays: inhibitory activities expressed as IC₅₀ and pIC₅₀.

Cpd	HDAC 3			
	IC ₅₀		pIC ₅₀	
	COMPANY	GRM	COMPANY	GRM
SAHA 1st SCREEN	2.00E-08	2.18E-09	7.70	8.66
SAHA 2nd SCREEN	2.00E-08	6.24E-10	7.70	9.20
SAHA 3rd SCREEN	2.00E-08	2.92E-09	7.70	8.54
SAHA 4th SCREEN	2.00E-08	6.17E-10	7.70	9.21
TUBASTATIN A	3.00E-05	8.20E-07	4.52	6.09
ENTINOSTAT	1.70E-06	7.73E-07	5.77	6.11
MC1716		2.34E-08		7.63
MC1723		7.29E-08		7.14
MC1739		3.72E-08		7.43
MC1742		2.36E-08		7.63
MC1746		3.81E-08		7.42
MC1862		4.58E-08		7.34
MC2122		6.81E-08		7.17
MC2126		1.52E-07		6.82
MC2129		1.15E-07		6.94
MC2195		1.20E-07		6.92
MC2427		1.57E-06		5.80
MC2625		1.94E-07		6.71
MC2664		7.25E-07		6.14
MC2726		2.75E-07		6.56
MC2727		2.73E-07		6.56
MC2776		1.32E-05		4.88
MC2780		1.01E-06		5.99
MC2984		4.32E-07		6.36
MC3004		1.63E-06		5.79
MC3031		6.31E-08		7.20
MC3050		7.25E-06		5.14
MC3079		1.14E-05		4.94
SD-L-148		1.44E-06		5.84
SD-L-256		4.23E-07		6.37
SDM 141		NAc		NAc
SDM 146		NAc		NAc

NAc: No inhibition/activity not converged in a IC₅₀ curve;

NAv: Not Available

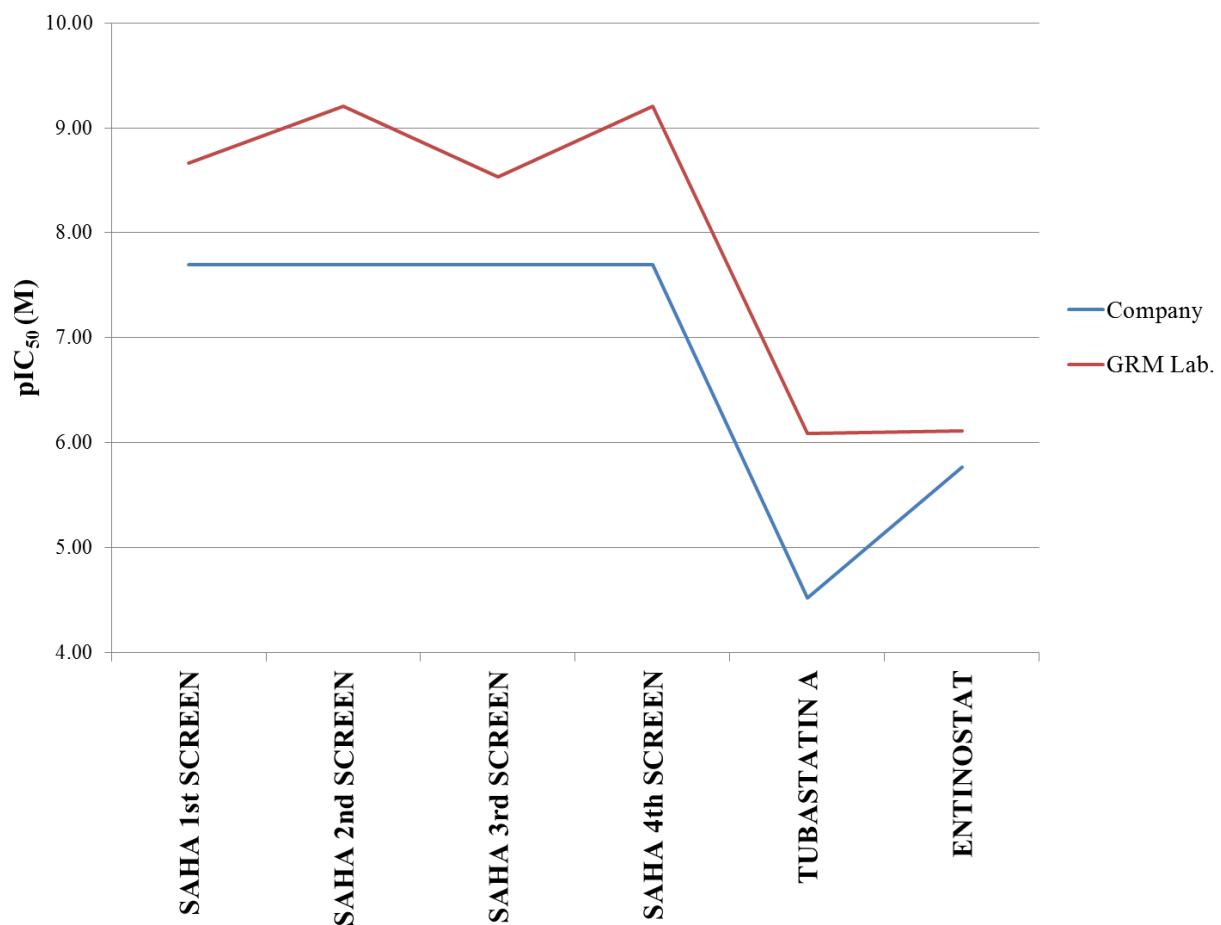


Figure 17. HDAC 3: standard compounds' inhibitory profile, pIC₅₀ values are reported on molar base.

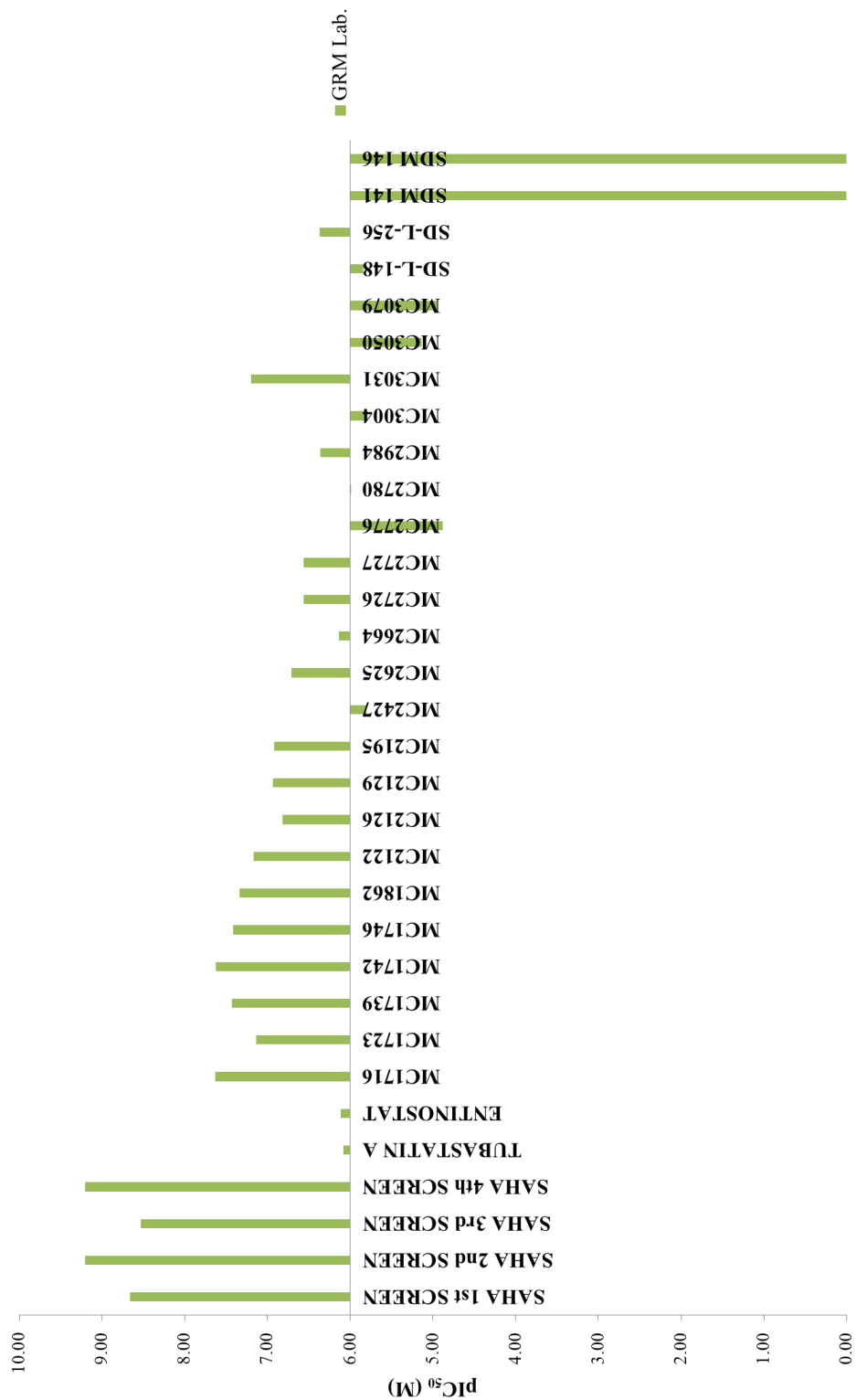


Figure 18. HDAC 3: compounds Inhibitory Profile, pIC₅₀ values are reported on molar base.

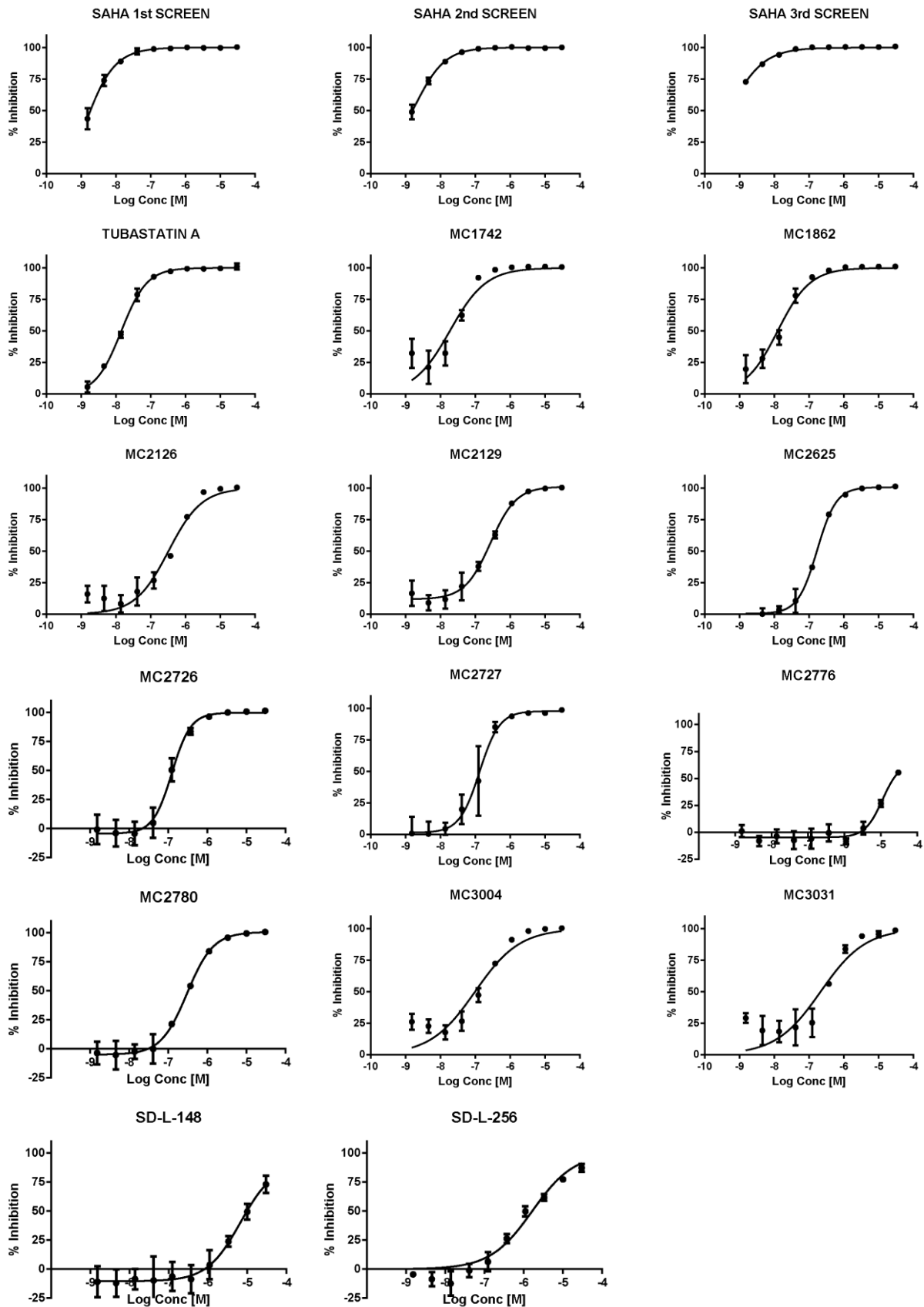


Figure 19. HDAC 6 inhibitory dose-response curves.

Table 7. HDAC 6 assays: inhibitory activities expressed as IC₅₀ and pIC₅₀.

Cpd	HDAC 6			
	IC ₅₀		pIC ₅₀	
	COMPANY	GRM	COMPANY	GRM
SAHA 1st SCREEN	NAv	1.02E-09	NAv	8.99
SAHA 2nd SCREEN	NAv	1.92E-09	NAv	8.72
SAHA 3rd SCREEN	NAv	1.31E-09	NAv	8.88
TUBASTATIN A	NAv	1.43E-08	NAv	7.85
ENTINOSTAT	NAv	NAc	NAv	NAc
MC1742		4.30E-08		7.37
MC1862		2.21E-08		7.66
MC2126		5.38E-07		6.27
MC2129		2.73E-07		6.56
MC2625		1.70E-07		6.77
MC2664		NAc		NAc
MC2726		1.24E-07		6.91
MC2727		1.32E-07		6.88
MC2776		1.11E-05		4.96
MC2780		3.03E-07		6.52
MC2984		NAc		NAc
MC3004		2.20E-07		6.66
MC3031		4.68E-07		6.33
MC3050		NAc		NAc
MC3079		NAc		NAc
SD-L-148		6.68E-06		5.18
SD-L-256		8.01E-07		6.10
SDM 141 ^a		NAc		NAc
SDM 146 ^a		NAc		NAc

NAc: No inhibition/activity not converged in a IC₅₀ curve;

NAv: Not Available

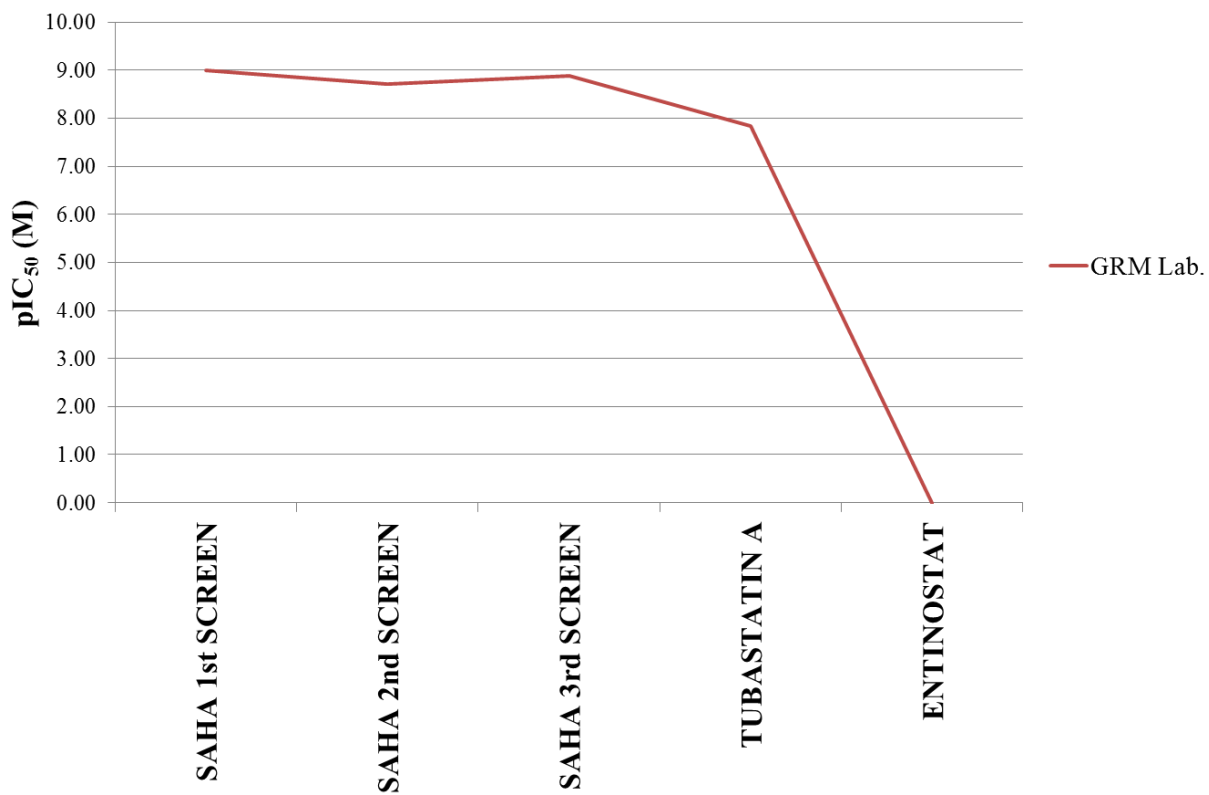


Figure 20. HDAC 6 Standard compounds' inhibitory profile, pIC₅₀ values are reported on molar base.

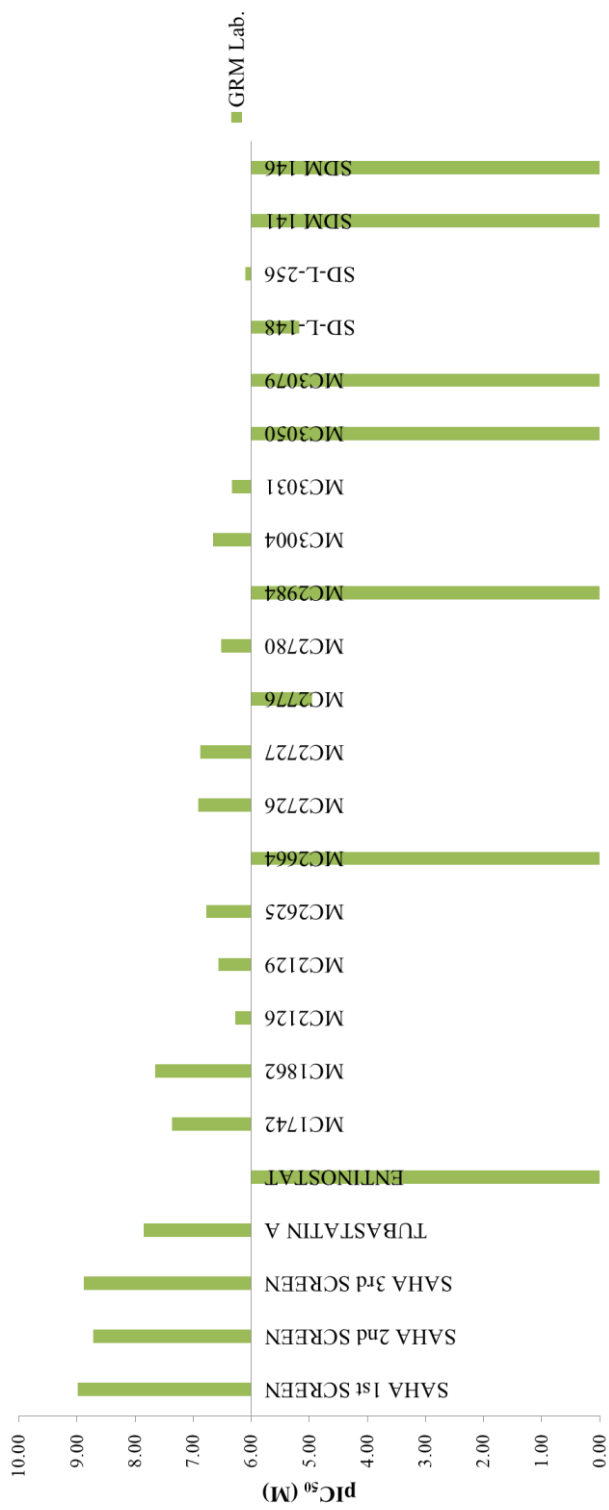


Figure 21. HDAC 6: Compounds Inhibitory Profile, pIC₅₀ values are reported on molar base.

Bibliography

1. Perrin, D.; Fremaux, C.; Shutes, A., Capillary microfluidic electrophoretic mobility shift assays: application to enzymatic assays in drug discovery. *Expert opinion on drug discovery* **2010**, *5*, 51-63.
2. Laurel M. Provencher, A. L. E., Seth P. Cohen *Characterization of HDAC Substrates and Inhibitors using a Microfluidic Mobility-Shift Assay*; Caliper Life Sciences, Hopkinton, MA USA.

Chapter XI

Scientific Production during the PhD scholarship

Journal publications:

1. L. Friggeri, **BALLANTE F.**, R. Ragno, I. Musmuca, D. De Vita, F. Manetti, M. Biava, L. Scipione, R. Di Santo, R. Costi, M. Feroci, S. Tortorella (2013). Pharmacophore assessment through 3-D QSAR: evaluation of the predictive ability on new derivatives by the application on a serie of antitubercular agents. *JOURNAL OF CHEMICAL INFORMATION AND MODELING*; p 1463–1474, ISSN: 1549-9596, doi: 10.1021/ci400132q
2. Enrico Perspicace, Valérie Jouan-Hureau, Rino Ragno, **BALLANTE F.**, Stefania Sartini, Concettina La Motta, Federico Da Settimo, Binbin Chen, Gilbert Kirsch, Serge Schneider, Béatrice Faivre, Stéphanie Hesse (2013). Design, synthesis and biological evaluation of new classes of thieno[3,2-d]pyrimidinone and thieno[1,2,3]triazine as inhibitor of vascular endothelial growth factor receptor-2 (VEGFR-2). *EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY*, vol. 63; p. 765-781, ISSN: 0223-5234, doi: 10.1016/j.ejmech.2013.03.022
3. **BALLANTE F.**, Rino Ragno (2012). 3-D QSAutogrid/R: An Alternative Procedure To Build 3-D QSAR Models. Methodologies and Applications. *JOURNAL OF CHEMICAL INFORMATION AND MODELING*, ISSN: 1549-9596, doi: 10.1021/ci300123x
4. Rotili D, Samuele A, Tarantino D, Ragno R, Musmuca I, **BALLANTE F.**, Botta G, Morera L, Pierini M, Cirilli R, Nawrozki MB, Gonzalez E, Clotet B, Artico M, Esté JA, Maga G, Mai A. (2012). 2-(Alkyl/aryl)amino-6-benzylpyrimidin-4(3H)-ones as inhibitors of wild-type and mutant HIV-1: enantioselectivity studies. *JOURNAL OF MEDICINAL CHEMISTRY*, vol. 55; p. 3558-3562, ISSN: 1520-4804, doi: 10.1021/jm201308v
5. **BALLANTE F.**, Ira Musmuca, Garland R. Marshall, Rino Ragno (2012). Comprehensive model of wild-type and mutant HIV-1 reverse transcriptases. *JOURNAL OF COMPUTER-AIDED MOLECULAR DESIGN*, vol. 26; p. 907-919, ISSN: 0920-654X, doi: 10.1007/s10822-012-9586-6
6. Laura Silvestri, **BALLANTE F.**, Antonello Mai, Garland R. Marshall, Rino Ragno (2012). Histone Deacetylase Inhibitors: Structure-Based Modeling and Isoform-Selectivity Prediction. *JOURNAL OF CHEMICAL INFORMATION AND MODELING*, vol. 52; p. 2215-2235, ISSN: 1549-9596, doi: 10.1021/ci300160y
7. **FLAVIO BALLANTE**, Antonia Caroli, Richard B. Wickersham III and Rino Ragno (2013) Hsp90 Inhibitors (I). Definition of 3-D QSAutogrid/R Models as a Tool for Virtual Screening. *JOURNAL OF CHEMICAL INFORMATION AND MODELING*, submitted
8. Antonia Caroli, **FLAVIO BALLANTE**, Richard B. Wickersham III, Federico Corelli and Rino Ragno (2013) Hsp90 Inhibitors (II). Combining ligand-based and structure-based approaches for Virtual Screening application. *JOURNAL OF CHEMICAL INFORMATION AND MODELING*, submitted

Published Abstract

- 1 Computational strategies to design new highly potential BSAO polyamine substrates. **F. BALLANTE**, G. Tempera, E. Agostinelli, R. Ragno. Abstracts presented at the 13th International Congress on Amino acids, peptides and proteins. Amino Acids 2013, 45, 563-612. ISSN: 0939-4451 doi: 10.1007/s00726-013-1540-y

Oral communications:

- 1 Computational strategies to design new highly potential BSAO polyamine substrates. **F. BALLANTE**, G. Tempera, E. Agostinelli, R. Ragno. 13th International Congress on Amino Acids, Peptides and Proteins (ICAPP). October 5-7, 2013. Galveston, Texas, USA

Seminars:

- 1 3-D QSAutogrid/R: behind the scenes of the new 3-DQSAR procedure. **F. BALLANTE**, CNRS CRM2, 13 December 2012. Université de Lorraine Faculté des Sciences et Technologies

Abstract/Posters:

1. Automatic generation of 3-D QSAR Models: a quasi-systematic approach. ADELE PIROLI, **FLAVIO BALLANTE**, MANUELA SABATINO, ALEXANDROS PATSILINAKOS, RINO RAGNO. XXII National Meeting on Medicinal Chemistry (NMMC), September 10-13, 2013, Rome, Italy

2. Design And Synthesis Of 2-Indolyl-Thieno[2,3-D]Pyrimidinones As Potential Inhibitors Of VEGFR-2. CHARLÈNE GADAIS, **FLAVIO BALLANTE**, STÉPHANIE HESSE, RINO RAGNO, GILBERT KIRSCH. International Conference on Medicinal Chemistry, RICT 2013, July 3-5 2013, Nice, France

3. VEGFR-2 Inhibitors. New perspectives from molecular modeling. **F. BALLANTE**, A. PIROLI, A. PATSILINAKOS, C. GADAIS, S. HESSE, G. KIRSCH, R. RAGNO (2013) 29-31 Maggio 2013, NPCF7, Savigliano (CN)

4. Molecular docking of biaryl tetrazolyl ureas as inhibitors of endocannabinoid metabolism PIROLI A. **BALLANTE F.**, PATSILINAKOS A., ORTAR G., MORERA E., RAGNO R. 29-31 Maggio 2013, NPCF7, Savigliano (CN)

5. Molecular docking of biaryl tetrazolyl ureas as inhibitors of endocannabinoid metabolism PIROLI A. **BALLANTE F.**, PATSILINAKOS A., ORTAR G., MORERA E., RAGNO R. Nine European Workshop in Drug Design. May 19-25 May 2013 Certosa di Pontignano, Siena (Italy)

6. Conception and Synthesis of New Inhibitors of VEGFR-2. C. GADAIS, **F. BALLANTE**, S. HESSE, G. KIRSCH, R. RAGNO JJC SCT 7 et 8 février, Biocitech (Romainville), 2013.

7. "Interactions between new polyamine analogs and bovine serum amine oxidase" GIAMPIERO TEMPERA, ANNA MINARINI, ANDREA MILELLI, VINCENZO TUMIATTI, TAICHI UESHIMA, MARTINA MERINGOLO, EMANUELA BONAIUTO, MARIA LUISA DI PAOLO, ANTONIO TONINELLO, RINO RAGNO,

FLAVIO BALLANTE, ROBERTO STEVANATO AND ENZO AGOSTINELLI.
International Congress on Polyamines, Istanbul (TURKEY), September 2nd to 7th, 2012

8. "Histone Deacetylase Inhibitors. The Development of a Model for Target-Selectivity Prediction." LAURA SILVESTRI, **FLAVIO BALLANTE**, GARLAND R. MARSHALL AND RINO RAGNO. CDDD Meeting. Computationally Driven Drug Discovery Meeting. November 21-23, 2011 - L'Aquila

9. "HDAC Inhibitors Pharmacophore Modelling using Ligand Scout 3.0." **FLAVIO BALLANTE**, ANNALUCIA FALLACARA, RENÉ WEISSENSTEINER, ANDREAS JURIK, CONSTANTINOS POTAMITIS, JENS PILGER, JONAS DEMEULEMEESTER, LAURA GUASCH PÀMIES, MAIK KINDERMANN VIII EWDD. Eighth European Workshop in Drug Design. May 22 - 28, 2011 - Certosa di Pontignano, Siena (Italy)

10. "Histone deacetylase inhibitors: Comparative Binding Energy and QSAR studies for target selectivity prediction" **FLAVIO BALLANTE**, RINO RAGNO VIII EWDD. Eighth European Workshop in Drug Design. May 22 - 28, 2011 - Certosa di Pontignano, Siena (Italy)

11. "βsecretase: quantitative structure-activity relationships (3D QSARs) and evaluation of predictive ability."
ALBERTO DE PETRIS, **FLAVIO BALLANTE**, RINO RAGNO VIII EWDD. Eighth European Workshop in Drug Design. May 22 - 28, 2011 - Certosa di Pontignano, Siena (Italy)

Courses, seminars, conferences attended during the PhD scholarship

Courses:

- 1 Bioinformatica Strutturale: Analisi E Predizione Della Struttura Delle Proteine E Delle Loro Interazioni Con Ligandi (2013, Sapienza Università di Roma). Dr.ssa Veronica Morea, Dr. Alessandro Paiardini

- 2 School Of Experimental Design (Section of Analytical Chemistry and Chemometrics of the Department of Chemistry and Food and Pharmaceutical Technologies of the University of Genoa, September 19-23, 2011). Teacher Prof. Riccardo Leardi

- 3 Design Of Subtype-Specific Histone Deacetylase Inhibitors As Epigenetic Modulators VIII EWDD May 22 - 28, 2011 - Certosa di Pontignano, Siena (Italy) University of Siena. Teachers: Sharon Bryant, Daniela Schuster, Christine Huynh Buu (Inte:Ligand GmbH – Austria); Gerhard Wolber (Inte:Ligand GmbH - Austria and Freie Universität Berlin – Germany) and Wolfgang Sippl (University of Halle – Germany)

- 4 Workshop on Drug Design- S-IN- Schrödinger – Istituto Ricerche Farmacologiche "Mario Negri", 2011. Supervisor: Jas Bhachoo

Conferences:

- 1 The 13th International Congress on Amino Acids, Peptides and Proteins (ICAPP). October 5-7, 2013. Galveston, Texas, USA

- 2 Nuove Prospettive in Chimica Farmaceutica (NPCF7), 29-31 Maggio 2013 Savigliano (CN)

- 3 Epigenetic Rome Training School. May 21-24, 2013 Rome (Italy)

-
- 4 International Congress on Polyamines: Biological and Clinical Perspectives, Istanbul (TURKEY), September 2nd to 7th, 2012

 - 5 Angiogenesi: Ricerca di base e applicazioni cliniche. March 27 2012. Accademia Medica di Roma, Rome (Italy)

 - 6 VIII EWDD. Eighth European Workshop in Drug Design. May 22 - 28, 2011 - Certosa di Pontignano, Siena (Italy)

 - 7 Innovative Targets and Drug Design Strategies in Cancer Therapy, Sapienza Universita' di Roma- 2011

Seminars:

- 1 Total Synthesis of Depsipeptide HDAC Inhibitors. Prof. A. Ganesan. Sapienza Universita' di Roma- 2012

 - 2 Chemistry of Fungi. Prof. Gilbert Kirsch. Sapienza Universita' di Roma- 2012

 - 3 L'utilizzo della piattaforma avidina/biotina per applicazioni di terapia con radionuclidi in oncologia Dott.ssa Rita De Santis. Sapienza Universita' di Roma- 2012

 - 4 Multipole vs Monopole Electrostatics - Comparison of Predicted NOE Spectra with Experimental Data. Prof. Garland Marshall Sapienza Universita' di Roma- 2011
-

Publications

2-(Alkyl/Aryl)Amino-6-Benzylpyrimidin-4(3*H*)-ones as Inhibitors of Wild-Type and Mutant HIV-1: Enantioselectivity Studies

Dante Rotili,^{†,‡} Alberta Samuele,^{‡,§} Domenico Tarantino,[†] Rino Ragno,[†] Ira Musmuca,[†] Flavio Ballante,[†] Giorgia Botta,[†] Ludovica Morera,[†] Marco Pierini,[†] Roberto Cirilli,[§] Maxim B. Nawrozki,^{||} Emmanuel Gonzalez,[⊥] Bonaventura Clotet,[⊥] Marino Artico,[†] José A. Esté,^{*,⊥} Giovanni Maga,^{*,‡} and Antonello Mai^{*,†}

[†]Istituto Pasteur—Fondazione Cenci Bolognetti, Dipartimento di Chimica e Tecnologie del Farmaco, Università degli Studi di Roma “La Sapienza”, P.le A. Moro 5, 00185 Rome, Italy

[‡]Istituto di Genetica Molecolare IGM-CNR, via Abategrasso 207, 27100 Pavia, Italy

[§]Dipartimento del Farmaco, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy

^{||}Volgograd State Technical University, prospekt Lenina, 28, 400131 Volgograd, Russia

[⊥]Retrovirology Laboratory IrsiCaixa, Hospital Universitari Germans Trias i Pujol, Universitat Autònoma de Barcelona, 08916 Badalona, Spain

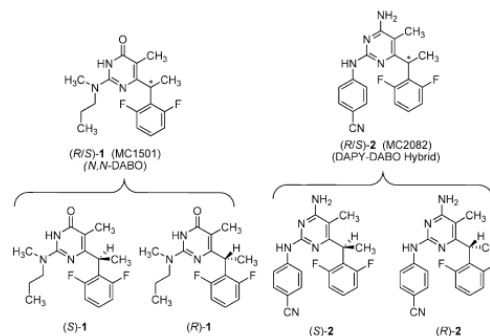
S Supporting Information

ABSTRACT: The single enantiomers of two pyrimidine-based HIV-1 non-nucleoside reverse transcriptase inhibitors, **1** (MC1501) and **2** (MC2082), were tested in both cellular and enzyme assays. In general, the *R* forms were more potent than their *S* counterparts and racemates and (*R*)-**2** was more efficient than (*R*)-**1** and the reference compounds, with some exceptions. Interestingly, (*R*)-**2** displayed a faster binding to K103N RT with respect to WT RT, while (*R*)-**1** showed the opposite behavior.

INTRODUCTION

Since 1992, our research team has discovered excellent dihydro-alkoxybenzyl-oxypyrimidine (DABO) classes of non-nucleoside reverse transcriptase (RT) inhibitors (NNRTIs) such as F₂-S-DABOs,¹ F₂-NH-DABOs,² and F₂-N,N-DABOs.³ Such compounds had inhibitory potencies in the (sub)nanomolar range against wild-type (WT) HIV-1 without significant cytotoxicity at higher concentrations, and with potencies in the submicromolar range against clinically relevant mutant strains. The F₂-N,N-DABO derivative **1** (MC1501),³ characterized by a *N*-methyl-*N*-propyl side chain at the C-2 pyrimidine ring position and by a double pyrimidine C-5/C-6 benzylic position methyl substitution, with its subnanomolar inhibitory potency against WT HIV-1 and the Y181C mutant strain, can be considered one of the most promising DABO compounds reported so far (Chart 1).

Diarylpyrimidine (DAPY) derivatives are one of the most successful classes of pyrimidine-based NNRTIs, as demonstrated by the recent approval of etravirine and rilpivirine for clinical use.^{4,5} SAR studies on the DAPY class had also shown how the *para*-cyanoaniline substitution at the C-2 pyrimidine ring position is crucial for a strong inhibitory activity. Recently, combining the 2,6-difluorobenzyl or 1-(2,6-difluorophenyl)-ethyl group at the C-6 position characteristic of DABOs with the *para*-cyanoaniline group at the C-2 of the pyrimidine ring typical of DAPYs, and inserting at the C-4 pyrimidine ring position groups found in either DABOs (–OH) or DAPYs (–Cl, –H, –NH₂), a new series of highly potent anti-HIV-1 agents called DAPY-DABO hybrids has been developed.⁶ Among those, the derivative **2** (MC2082) (Chart 1) was the

Chart 1. Pyrimidine-Based NNRTIs That Are the Object of the Present Study

most potent due to its (sub)nanomolar activity against WT and clinically relevant mutant HIV-1 strains, such as K103N and Y188L.⁶

Both the pyrimidine-based NNRTIs **1** and **2** contain a stereogenic center at the C-6 benzylic position and, therefore, exist as racemic mixtures. Because of the high potency of these two lead compounds, we decided to perform a systematic investigation on their potential enantioselective anti-HIV-1 activity. Here we describe, through a multidisciplinary approach, the enantioselective separation, the absolute configuration assignment, and the enzymatic and cellular evaluation of the

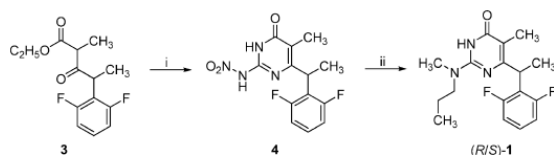
Received: October 11, 2011

Published: March 19, 2012

HIV-1 inhibitory activity of the single enantiomers in comparison with the corresponding racemic mixtures of both compounds. The application of a comparative binding energy (ComBinE) model built to quantitate the influence of single point HIV-RT mutations to modeled bound conformations of both **1** and **2** was in good agreement with the experimental findings and gave insights on the absolute configuration assignment of both **1** and **2** separated enantiomers.

Chemistry. The racemic mixture of **2** was prepared as previously described.⁶ The racemic mixture of **1** was prepared starting from the β -oxoester **3**¹ and using, differently from previously reported,³ a route characterized by only two steps and by a final nucleophilic displacement under mild conditions and with high yield. After condensation of the β -oxoester **3** with nitroguanidine, the resulting 6-[1-(2,6-difluorophenyl)-ethyl]-5-methyl-2-nitroamino-3H-pyrimidin-4-one **4** was heated in a sealed tube of a Parr high pressure reactor in the presence of an excess of *N*-methyl-*N*-propylamine to give directly the racemic **1** (Scheme 1).

Scheme 1^a



^a(i) EtONa, nitroguanidine, dry EtOH, reflux; (ii) *N*-methyl-*N*-propylamine, 120 °C, Parr high pressure reactor.

RESULTS AND DISCUSSION

The direct resolution of **1** and **2** was achieved by HPLC on the coated-type Chiralpak AD chiral stationary phase (CSP) using pure ethanol in mixture with a small percentage (0.1%) of diethylamine (DEA) as eluent (see Supporting Information). Efforts to obtain the absolute configuration of **1** and **2** by X-ray crystallography were not successful, therefore stereochemical information on the four chiral compounds were obtained by analyzing their chiroptical properties. The (*S*) configuration was assigned to the dextrorotatory enantiomers of **1** and **2** ((*S*-

(+)-**1** and (*S*)-(+)-**2**, first-eluted enantiomers on the Chiralpak AD CSP) and the (*R*) configuration assigned to the levorotatory counterparts ((*R*)-(–)-**1** and (*R*)-(–)-**2**, second-eluted enantiomers).

The racemic compounds **1** and **2** and their enantiomers were tested in MT-4 cells to evaluate their cytotoxicity and their capability to inhibit by 50% the HIV-induced cytopathic effect (HIV-1 strain: NL4–3). The compounds were also tested against a panel of clinically relevant HIV-1 mutant strains (K103N, Y181C, and Y188L). Nevirapine (NVP), efavirenz (EFV), and dapivirine (TMC120) were also tested as reference drugs (Table 1).

The enantiomer (*R*)-**1** was much more potent (ranging from about 2800- to 11000-fold) than the other enantiomer (*S*)-**1** in inhibiting WT HIV-1 and the three tested mutant strains. (*R*)-**1** was slightly better, both in terms of activity and of selectivity, than the corresponding racemic mixture. The enantiomer (*R*)-**2** inhibited WT and mutant HIV-1 strains from 900- to 10000-fold more efficiently than the other enantiomer and was slightly more potent than the racemic mixture. With the only exception of the mutant Y181C, (*R*)-**2** was always more efficient than (*R*)-**1** and TMC120 in the inhibition of both WT and mutant (K103N and Y188L) HIV-1 strains (Table 1). Moreover, (*R*)-**2** was 10-fold less toxic than TMC120, yielding an highly improved selectivity index ($SI_{(R)-2} = 300000$; $SI_{TMC120} = 5000$).

We evaluated the RNA-dependent DNA polymerase activity of WT RT and those enzymes carrying the most common NNRTI-resistance mutations (K103N, L100I, Y181I, V106A, Y188L) in the presence of increasing concentrations of the **1** and **2** enantiomers (Table 2). In the inhibition assays, the enantiomers (*R*)-**1** and (*R*)-**2** showed higher potency than their corresponding *S* isomers toward RT WT and mutants. Moreover, (*R*)-**2** showed the lowest ID_{50} values with respect to the majority of the mutants evaluated, such as K103N, V106A, and Y188L, whereas its efficacy appeared slightly lesser than that observed for (*R*)-**1** in the case of L100I and Y181I. As illustrated in Table 2, in all cases both racemic compounds presented intermediate levels of potency compared to the corresponding pairs of enantiomers.

Steady-state kinetic assays were performed for the two most potent isomers, (*R*)-**1** and (*R*)-**2**, to assess the inhibition mechanism against RT WT and RT carrying K103N mutation. The enzyme activity was measured in the presence of fixed

Table 1. Cytotoxicity and Anti-HIV-1 Activity against WT (NL4-3) and Clinically Relevant HIV-1 Mutant Strains of Racemic Compounds **1 and **2** and their Enantiomers^a**

compd	EC_{50}^b nM (fold resistance) ^c				CC_{50}^d nM	selectivity index ^e
	NL4–3	K103N	Y181C	Y188L		
1	1.9 ± 0.2	16.5 ± 0.5 (9)	17 ± 1 (9)	49 ± 4 (26)	>78000	>41000
(<i>S</i>)- 1	15000 ± 50	>78000	>78000	>78000	>78000	>5.2
(<i>R</i>)- 1	1.3 ± 0.1	11 ± 0.7 (8)	13 ± 1 (10)	28 ± 2 (22)	75000 ± 50	58000
2	0.6 ± 0.1	35 ± 3 (60)	40 ± 2 (70)	27 ± 3 (40)	40000 ± 70	70000
(<i>S</i>)- 2	1000 ± 10	4500 ± 10 (4)	>38000	6500 (6)	38000 ± 50	40
(<i>R</i>)- 2	0.1 ± 0.02	0.8 ± 0.01 (8)	31 ± 2 (300)	7.1 ± 0.5 (70)	31000 ± 50	300000
TMC120	0.6 ± 0.07	1.2 ± 0.2 (2)	1.2 ± 0.2 (2)	330 ± 10 (500)	3000 ± 30	5000
NVP	130 ± 1	6100 ± 20 (47)	>7500	>7500	7500 ± 20	58
EFV	7.3 ± 0.5	340 ± 4 (47)	10 ± 1 (1.4)	1600 ± 5 (220)	3200 ± 8	440

^aValues are means ± SD determined from at least three experiments. ^bEffective concentration 50, concentration needed to inhibit 50% HIV-induced cytopathic effect, evaluated with the MTT method in MT-4 cells. ^cFold resistance: ratio of EC_{50} value against drug-resistant strain and EC_{50} of the WT NL4–3 strain. ^dCytotoxic concentration 50, concentration to induce 50% death of noninfected cells, evaluated with the MTT method in MT-4 cells. ^eSelectivity index, CC_{50}/EC_{50} .

Table 2. Comparison of Inhibition Potencies, Expressed as ID₅₀ (nM), between the Racemic Compounds 1 and 2 and their Enantiomers, Tested against RT Wild-Type (WT) and K103N, L100I, Y181I, V106A, Y188L RT Mutants^a

compd	ID ₅₀ ^b nM (fold resistance) ^c					
	WT	K103N	L100I	Y181I	V106A	Y188L
1	30 ± 3	174 ± 4 (6)	177 ± 9 (6)	1260 ± 8 (40)	19 ± 1 (0.6)	1430 ± 20 (50)
(S)-1	975 ± 5	>20000	350 ± 8 (0.4)	>20000	120 ± 8 (0.1)	>20000
(R)-1	2 ± 0.2	96 ± 9 (48)	96 ± 1 (50)	450 ± 9 (200)	3 ± 0.2 (1.5)	195 ± 5 (100)
2	14 ± 2	22 ± 3 (1.6)	280 ± 3 (20)	1700 ± 10 (120)	ND ^d	150 ± 6 (11)
(S)-2	180 ± 8	240 ± 9 (1.3)	>15000	>20000	64 ± 4 (0.3)	500 ± 8 (3)
(R)-2	8 ± 1	8 ± 1 (1)	150 ± 9 (20)	750 ± 8 (90)	0.3 ± 0.02 (0.04)	3 ± 0.2 (0.4)
TMC120	7 ± 1	100 ± 10 (14)	90 ± 1 (13)	>1000	ND	380 ± 3 (50)
NVP	400 ± 10	7000 ± 100 (20)	ND	35000 ± 100 (87)	ND	ND
EFV	30 ± 3	3000 ± 50 (100)	ND	80 ± 3 (3)	ND	ND

^aValues are means ± SD determined from at least three experiments. ^bInhibitory dose 50, compound dose required to inhibit the HIV-1 rRT activity by 50%. ^cFold resistance: ratio of ID₅₀mut/ID₅₀WT values. ^dND: not determined.

concentrations of the compounds and variable concentrations of one of the two substrates (either nucleic acid or nucleotide), while the other was maintained at saturating doses. The kinetic parameters V_{max} and K_m were derived by fitting the experimental data to the appropriate rate equations (Table S1 in Supporting Information). For both compounds, either varying the nucleic acid (NA) or the nucleotide (dNTP) concentrations, the V_{max} values of RT WT or K103N decreased as the inhibitor concentrations increased while the K_m values did not change. Thus, the mechanism of action of (R)-1 and (R)-2 against RT WT and K103N was fully noncompetitive. The (R)-2 showed K_i values lower than those of (R)-1. Binding experiments were subsequently carried out, with RT WT and K103N to evaluate the association (k_{on}) and the dissociation rate (k_{off}) values for each inhibitor with respect to the different catalytic forms of the enzyme (free, binary complex, ternary complex) along the reaction pathway (Table S2 in Supporting Information). The k_{on} rate values were used to calculate the relative association index ($RAI = k_{on} WT/k_{on} mut$). RAI values were >1 for (R)-1 toward the RT binary and ternary complexes, while they scored <1 for (R)-2 (toward the free RT and binary complex). This indicated a different interaction mechanism of the two compounds with RT: (R)-1 experienced a slower binding to the mutated form with respect to RT WT, while (R)-2, surprisingly, associated faster to the mutant RT (Figure 1A). Such unexpected behavior could be related to the presence of the cyano group into the (R)-2 structure and its ability to disrupt the K103–Y188 hydrogen bond in the apo form of the mutated enzyme.⁷ k_{off} values were used to derive the relative dissociation index ($RDI = k_{off} mut/k_{off} WT$). RDI values were, for both inhibitors, >1 toward the free enzyme and <1 in the case of the binary and ternary complexes, indicating that the binding of the compounds to the K103N RT was more stable than with WT RT once the enzymes were bound to the substrates (Figure 1B).

Biosensor-based analysis of binding of different NNRTIs to HIV-1 RT either WT or carrying drug-resistance mutants^{8,9} supported the notion that the inherent flexibility of HIV-1 RT might explain the existence of high-affinity and low-affinity forms of RT with respect to NNRTI binding. Moreover, those studies revealed that the resistance induced by some mutations (such as K103N) was due to effects on both the association (k_{on}) and dissociation (k_{off}) rates of NNRTIs in an inhibitor-specific manner. In particular, the K103N mutation was found to stabilize the low-affinity conformation of the enzyme, thereby increasing both the k_{on} and k_{off} rates of some NNRTIs.

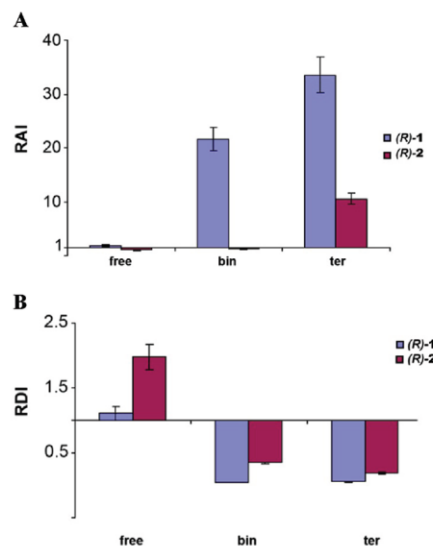


Figure 1. (A) Relative association index (RAI) derived from the association rates (k_{on}) of the inhibitors (R)-1 and (R)-2 to RT K103N and RT WT, expressed as ratio $k_{on} WT/k_{on} mut$. (B) Relative dissociation index (RDI) derived from the dissociation rates (k_{off}) of the inhibitors from RT K103N and RT WT, expressed as ratio $k_{off} mut/k_{off} WT$. free, free enzyme; bin, binary complex; ter, ternary complex. Error bars indicate the standard deviations (\pm SD).

Binding of substrates to RT is known to stabilize specific conformations of the enzyme, thus narrowing its range of inherent flexibility. According to the studies mentioned above, this implies that a given mutation will exert its effects on a different set of possible enzyme conformations, depending on whether the enzyme is unliganded, bound to nucleic acid, or in the ternary complex conformation. Our data suggest that such substrate-induced conformational changes are capable of influencing the ability of a certain mutation to affect the different kinetic steps of NNRTI interaction.

A structure-base study by means of ComBinE¹⁰-like protocol was undertaken (see Supporting Information) to further support the experimental evidence on the absolute configuration of the most- and the least-active 2 enantiomers and to provide a quantitative assessment of the mutation influence on the activity potency. The statistically most robust ComBinE model CM4, (previously developed without consideration of these compounds, Ballante, F., et al. *J. Comput.-Aided Mol. Des.*,

submitted for publication) was used to verify the mutation role on the activity profile of (*R*)- and (*S*)-1 and of (*R*)- and (*S*)-2.

It could be argued that the use of static unrefined complexes cannot account for the dynamic aspect of ligand binding, nevertheless the CM4 models is able to highlight some aspects that would be otherwise ignored to interpret the experimental data. Application of the CM4 model to the inhibitors successfully reproduced the experimental eudismic ratio of (*R*) and (*S*) enantiomers as well as the increase of potency of **2** compared to **1**. In particular, among the residues around the RT NNBP a main role can be quantitatively attributed to Lys101 that was not considered in the RT mutated list. Other important residues for the binding strength were the highly conserved Trp229 and Tyr188 residues. The latter (Tyr188 in the WT isoform) serves as a steric anchor point to fulfill the concomitant lack of interactions that occurs in other mutation points such as in the case of V106A (Figure 2).

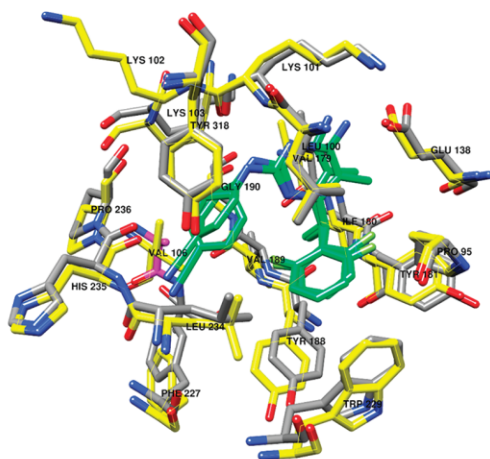


Figure 2. Comparison between (*R*)-2 (green carbon atoms) docked into WT (yellow) and V106A mutated (gray) RT. The only residue that appreciably moves is Tyr188. The mutated Ala106 is displayed in magenta.

CONCLUSION

Among the most potent pyrimidine derivatives described by us as anti-HIV-1 agents, the *N,N*-DABO **1** and the DABO–DAPY hybrid **2**, both carrying a stereogenic center at the C6-benzylic position and previously prepared and tested as racemates, were resolved and characterized for their enantioselective activities in HIV-1-infected cells as well as by enzyme assays. In both cellular and enzyme assays, the *R* enantiomers of the two compounds were significantly more potent than the *S* counterparts, the racemates having an intermediate behavior compared to the corresponding single enantiomers. (*R*)-2 was typically more efficient than (*R*)-1 as well as than the reference compounds TMC120, NVP, and EFV in inhibiting the cytopathic effect of HIV-1 strains in MT-4 cells with the exception of the Y181C mutant, against which TMC120 and EFV showed the highest effects. However, (*R*)-2 was one magnitude order less toxic than TMC120 and EFV, thus reaching a very high selectivity index. In enzyme assays, (*R*)-2 displayed the highest inhibitory activities against RT WT, K103N, V106A, and Y188L, while against L100I and Y181I (joined in this case to EFV) (*R*)-1 was more efficient. The

effect of K103N RT mutation on the (*R*)-1 and (*R*)-2 activities was further characterized. A ComBinE model based on 14 complexes between RT and EFV and NVP without any knowledge of the herein reported compounds, reproduced with acceptable errors of prediction quantitatively the differences between different pyrimidinone derivatives and their enantiomers and was used to gain insight into the role of RT mutations on the inhibitors' biological activities.

EXPERIMENTAL SECTION

Chemistry. Melting points were determined on a Buchi 530 melting point apparatus and are uncorrected. ¹H NMR spectra were recorded at 400 MHz on a Bruker AC 400 spectrometer; chemical shifts are reported in δ units relative to the internal reference tetramethylsilane (Me₄Si). All of the compounds were routinely checked by TLC and ¹H NMR. TLC was performed on aluminum backed silica gel plates (Merck DC, Alufolien Kieselgel 60 F254) with spots visualized by UV light. All of the solvents were reagent grade and, when necessary, were purified and dried by standard methods. Concentration of solutions after reactions and extractions involved the use of a rotary evaporator operating at a reduced pressure of ca. 20 Torr. Organic solutions were dried over anhydrous sodium sulfate. Elemental analysis has been used to determine purity of the described compounds, that is >95%. Analytical results are within $\pm 0.40\%$ of the theoretical values. All chemicals were purchased from Aldrich Chimica, Milan (Italy), or from Lancaster Synthesis GmbH, Milan (Italy), and were of the highest purity.

Preparation of 6-[1-(2,6-Difluorophenyl)-ethyl]-5-methyl-2-nitroamino-3*H*-pyrimidin-4-one (4**).** See Supporting Information.

Preparation of 6-[1-(2,6-Difluorophenyl)-ethyl]-5-methyl-2-(methyl-*n*-propyl-amino)-3*H*-pyrimidin-4-one (1**).** A mixture of 6-[1-(2,6-difluorophenyl)-ethyl]-5-methyl-2-nitroamino-3*H*-pyrimidin-4-one **4** (100 mg, 0.322 mmol) and *N*-methyl-*N*-propylamine (707.2 mg, 9.66 mmol, 0.99 mL) was heated in a sealed tube of a Parr apparatus at 120 °C for 5 h. After cooling, the crude residue was dissolved in ethyl acetate (15 mL) and water (15 mL). The aqueous phase was extracted with ethyl acetate (3 \times 15 mL). The organic extracts were washed with brine (1 \times 20 mL), dried, evaporated under reduced pressure, and purified by column chromatography (silica gel, ethyl acetate/chloroform 1:2) to give the desired product **1** as a white powder (75.1 mg, 72.6%); mp 128–130 °C (dichloromethane/diethyl ether). ¹H NMR (CDCl₃) δ 10.70 (s, 1H, NH), 7.13 (m, 1H, *H* benzene ring), 6.81 (m, 2H, *H* benzene ring), 4.54 (q, 1H, CHCH₃), 3.50 (m, 1H, NCHHCH₂CH₃), 3.37 (m, 1H, NCHHCH₂CH₃), 3.06 (s, 3H, NCH₃), 1.93 (s, 3H, CH₃), 1.64 (d, 3H, CHCH₃), 1.54 (m, 2H, NCH₂CH₂CH₃), 0.88 (s, 3H, NCH₂CH₂CH₃). Anal. C, H, N, F: % Calcd C, 63.54; H, 6.59; N, 13.08; F, 11.82, Percent found C, 63.22; H, 6.48; N, 13.25; F, 12.01.

Enantioseparation and Chiroptical Characterization. See Supporting Information.

Molecular Modeling: ComBinE and Docking Calculations. See Supporting Information.

Biology: Anti-HIV Activity in Lymphoid Cells. Biological activity of the compounds was tested in the lymphoid MT-4 cell line (received from the NIH AIDS Reagent Program) against the WT HIV-1 NL4–3 strain and three different HIV-1 strains, as described before.^{11,12} For a brief description, see Supporting Information.

Anti-HIV Reverse Transcriptase Assays. RNA-dependent DNA polymerase activity was assayed as described¹³ in the presence of 0.5 μ g of poly(rA)/oligo(dT)_{10:1} (0.3 μ M 3'-OH ends), 10 μ M [³H]-dTTP (1 Ci/mmol), and 2–4 nM RT in the presence of 8% final concentration of DMSO. For a brief description, see Supporting Information.

ASSOCIATED CONTENT

Supporting Information

Details on enantioseparation and absolute configuration assignment, biochemistry and molecular modeling studies,

and experimental section. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*For J.A.E.: phone, +34-934656374; fax, +34-934653968; E-mail, jaeste@irsicaixa.es. For G.M.: phone, +39 0382 546354; fax, +39 0382 422286; E-mail: maga@igm.cnr.it. For A.M.: phone, +39 06 49913392; fax, +39 06 49693268; E-mail, antonello.mai@uniroma1.it.

Author Contributions

#Equal contribution.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was partially supported by MICINN (BFU2009-06958 and SAF2010-21617-C02 to J.A.E. and B.C.) and by the Italian National Program for AIDS Research Grant 40H26 (to G.M.).

ABBREVIATIONS USED

CC₅₀, compound concentration toxic for 50% of cells; CSP, chiral stationary phase; DABOs, dihydro-alkoxy-benzyl-oxopyrimidines; DAPYs, diarylpurimidines; DEA, diethylamine; EC₅₀, effective concentration able to protect 50% of cells from the HIV-1 induced cytopathogenicity; EFV, efavirenz; F₂-N,N-DABOs, 5-alkyl-2-(N,N-disubstituted)amino-6-(2,6-difluorophenylalkyl) pyrimidin-4(3H)-ones; HIV, human immunodeficiency virus; MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; NH-DABOs, dihydro-alkylamino-benzyl-oxopyrimidines; NNBS, non-nucleoside binding site; NNRTIs, non-nucleoside reverse transcriptase inhibitors; NVP, nevirapine; RT, reverse transcriptase; SAR, structure-activity relationship; S-DABOs, dihydro-alkylthio-benzyl-oxopyrimidines; WT, wild-type

REFERENCES

- (1) Mai, A.; Sbardella, G.; Artico, M.; Ragno, R.; Massa, S.; Novellino, E.; Greco, G.; Lavecchia, A.; Musiu, C.; La Colla, M.; Murgioni, C.; La Colla, P.; Loddo, R. Structure-based design, synthesis, and biological evaluation of conformationally restricted novel 2-alkylthio-6-[1-(2,6-difluorophenyl)alkyl]-3,4-dihydro-5-alkylpyrimidin-4(3H)-ones as non-nucleoside inhibitors of HIV-1 reverse transcriptase. *J. Med. Chem.* **2001**, *44*, 2544–2554.
- (2) Mai, A.; Artico, M.; Ragno, R.; Sbardella, G.; Massa, S.; Musiu, C.; Mura, M.; Marturana, F.; Cadeddu, A.; Maga, G.; La Colla, P. 5-Alkyl-2-alkylamino-6-(2,6-difluorophenylalkyl)-3,4-dihydropyrimidin-4(3H)-ones, a new series of potent, broad-spectrum non-nucleoside reverse transcriptase inhibitors belonging to the DABO family. *Bioorg. Med. Chem.* **2005**, *13*, 2065–2077.
- (3) Mai, A.; Artico, M.; Rotili, D.; Tarantino, D.; Clotet-Codina, L.; Armand-Ugon, M.; Ragno, R.; Simeoni, S.; Sbardella, G.; Nawrozkij, M. B.; Samuele, A.; Maga, G.; Este, J. A. Synthesis and biological properties of novel 2-aminopyrimidin-4(3H)-ones highly potent against HIV-1 mutant strains. *J. Med. Chem.* **2007**, *50*, 5412–5424.
- (4) Johnson, L. B.; Saravolatz, L. D. Etravirine, a next-generation nonnucleoside reverse-transcriptase inhibitor. *Clin. Infect. Dis.* **2009**, *48*, 1123–1128.
- (5) de Bethune, M. P. Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989–2009). *Antiviral Res.* **2010**, *85*, 75–90.

- (6) Rotili, D.; Tarantino, D.; Artico, M.; Nawrozkij, M. B.; Gonzalez-Ortega, E.; Clotet, B.; Samuele, A.; Este, J. A.; Maga, G.; Mai, A. Diarylpurimidine-dihydrobenzoxopyrimidine hybrids: new, wide-spectrum anti-HIV-1 agents active at (sub)-nanomolar level. *J. Med. Chem.* **2011**, *54*, 3091–3096.

- (7) Rodriguez-Barrios, F.; Balzarini, J.; Gago, F. The molecular basis of resilience to the effect of the Lys103Asn mutation in non-nucleoside HIV-1 reverse transcriptase inhibitors studied by targeted molecular dynamics simulations. *J. Am. Chem. Soc.* **2005**, *127*, 7570–7578.

- (8) Geitmann, M.; Unge, T.; Danielson, U. H. Interaction kinetic characterization of HIV-1 reverse transcriptase non-nucleoside inhibitor resistance. *J. Med. Chem.* **2006**, *49*, 2375–2387.

- (9) Geitmann, M.; Unge, T.; Danielson, U. H. Biosensor-based kinetic characterization of the interaction between HIV-1 reverse transcriptase and non-nucleoside inhibitors. *J. Med. Chem.* **2006**, *49*, 2367–2374.

- (10) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem.* **1995**, *38*, 2681–2691.

- (11) Gonzalez, E.; Ballana, E.; Clotet, B.; Este, J. A. Development of resistance to VIR-353 with cross-resistance to the natural HIV-1 entry virus inhibitory peptide (VIRIP). *AIDS* **2011**, *25*, 1557–1583.

- (12) Gonzalez-Ortega, E.; Mena, M. P.; Permanyer, M.; Ballana, E.; Clotet, B.; Este, J. A. ADS-J1 inhibits HIV-1 entry by interacting with gp120 and does not block fusion-active gp41 core formation. *Antimicrob. Agents Chemother.* **2010**, *54*, 4487–4492.

- (13) Maga, G.; Amacker, M.; Ruel, N.; Hubscher, U.; Spadari, S. Resistance to nevirapine of HIV-1 reverse transcriptase mutants: loss of stabilizing interactions and thermodynamic or steric barriers are induced by different single amino acid substitutions. *J. Mol. Biol.* **1997**, *274*, 738–747.

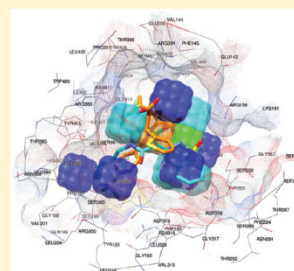
3-D QSAutogrid/R: An Alternative Procedure To Build 3-D QSAR Models. Methodologies and Applications

Flavio Ballante[†] and Rino Ragno^{*,†}

[†]Rome Center for Molecular Design, Dipartimento di Chimica e Tecnologie del Farmaco, Sapienza Università di Roma, P. le A. Moro 5, 00185, Rome, Italy

Supporting Information

ABSTRACT: Since it first appeared in 1988 3-D QSAR has proved its potential in the field of drug design and activity prediction. Although thousands of citations now exist in 3-D QSAR, its development was rather slow with the majority of new 3-D QSAR applications just extensions of CoMFA. An alternative way to build 3-D QSAR models, based on an evolution of software, has been named 3-D QSAutogrid/R and has been developed to use only software freely available to academics. 3-D QSAutogrid/R covers all the main features of CoMFA and GRID/GOLPE with implementation by multiprobe/multiregion variable selection (MPGRS) that improves the simplification of interpretation of the 3-D QSAR map. The methodology is based on the integration of the molecular interaction fields as calculated by AutoGrid and the R statistical environment that can be easily coupled with many free graphical molecular interfaces such as UCSF-Chimera, AutoDock Tools, JMol, and others. The description of each R package is reported in detail, and, to assess its validity, 3-D QSAutogrid/R has been applied to three molecular data sets of which either CoMFA or GRID/GOLPE models were reported in order to compare the results. 3-D QSAutogrid/R has been used as the core engine to prepare more than 240 3-D QSAR models forming the very first 3-D QSAR server (www.3d-qsar.com) with its code freely available through R-Cran distribution.



INTRODUCTION

The main requirements of molecular analyses today are as follows: speed, automation, optimization, and economy. Three-Dimensional Quantitative Structure–Activity Relationships (3-D QSARs) approaches are widely used and represent a viable medicinal chemistry tool whose application domain range from rationalizing a structure–activity relationship quantitatively and retrospectively to prioritizing the synthesis of molecules for synthesis and testing; its development considering the actual technology and insight sought becomes important. Till recently,¹ the well-known CoMFA² technique and the GRID³/GOLPE^{4,5} approaches were the 3-D QSAR tools most widely used in the last two decades; although successful, both these methods utilize proprietary software and require significant user interaction. The classical flowchart of a 3-D QSAR can be summarized as reported in Figure 1, and it lists the following:

- 1) Selection and alignment of Training and Test Sets
- 2) Calculation of Molecular Interaction Fields (MIFs)
- 3) Importing of Bioactivities and linking to the MIFs
- 4) Statistical evaluation
- 5) Interpretation of results by means of 2-D and 3-D plots

Excellent reviews^{6,7} of 3-D QSAR methods have been recently published; for any further details, the reader is referred to them.

Herein is described an alternative approach based on the use of open-source software to perform 3-D QSAR studies fully optimized to minimize costs, calculation time, user/computer

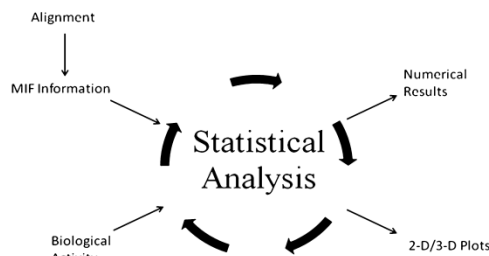


Figure 1. Overview of a classical 3-D QSAR.

interactions, and random and systematic errors. As conceived, the procedure needs only the prealigned training and test sets molecules. The protocol allows the iterative generation of hundreds/thousands 3-D QSAR models and selection of the best on the basis of conventional squared correlation (r^2), predictive cross-validation squared correlation (q^2), and standard deviation error of prediction (SDEP or root mean squared error of prediction, RMSEP) statistical coefficients.

After this project was started, another open-source method, namely OPEN3DQSAR, was reported by Tosco and Balle.¹

Received: March 7, 2012

Published: May 29, 2012

OVERVIEW OF THE NEW PROCEDURE

The new procedure is characterized by a set of new R⁸-based packages that make it possible to carry out, automatically and in parallel, 3-D QSAR studies (like CoMFA and GRID/GOLPE). Different from the previously mentioned methods, the united atom force field (FF) implemented in the AutoGrid program (AutoDock Suite⁹) was used to generate the Molecular Interaction Fields (MIFs). Upon statistic treatment of the merged biological activity and MIFs by means of the R environment and molecular graphic softwares (UCSF Chimera,¹⁰ Python Molecular Viewer¹¹ (PMV), Autodock Tools¹² (ADT)) useful 2-D (actual vs recalculated (fitting), actual vs predicted (cross validation), principal component analysis (PCA scores, loadings and scores/loadings), partial least-squares (PLS t - u and weights), and 3-D plots (actual fields, PLS-Coefficients, Activity contribution, CoMFA-Like (PLS-Coeff*StDev), and various related PLS parameters) are generated to graphically inspect, analyze, and interpret the 3-D QSAR models. Each R based package was conceived to perform specific steps ensuring high specificity, versatility, and, compared to other methods, deep optimization of the models.

Worthy of note is the included ability to determine, through a combinatorial calculation, the most appropriate pretreatment values to get preoptimized 3-D QSAR models; therefore, particular effort was given to data pretreatment and variable selection. To this aim, heavy use of the cross-validation (CV) techniques such as leave-one-out (LOO), leave-some-out (LSO), k -fold (KF), and Monte Carlo (MC) based CVs were applied either in standalone or in conjunction with a genetic algorithm (GA) as implemented in the genalg R package.¹³ Guided Region Selection (GRS) using just one probe, or a compilation of different probes (Multi Probe Guided Region Selection (MPGRS)), is a further available variable-selection method as previously reported.^{14–16}

The whole approach is described in detail below with its applications, either on ligand-based¹⁷ or structure-based¹⁸ prealigned, to molecular data sets previously reported using CoMFA and GRID/GOLPE, respectively.

At the time all the models were completed, comparison with OPEN3DQSAR software¹ was not possible due to patent restrictions that prevented CoMFA's free release in Italy.

COMPUTATIONAL METHODS

All calculations used a 6 blades (8 Intel-Xeon E5520 2.27 GHz CPU and 24 GB DDR3 RAM each) cluster (48 CPU total) running Debian GNU/Linux 5.03 64 bit operating system. The entire sequence was automated; to obtain 3-D QSAR models, the user needed only to input the prealigned data set and the values of the corresponding experimental parameters (i.e., biological activity).

Alignment Rules. The described methodology does not include an alignment engine; therefore, all the molecules contained in the data sets were used prealigned. Alignment procedures using several molecular superimposition programs are currently under investigation at the Rome Center for Molecular Design (RCMD, www.rcmd.it).

MIF Calculation. MIFs were generated using AutoGrid Software (based on the AMBER united-atom Force Field), although almost any probe can be used, in the current implementation 8 different probes (Table 1) were used in agreement with the most common residue atomic composition.

Table 1. List of the AutoGrid Probes Employed for MIF Calculation

probe type	description
A	aromatic carbon
C	aliphatic (sp ³) carbon
OA	hydrogen-bond-accepting oxygen
HD	hydrogen bonded to heteroatom
NA	hydrogen-bond-accepting amine nitrogen
N	amide nitrogen
e	electrostatic
d	desolvation

The sulfur probe (SA) was eliminated due to its close similarity to the OA (hydrogen-bond accepting oxygen) probe.

The calculated AutoGrid MIFs were imported in the R environment by means of the D2M R package (see below).

Statistical Analysis. The actual construction of statistical models was performed by a series of dedicated R packages as listed in Table 2 and arranged in Figure 2.

Table 2. List of the R Compiled Packages

R package	description
D2M	data to model
CAPP	combinatorial analysis of pretreatment parameters
MDP	model data pretreatment
CV	cross validation
VS	variable selection
GRS	guided region selection
MPGRS	multi probe guided region selection
ESP	external set prediction
YS	Y-scrambling

For a given Training Set, 3-D QSAR PLS¹⁹ models were generated according to the MIF calculations described above. Different MIFs calculated with other softwares³ could easily be imported as well.

While running the 3-D QSAR procedure, each package (named using the acronym build on the particular stage performed) was designed to achieve a statistical objective while saving the statistical information, workspace, and logs.

Package "D2M" (Data to Model). Two main steps are achieved by D2M: (1) merging of MIF(s) data with biological activities; (2) building as many raw 3-D QSAR models as the number of user-defined principal components (PC), saving the reloadable workspace and spreadsheet files containing the conventional correlation coefficients (r^2), the standard deviation errors of recalculation (SDEC), and the PLS recalculations for each PC.

Package "CV" (Cross-Validation). While the application of PLS, or other statistical techniques, to the training set are necessary to obtain a set of 3-D QSAR models, internal validation by CV is essential to assess chance correlation,²⁰ select the optimal model dimensionality (number of PCs), and measure the internal predictive ability by means of statistical coefficients such as cross-validated correlation coefficient (q^2) and standard deviation error of prediction (SDEP).

Different validation methods were included: LOO (leave-one-out), LSO (leave-some-out), KF (k -fold), and MC (Monte Carlo). For any implemented CV, several PLS calculations are performed as outlined in Table 3. All partial and final statistical

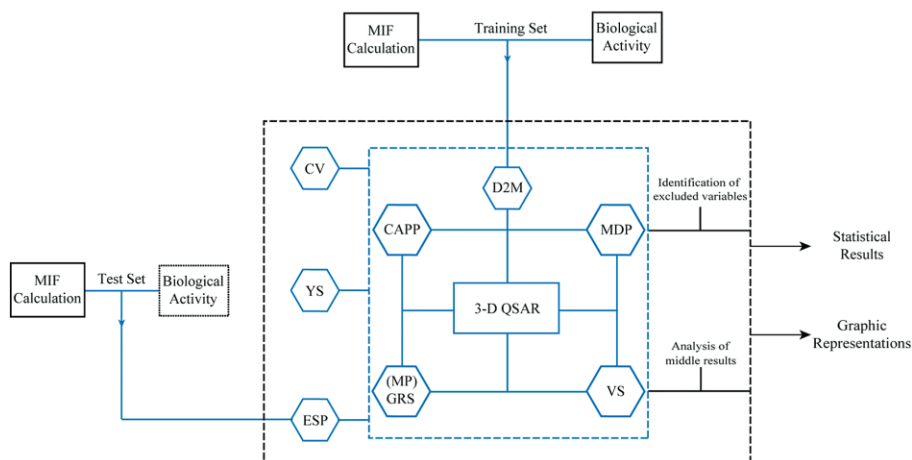


Figure 2. 3-D QSAutogrid/R process workflow. Acronyms inside the blue meshed square refer to the packages (steps) that effectively build or optimize the statistical PLS 3-D-QSAR models; others are related to packages that perform analysis on these created models (see below for the description of each package). D2M: “Data to Model”; CAPP: “Combinatorial Analysis of Pretreatment Parameters”; MDP: “Model Data Pretreatment”; CV: “Cross-Validation”; VS: “Variable Selection”; (MP)GRS: “(Multi Probe) Guided Region Selection”; ESP: “External Set Prediction”; YS: Y-scrambling.

coefficients to exhaustively analyze the cross-validation process for each method are stored and can be inspected.

Table 3. Cross-Validation (CV) Methods Implemented in the “CV” Package, with the Relative Number of PLS Calculations^a

CV method	LOO	LTO	KF	MC
CV combinations	N	$\frac{N!}{2!(N-2)!}$	$K * \text{iterations}$	iterations

^aLOO: Leave One Out; LTO: Leave Two Out, KF: k -Fold, MC: Monte Carlo; N : no. of molecules in the Training Set, K : no. of k subsamples (folds).

Package “CAPP” (Combinatorial Analysis of Pretreatment Parameters). Raw data usually need to be pretreated to minimize redundancy,²¹ even though pretreatment parameters’ values are generally arbitrarily assigned without any systematic approach, thus ignoring a possible refinement based on the specific statistical model under development. To face this issue, a methodology was developed to systematically seek the more efficient data pretreatment values (energy cutoff, zeroing of very low data points, and minimum standard deviation cutoff). The CAPP package, through combinatorial analysis for each combination of parameters’ values, builds the relative 3-D QSAR PLS model readily evaluated for each PC by different cross-validations (choice between LOO, LSO or KF). The optimal pretreatment combination is then selected according to the maximum q^2 while considering the percentage decrement value of sPRESS for each PC (Figure 3) as suggested by Gillet.²²

Package “MDP” (Model Data Pretreatment). By the means of MDP, the user filters the data set values (MIFs) either setting the pretreatment parameters in an arbitrary way or as supplied from the CAPP procedure. A further parameter not

included in the CAPP package is recognition of 1N kind of 2-level variable elimination (variables which take only 2 values in all of the data file, one of which appears only in one object). In the data pretreatment, the user can freely set which pretreatment to switch on or off. Although the logical sequence should be Field Cut-Off \rightarrow Zeroing \rightarrow SD Cut-Off \rightarrow 1N kind of 2-level variable elimination, no restriction is set to the chosen data pretreatment sequence. At each chosen pretreatment stage, the PLS is applied while saving the r^2 , SDEC, and all the recalculated vs experimental responses for each extracted PC.

Package “VS” (Variable Selection). In order to improve the predictability of the statistical model, different variable-selection procedures like D-optimal design²³ (DOD), Fractional Factorial Design²⁴ (FFD), simulated annealing²⁵ (SA), and Genetic Algorithm²⁶ (GA) are currently used in QSARs^{27–29} and 3-D QSARs.³⁰ In the VS package only the GA was actually implemented, while DOD, FFD, and SA are currently under development.

In the VS package, the GA-selection variable was implemented by the means of the R-binary genetic algorithm (genalg R package)¹³ in combination with an *ad hoc* fitness evaluation R script in which SDEP/RMSEP was used as the discriminator.

Package “GRS” (Q2-Guided Region Selection). Along with the above variable selection procedures, other approaches to improve both the robustness and goodness of the models, such as q^2 -guided region selection (q^2 -GRS) and smart-region definition (SRD), were developed.^{14,31} The q^2 -GRS procedure was implemented in this approach and, for each separate probe field, was performed by the following steps: (1) the box is divided into a user-defined number of subregions leading to many PLS submodels; (2) each generated submodel is automatically validated through one the above-described cross-validation methods (LOO, LSO, MC, or KF); (3) for each PC, only those regions displaying a q^2 value higher than a user-defined threshold value are selected, and a new 3-D QSAR

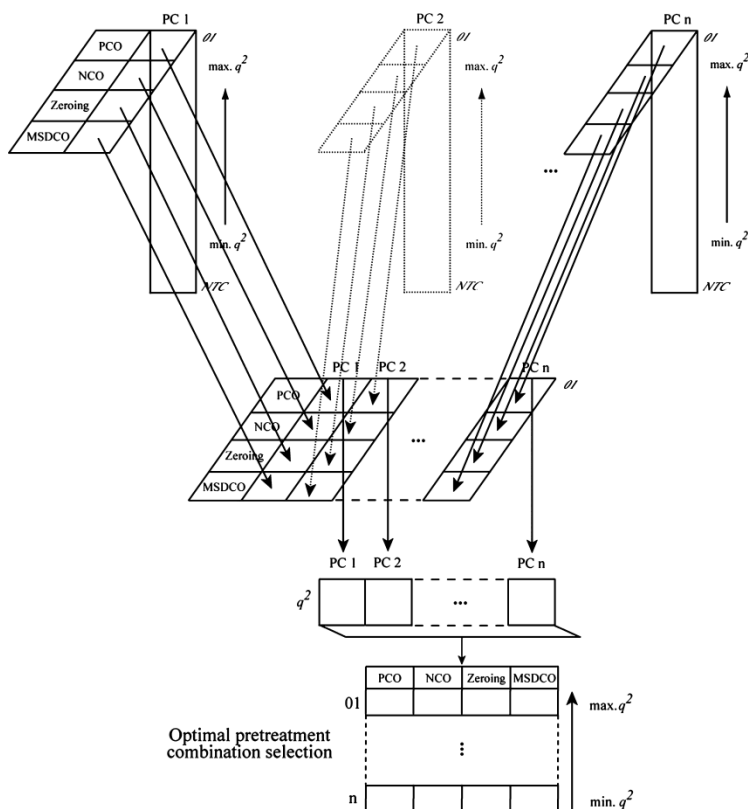


Figure 3. The CAPP Process. Through a combinatorial analysis according to predetermined setting values (as listed in Table 4), a certain number of combinations is investigated; after evaluation for each PC, the best pretreatment setting, the optimal overall-pretreatment combination is selected considering both the maximum q^2 and the percentage decrement sPRESS values.²²

model is built and cross-validated; (4) the final model is chosen at the optimal number of PC.²²

The q^2 -GRS method implemented differs from the previously reported³² as step 3 uses a bidimensional approach to select the more informative regions, so that for each dimension (PC) a different number of boxes may be retained to define the best model.

The maximum number of regions is only limited by the available amount of computer memory. Work is in progress to reduce this limitation.

Package "MPGRS" (Multi Probe Guided Region Selection). This package, an extension of the above single-probe q^2 -GRS variable selection, considers as a further dimension probe interexchange, resembling what was previously described.¹⁵ In particular, the following steps were considered: (1) starting from a series of monoprobe MIFs, for each subregion, a PLS model was built and cross-validated for i numbers of PCs (Figure 4, Step 1); (2) for the considered probe list, the obtained q^2 values (herein referred as first level q^2 , q^2_{FL}) belonging to the same region and PC were sorted. The q^2_{FL} maximum values indicated the optimal probe for each PC (called "first level" PC, PC_{FL}) and region; all values constituted

the PC_{FL} spreadsheet (Figure 4, Step 2); (3) as indicated in the PC_{FL} spreadsheet for each PC, the relative final multiprobe MIF and 3-D QSAR model were generated for q numbers of "second level" PCs (PC_{SL}) with associated q^2_{SL} ; thus, each multiprobe model was indicated by two indexes recalling both PC_{FL} and PC_{SL} ($PC_{FL:SL}$) to which corresponded a $q^2_{FL:SL}$ value (Figure 4, Step 3); (4) the optimal multiprobe 3-D QSAR model was selected according to the $q^2_{FL:SL}$ values applying the percentage decrement value of sPRESS in a bidimensional way. First were selected models at fixed first-level PCs ($PC_{FL:n}$) (first dimension), then the PC_{SL} index (second dimension) was directly retrieved from the relative q^2_{SL} values. (Figure 4, Step 4). As a result, the optimal MPGRS model was characterized by two determined values of PCs ($PC_{FL:SL}$) that implicitly contained the most informative probe for each subregion and their best combination. Notably, the final model obtained by merging the selected subregions back into a single multiprobe MIF represented a very useful tool to derive advanced 3-D QSAR studies. The same $FL:SL$ notation can also be applied to the simple GRS procedure described above.

Package "YS" (Y-Scrambling). Elimination of chance correlations of generated models was checked via the

Table 4. List of the Pretreatment Parameters (With Relative Editable Values and Number of Combinations) Analyzed by CAPP^a

parameter	max value	min value	step value	no. comb
PCO	Max _{PCO}	Min _{PCO}	Step _{PCO}	$PCO_{CB} = \left[\left(\frac{Max_{MCO} - Min_{MCO}}{Step_{MCO}} \right) + 1 \right]$
NCO	Max _{NCO}	Min _{NCO}	Step _{NCO}	$NCO_{CB} = \left[\left(\frac{ Max_{NCO} - Min_{NCO} }{Step_{NCO}} \right) + 1 \right]$
Zeroing	Max _Z	Min _Z	Step _Z	$Z_{CB} = \left[\left(\frac{Max_z - Min_z}{Step_z} \right) + 1 \right]$
MSDCO	Max _{MSDCO}	Min _{MSDCO}	Step _{MSDCO}	$MSDCO_{CB} = \left[\left(\frac{Max_{MSDCO} - Min_{MSDCO}}{Step_{MSDCO}} \right) + 1 \right]$

$$\text{no. total combinations} = PCO_{CB} \times NCO_{CB} \times Z_{CB} \times MSDCO_{CB}$$

^aPCO (Positive Cut Off); Max_{PCO} (maximum PCO value); Min_{PCO} (minimum PCO value); Step_{PCO} (incremental PCO value); PCO_{CB} (number of combinations that origins only for the PCO analysis); NCO (Negative Cut Off); Max_{NCO} (maximum NCO value); Min_{NCO} (minimum NCO value); Step_{NCO} (incremental NCO value); NCO_{CB} (number of combinations that origins only for the NCO analysis); Zeroing; zeroing of very low data points; Max_Z (maximum zeroing value); Min_Z (minimum zeroing value); Step_Z (incremental zeroing value); Z_{CB} (number of combinations that origins only for the zeroing analysis); MSDCO (Minimum SD cut-off); Max_{MSDCO} (maximum MSDCO value); Min_{MSDCO} (minimum MSDCO value); Step_{MSDCO} (incremental MSDCO value); MSDCO_{CB} (number of combinations that origins only for the MSDCO analysis).

experimental response scrambling approach.³³ The YS package allowed a user-defined number of iterations randomly coupled property/activity values to evaluate the risk of chance correlation.²⁰

Package "ESP" (External Set Prediction). As predictions are the main purpose for any QSAR-related model, validation through external test sets is mandatory. Furthermore, ESP was compiled as an independent program in place of internal validation (CV) to select the optimal number of PCs and as an extension for CAPP, GRS, and MPGRS variable selections. Such an approach allowed model optimization for external prediction.

In the current version, the ESP applied the same training-set pretreatment. In the case of (MP)GRS models, the training set selected/merged regions were retained and applied to the test set.

Graphical Results. Besides the essential role of PLS, a successful 3-D QSAR is also due to the number of graphical insights that can be generated to help interpretation of numerical results. Without graphical analyses, 3-D QSAR would be reduced to QSAR with a great number of parameters. CoMFA success measurable in more than 3340 papers citing it (SciFinder accessed February 2012) is surely due to the fact that SYBYL allowed depiction of user-friendly 3-D plots correlating structure with activity. Regarding the current method, the gnuplot style implemented in R³⁴ through the ggplot library³⁵ allowed creation of 2-D graphics score plots, loading plots, regression plots, inner-correlation plots, biplots, and many others in a straightforward way. Through an ad-hoc "in house" utility, the MIF, PLS coefficient, activity contribution 3-D plots were written in a format to be used by molecular viewers such as UCSF Chimera,¹⁰ Python Molecular Viewer (PMV),¹¹ Autodock Tools (ADT),¹² and Jmol³⁶ to generate high-quality colored molecular maps. To better interpret the 3-D QSAR model, more than one map can be overlapped to generate a complete scenario (see Figures in the application section).

RESULTS AND DISCUSSION

The application of the new 3-D QSAR procedure to a data set of aligned opioid-receptor antagonists¹⁷ (LB data set) and two data sets of HCV NSSB allosteric inhibitors¹⁸ (SB data sets) is reported.

Ligand-Based Case Studies: Opioid-Receptor Antagonists. To test the new 3-D QSAR procedure, a series of opioid-receptor antagonists, previously described by Peng et al.¹⁷ in a CoMFA application, were used to build several 3-D QSAR models. The data set was comprised of prealigned 74 compounds, separated into training and test sets with associated δ , μ , κ opioid-binding affinities.

The new method was conducted maintaining the molecular alignment used in the original paper.¹⁷ Applying a spacing grid of 1 Å and considering the binding-affinity sets of data (δ , μ , κ) with 8-molecular probes, 24 3-D QSAR models were obtained.

During the model definitions the CV was conducted via (1) Leave-One-Out (LOO), (2) Leave-Two-Out (LTO), (3) k-Fold (KF), and (4) Monte Carlo (MC) methodologies.

Initially, the raw models (Tables S1, S15, S28) were optimized through the CAPP package setting the pretreatment intervals as listed in Table 5 using the *k*-fold cross-validation with 5-random groups and 100 iterations and monitoring the q^2 and SDEP values. For the ϵ probe, several trials (data not shown) led to set a fixed Negative Cut Off (NCO) value, equal to -0.5.

A total of 12,221 combinations for each 3-D QSAR model were processed using 5% sPRESS reduction as suggested by Gillet²² to select the best combination and derive the pretreated PLS model; this led to an average q^2_{KSFVCV} value increment ranging from 26% to 55% (Tables S7, S21, S34).

The best pretreated models were then optimized through the GA-variable selection (VS package), setting the number of chromosomes, number of generation, percentage of mutation chance, and % of best individuals that are kept into the next generation to 50, 100, 0.005, and 20, respectively (Tables S6, S20, and S33). As reported in Table 6, the statistical quality of the models is similar to those obtained by the original CoMFA¹⁷ (Table 7), although the 3-D QSAutogrid/R models

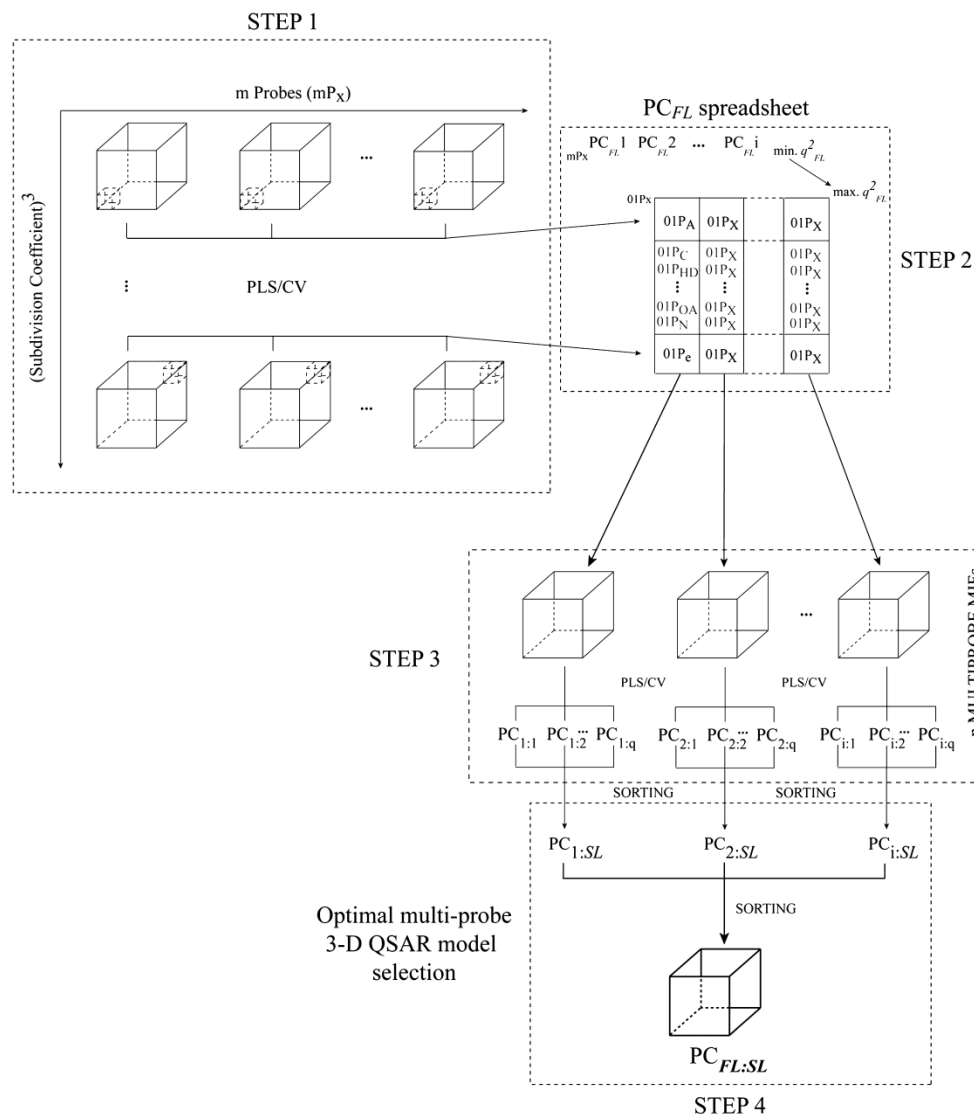


Figure 4. The MPGRS Process. For each subregion a PLS model was built and cross-validated (Step 1) in order to obtain, for each "first level" PC (PC_{FL}), the relative optimal multiprobe list (Step 2) and 3-D QSAR model characterized by an established number of "second level" PCs (PC_{SL} , Step 3). Applying a bidimensional criterion (Step 4), the optimal multiprobe 3-D QSAR model was then selected.

seem slightly more robust being described by a fewer number of PCs. For direct comparison, a further 3-D QSAR model was built using only aliphatic carbon-atom and the electrostatic probes in a single model applying only the CAPP procedure. As reported in Table 7 and Table 8, the CoMFA and the Autogrid/R double-probe models (DP) were statistically similar. Furthermore, the YS package was applied leading to

low r^2_{YS} and negative q^2_{YS} values thus supporting the lack of chance correlation in reported models.

Along with numerical output, a series of plots (scores, loadings, actual field, PLS coefficients, activity contribution and CoMFA-like) were generated to allow interpretation of the 3-D QSAutogrid/R models. Analyses of the 3-D QSAR models were conducted using all the graphical plots. Regarding the δ -opioid receptors, the CoMFA-like plots (the default TRIPOS

Table 5. CAPP Settings Adopted for the δ -, μ -, κ -Opioid Receptor Antagonists 3-D QSAR Models^a

min value	parameter	max value	step
0	PCO	50	5.0
0	Zeroing	0.1	0.01
0	MSDCO	5	0.05

^aPCO: Positive Cut Off; Zeroing: zeroing of very low data points; MSDCO: Minimum SD Cut Off.

Table 6. Opioid-Receptor Antagonists: Autogrid/R PLS Models Statistical Results (CAPP and GA Processes Were Applied)^b

model	OR	P	PC	r^2	q^2_{LOO}	q^2_{KSFCV}	r^2_{YS}	q^2_{YS}
1	δ	A	2	0.81	0.73	0.70	0.27	-0.37
2	δ	C	2	0.82	0.74	0.71	0.32	-0.35
3	δ	HD	2	0.83	0.75	0.72	0.33	-0.34
4	δ	NA	2	0.83	0.75	0.73	0.31	-0.34
5	δ	N	2	0.83	0.76	0.72	0.29	-0.32
6	δ	OA	2	0.83	0.74	0.71	0.32	-0.37
7	δ	e	3	0.69	0.58	0.56	0.22	-0.19
8	δ	d	3	0.70	0.59	0.55	0.24	-0.30
9	μ	A	3	0.91	0.82	0.76	0.57	-0.50
10	μ	C	3	0.90	0.81	0.78	0.59	-0.50
11	μ	HD	3	0.90	0.81	0.75	0.47	-0.49
12	μ	NA	3	0.91	0.81	0.78	0.59	-0.50
13	μ	N	3	0.91	0.83	0.78	0.52	-0.61
14	μ	OA	3	0.91	0.83	0.77	0.51	-0.61
15	μ	e ^a	1	0.31	0.21	0.20	0.06	-0.10
16	μ	d	3	0.72	0.60	0.52	0.27	-0.39
17	κ	A	2	0.78	0.58	0.49	0.42	-0.37
18	κ	C	3	0.81	0.62	0.55	0.54	-0.53
19	κ	HD	3	0.82	0.72	0.65	0.34	-0.41
20	κ	NA	3	0.80	0.62	0.54	0.55	-0.47
21	κ	N	3	0.80	0.61	0.52	0.54	-0.48
22	κ	OA	3	0.82	0.65	0.59	0.54	-0.44
23	κ	e ^a	2	0.35	0.20	0.18	0.13	-0.18
24	κ	d	3	0.58	0.38	0.34	0.29	-0.36

^aThe e models 15 and 23 reported were only pretreated due to too few variable after GA selection. ^bOR: Opioid-receptor data, P: Autogrid Probe, PC: optimal number of principal components/latent variables, r^2 : conventional square-correlation coefficient; q^2_{LOO} : cross-validation correlation coefficient using the leave-one-out method; q^2_{KSFCV} : cross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations; r^2_{YS} : average square correlation coefficient obtained after Y-scrambling process using 100 iterations; q^2_{YS} : average cross-validation correlation coefficient using the leave-one-out method obtained after Y-scrambling process using 100 iterations.

Table 7. Opioid-Receptor Antagonists: Original CoMFA Models Statistical Results^a

model	OR	P	PC	r^2	q^2_{LOO}	q^2_{KSFCV}
25	δ	CoMFA	4	0.91	0.69	-
26	μ	CoMFA	4	0.92	0.67	-
27	κ	CoMFA	6	0.96	0.60	-

^aOR: Opioid-receptor data; P: standard CoMFA Probe $C_{\text{sp}3}$, PC: optimal number of principal components/latent variables, r^2 : conventional square-correlation coefficient; q^2_{LOO} : cross-validation correlation coefficient using the leave-one-out method; q^2_{KSFCV} : cross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations.

Table 8. Opioid-Receptor Antagonists: Autogrid Double-Probe (DP) PLS Models Statistical Results (Only the CAPP Process Was Applied)^a

model	OR	P	PC	r^2	q^2_{LOO}	q^2_{KSFCV}	r^2_{YS}	q^2_{YS}
28	δ	Autogrid DP	3	0.83	0.70	0.67	0.41	-0.50
29	μ	Autogrid DP	4	0.85	0.65	0.63	0.52	-0.53
30	κ	Autogrid DP	3	0.84	0.67	0.63	0.50	-0.53

^aOR: Opioid-receptor data; P: Autogrid double probe (DP, C, and e probes); PC: optimal number of principal components/latent variables; r^2 : conventional square-correlation coefficient; q^2_{LOO} : cross-validation correlation coefficient using the leave-one-out method; q^2_{KSFCV} : cross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations; r^2_{YS} : average square correlation coefficient obtained after Y-scrambling process using 100 iterations; q^2_{YS} : average cross-validation correlation coefficient using the leave-one-out method obtained after Y-scrambling process using 100 iterations.

StdDev*PLS Coeff contour plots) for the C probe that allowed highlighting the molecular features indicating where sterically bulky groups were favorable (green) or unfavorable (yellow) are reported in Figure 5. A CoMFA model of the data was reproduced with a recent SYBYL version, and the related contour plots confirmed that the new procedure generated similar graphical information (compare Figures 5 and S2 with Figure S4).

Model 2 (Table 6) CoMFA-like maps (Figure 5) were in good agreement with those reported by Peng¹⁷ and displayed two green areas indicating that around bulky groups in positions 5', 6', 7' (R5 substituents of Core 1 as in ref 17) of 18, 20, 22, and 50 (Figure 5) were well tolerated. On the other hand, two yellow contours (unfavorable steric interactions) were present and give some hints to explain the reduced activities of 30 and 67 that bear bulky groups in R1 (Core 1, 30) and R3 (Core 4, 67) and that of 68 (one of the least active) which to some extent occupies both regions. The latter region was not viewable in Figure S4, likely due to differences in force field and molecular formats.

Slightly less agreement was observed between the electrostatic probe-derived plots (compare Figure 6 and Figure S3 with Figure S5 in the Supporting Information) and its CoMFA counterparts; differences were mainly located on a supplemental region that was found on the NTI indole group. Likely, these differences were surely due to the force-field differences; AutoGrid uses a united-atom force field, while CoMFA uses the all-atom TRIPOS force-field.

Similar results were obtained analyzing the models for the μ - and κ -opioid receptors; therefore, to avoid redundancy, the analyses are reported as Supporting Information (pages S13–S32).

As in the original CoMFA paper,¹⁷ the 3-D QSAutogrid/R models were externally validated using test sets (TS1) compiled from the original data sets and external test sets (TS2) compiled from different literature sources (Table 9). The predictions were similar to the original CoMFA paper and, therefore, are not commented on in detail (see pages S51–S52 in the Supporting Information).

Structure-Based Case Studies: Hepatitis C Virus NS5B-Polymerase Inhibitors. A detailed GRID/GOLPE application was reported on HCV NS5B non-nucleoside inhibitors (NNI)¹⁸ binding at two distinct allosteric sites (thumb and

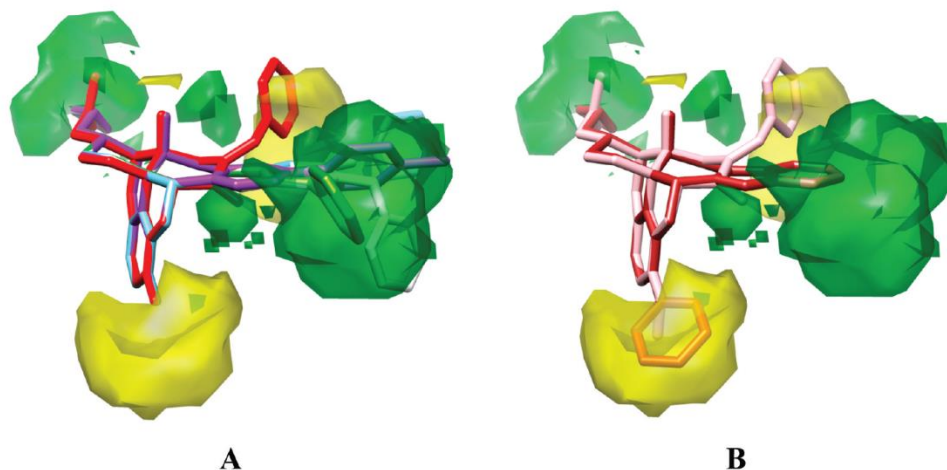


Figure 5. CoMFA-like steric-contour map derived from the C probe for the δ -opioid receptors. A: compounds 18 (sky blue), 20 (white), 22 (green), 50 (purple), and 67 (red). B: compounds 30 (brown) and 68 (pink). Contour levels: 85% (positive green, negative yellow). Hydrogen atoms are omitted for the sake of clarity. In A and B are reported the same contour maps.

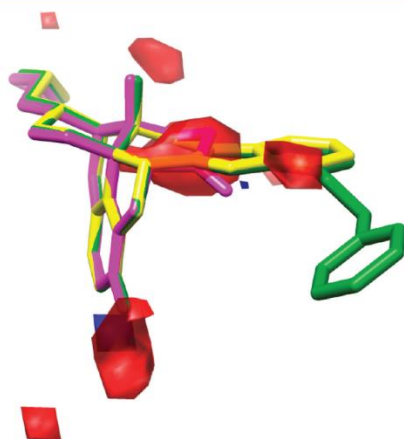


Figure 6. CoMFA-like electrostatic contour map derived from e probe for the δ -opioid receptors. Compounds: Naltrexone in magenta, NTI in yellow, 24 in green. Contour levels: 85% (positive blue, negative red). Hydrogen atoms are omitted for clarity.

palm). Thumb and palm NNIs data sets were chosen, as training sets to derive 3-D QSAR models using the new procedure with the purpose of comparing the results with those obtained by the well-established GRID/GOLPE method.¹⁸

Interaction energies between the eight probes and each molecule were computed using a grid spacing of 1 Å; thus a total of 16 3-D QSAR initial raw models were built. To build optimized models, CAPP analyses were conducted analyzing 27,775 combinations (Table S47) with the best ones chosen applying the same criteria for the previous opioid case. The CAPP procedure for both thumb- and palm-training sets

Table 9. δ Test Set Predictions Indicated by SDEP Values^a

OR model	P	PC	SDEP _{TS1}	SDEP _{TS2}
1	A	2	0.66	0.80
2	C	2	0.64	0.77
3	HD	2	0.62	0.74
4	NA	2	0.64	0.82
5	N	2	0.64	0.76
6	OA	2	0.65	0.75
7	e	3	0.81	1.20
8	d	3	0.90	1.12

^aOR model: Opioid-receptor model of Table 6; P: Autogrid probe; PC: optimal number of principal components/latent variables; SDEP_{TS1}: standard deviation error of prediction for the original test set; SDEP_{TS2}: standard deviation error of prediction for the external test set.

(Table 10) led to models comparable to those previously reported.¹⁸

Furthermore, similarly as for the LB case study, the YS package was applied leading to low r^2_{YS} and negative q^2_{YS} values thus supporting the lack of chance correlation in reported models.

Activity contribution, PLS coefficients, and CoMFA-like maps were generated, and their interpretation was in full agreement with those reported for the previous GRID/GOLPE models.

For comparison purposes, the PLS-coefficients plots obtained are shown in Figure 7 with the two methodologies, and their similarity and information content are clearly visible.

Furthermore for each SB model, the reduced test set of 21 (thumb) and 23 (palm) compounds in the original paper¹⁸ were employed to compare the predictive ability of AutoGrid/R and GRID/GOLPE. As for the statistical values, predictions were also very similar with no further comments (Table 11).

Application of Multi Probe Guided Region Variable Selection. Variable selection is an important task in 3-D QSAR in order to achieve models with an enriched data/noise ratio

Table 10. PLS Analysis Results for the Thumb- and the Palm-Structure Based AutoGrid/R and Original GRID/GOLPE C1= 3-D QSAR Models^a

data set	P	PC	r^2	q^2_{LOO}	q^2_{KSFVC}	r^2_{YS}	q^2_{YS}
thumb A		2	0.90	0.67	0.64	0.70	-0.63
thumb C		2	0.90	0.68	0.65	0.70	-0.60
thumb HD		2	0.92	0.75	0.73	0.68	-0.69
thumb NA		3	0.95	0.75	0.73	0.79	-0.66
thumb N		3	0.95	0.76	0.73	0.78	-0.67
thumb OA		3	0.95	0.77	0.73	0.77	-0.54
thumb e		3	0.98	0.58	0.52	0.92	-0.55
thumb d		1	0.58	0.36	0.36	0.27	-0.38
thumb GRID/GOLPE/ C1=		3	0.99	-	0.69	-	-
palm A		3	0.96	0.73	0.62	0.68	-1.62
palm C		3	0.96	0.73	0.62	0.69	-1.59
palm HD		1	0.90	0.75	0.71	0.44	-0.76
palm NA		2	0.97	0.62	0.52	0.84	-0.76
palm N		2	0.97	0.62	0.55	0.85	-0.87
palm OA		1	0.86	0.67	0.64	0.32	-0.66
palm e		3	0.96	0.85	0.82	0.73	-1.01
palm d		3	0.93	0.62	0.39	0.73	-1.80
palm GRID/GOLPE/ C1=		3	0.99	-	0.55	-	-

^aP: AutoGrid Probe or GRID C1= probe; PC: optimal number of principal components/latent variables; r^2 : conventional square-correlation coefficient; q^2_{LOO} : cross-validation correlation coefficient using the leave-one-out method; q^2_{KSFVC} : cross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations; r^2_{YS} : average square correlation coefficient obtained after Y-scrambling process using 100 iterations; q^2_{YS} : average cross-validation correlation coefficient using the leave-one-out method obtained after Y-scrambling process using 100 iterations.

and predictability.^{16,37} The default 3-D QSAR approaches use one or more probes distributed on regularly spatial grids without the possibility of mixing probe information into one single grid leading to a multiprobe (MP) grid. This was achieved by selecting the most informative subregions (guided region selection, q^2 -GRS package) for each considered probe so that the whole grid was reconstituted with pieces from several MIFs as described above (MPGRS package). This approach

Table 11. Thumb- and Palm-External Test Set Prediction Obtained from Structure Based AutoGrid/R and Original GRID/GOLPE C1= 3-D QSAR Models^a

data set	P	PC	SDEP _{ext}
thumb A		2	0.69
thumb C		2	0.69
thumb HD		2	0.76
thumb NA		3	0.66
thumb N		3	0.66
thumb OA		3	0.67
thumb e		3	0.63
thumb d		1	0.67
thumb GRID/GOLPE/C1=		3	0.59
palm A		3	1.14
palm C		3	1.11
palm HD		1	1.29
palm NA		2	1.04
palm N		2	1.04
palm OA		1	1.03
palm e		3	1.18
palm d		3	1.18
palm GRID/GOLPE/C1=		3	1.08

^aP: AutoGrid Probe or GRID C1= probe; PC: optimal number of principal components/latent variables; SDEP_{ext}: standard deviation error of prediction for the external test set.

was initially reported by Tropsha (modified q^2 GRS)¹⁵ and implemented in the new approach designated multi probe guided region selection (MPGRS). The MPGRS as conceived, if correctly applied, should result in a powerful modified 3-D QSAR technique; therefore, with the aim of optimizing the 3-D QSAR models, the MPGRS procedure was implemented and applied to the above case studies to test its validity and potentiality.

The MPGRS was thus applied to the three case studies [Table S12, S14, S25, S27, S38, S40, S54, S55, and S56], and as reported in Table 12, the mixed-probes models maintained a comparable level of statistical coefficients (compare Table 12 data with those of the above monoprobe 3-D QSARs: Tables S42 and S45 and Table 11).

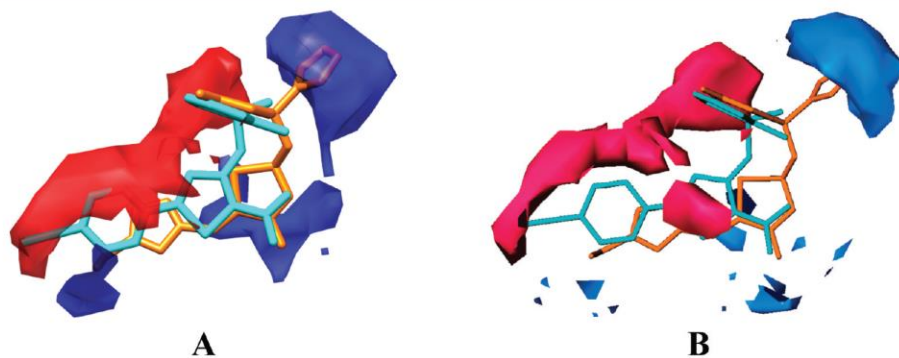


Figure 7. PLS-coefficients contour maps using the thumb-training set; only the highest active (6 in cyan) and one of the lowest active (11 in orange) compounds are shown. A: AutoGrid/R PLS coefficients contour maps derived from A probe analysis (contour levels: 60%, positive red, negative blue). B: GRID/GOLPE PLS coefficients contour maps derived from C1= GRID probe analysis (contour levels: 0.0008 red, -0.0008 blue).

Table 12. Statistical Results Obtained from MPGRS Analysis for the Thumb- and the Palm-HCV Training Sets^a

data set	MPGRS 3-D QSAR					SDEP _{ext}
	PC _{FL,SL}	r ²	q ² _{KSFVCV}	r ² _{YS}	q ² _{YS}	
thumb	2:2	0.95	0.90	0.50	-0.67	0.74
palm	1:2	0.99	0.91	0.61	-0.93	1.06

^aPC_{FL,SL}: optimal number of principal first level (FL) and second level (SL) components/latent variables for the MPGRS model; r²: conventional square-correlation coefficient; q²_{LOO}: cross-validation correlation coefficient using the leave-one-out method; q²_{KSFVCV}: cross-validation correlation coefficient using the *k*-fold cross-validation with 5 random groups and 100 iterations; r²_{YS}: average square correlation coefficient obtained after Y-scrambling process using 100 iterations; q²_{YS}: average cross-validation correlation coefficient using the leave-one-out method obtained after Y-scrambling process using 100 iterations.

All the MPGRS models were analyzed; of particular interest were the SB-derived alignments that checked for the MPGRS ability to propose a pseudoreceptor. Therefore, the detailed analyses is reported for the HCV palm-training set.

Applying q² threshold value of 0.4, 11-MIFs subregions were selected (Figure 8) to build the multiprobe MIF, and were color coded according to that reported in Table S53. In particular, 5 regions were taken from the N MIF, 4 from the NA, and the last two from HD and e probes, respectively. High agreement between the selected regions and the HCV NSSB-palm binding pocket surface was observed. These 11 subregions

were highly informative to allow a very detailed interpretation of the final MPGRS 3-D QSAR model (Figure 8).

Furthermore, by analyzing the selected subregions' PLS coefficients, a series of pharmacophoric-like points were extrapolated (Figures 9 and S20). According with the relative

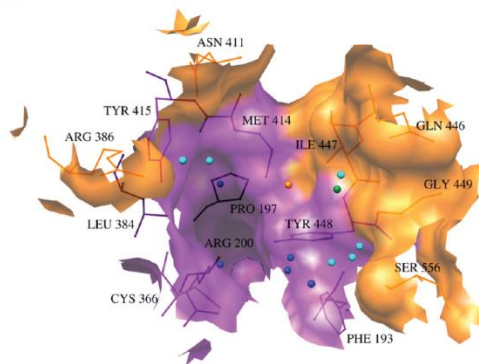


Figure 9. MPGRS 3-D QSAR palm model key points. The points are color coded: in blue N (amide nitrogen) probe key points, in cyan those from NA (hydrogen acceptor nitrogen) probe, in green and orange those from HD (hydrogen donor) and e (electrostatic) probes, respectively.

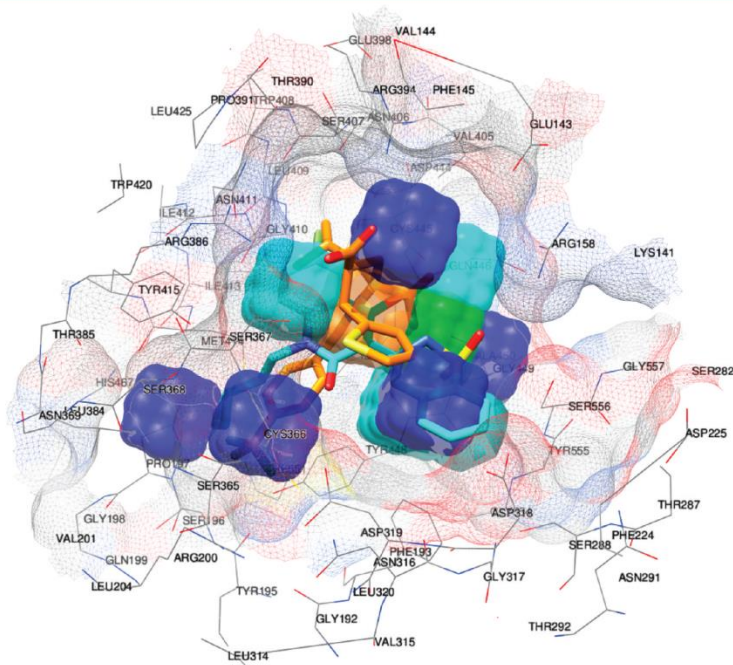


Figure 8. Most informative subregions derived from the final MPGRS 3-D QSAR palm model. The regions are color coded: in blue N, in cyan NA, in green HD, and in orange e.

probe subregion, similar signs, as in the GRID/GOLPE approach, could have a different meaning, and correct interpretation remains problematic.

The five N (amidic nitrogen) probe subregions in Figure 8 indicate the localization of hydrophobic interactions, and their positions in the palm allosteric-binding site were indeed characterized by the nonpolar residues Phe193, Pro197, Arg200, Cys366, Leu384, Met414, Tyr415, and Tyr448, forming a hydrophobic pocket (violet surface in Figure 9) in agreement with that previously reported.^{38–40}

The 4 NA (hydrogen acceptor nitrogen)-associated key points overlapped with residues bearing hydrogen-acceptor groups^{38–40} (Arg386, Asn411, Gln446, Ile447, Gly449, Tyr448, Ser555, orange surfaces in Figure 9), while the single green HD (hydrogen donor) region was in proximity with the Tyr448 and Gly449 main-chain nitrogens.^{39,40} The last electrostatic subregion presented difficulties of interpretation, and no specific role was assigned.

Regarding the internal predictive ability of the MPGRS models, the multiprobe approach in general was not improved; nevertheless, the interpretation of the model was greatly enhanced. MPGRS allowed focusing on the most informative regions around the ligands and used all the probes together to reduce the chances of missing important correlations when using single probe 3-D QSARs.

CONCLUSION

The use of the AutoGrid software coupled with ad-hoc R-based scripts allows an alternative procedure (3-D QAutogrid/R) to generate 3-D QSAR analyses, similarly as the well-established CoMFA and GRID/GOLPE techniques, improving both the use of chemical data and minimizing time and human-machine interactions. The procedure was validated with three data sets, covering both ligand-based and structure-based alignment methodologies. The main features of the new procedure are automation and flexibility that permit the iterative generation of hundreds/thousands of 3-D QSAR models selecting the best one in a completely independent way and improving the amount of important information generated from detailed 3-D QSARs analyses. Furthermore, the possibility to extrapolate/merge the more informative interactions from different probe fields into a single multiprobe MIF lead to more comprehensive interpretations. Case studies results and comparisons with the other mentioned methods show how the new procedure should be a useful tool, based on free software, to conduct advanced 3-D QSAR analyses. The implementation of the MPGRS, although not improving the models overall predictive abilities, greatly enhanced their interpretation. To the best of the author's knowledge, this is the very first free MPGRS implementation. Furthermore, 3-D QAutogrid/R has been recently used as core engine to prepare more than 240 3-D QSAR models used to generate the first 3-D QSAR server^{41–43} (www.3d-qsar.com).

All the described R packages are available through the CRAN-package repository; tutorials with example files are also available through the www.3d-qsar.com and www.rcmd.it Web sites.

ASSOCIATED CONTENT

Supporting Information

Full statistical results for all the data sets with example of GRS, 3-D QAutogrid/R double-probe, and bidimensional plots. Graphical comparisons between the new method and the

original CoMFA analyses for the opioid data sets. MPGRS statistical values for all the data sets and related plots. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: rino.ragno@uniroma1.it.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Many thanks to Prof. Garland R. Marshall (Department of Biochemistry and Molecular Biophysics, Washington University in St. Louis) for critical reading of the manuscript, helpful discussions, and encouragement and also for the use of TRIPOS software in his laboratory. Many thanks are also due to Youyi Y. Peng for having provided the opioid-receptor antagonists data sets. This study was supported by grants from Italian Ministry of University and Research (MIUR Grant 2008F8T894_002 and 2008ZTN724_003).

REFERENCES

- (1) Tosco, P.; Balle, T. Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. *J. Mol. Model.* **2011**, *17*, 201–208.
- (2) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, S959–S967.
- (3) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- (4) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear Pls Estimations (Golpe) - an Advanced Chemometric Tool for Handling 3d-Qsar Problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9–20.
- (5) Cruciani, G.; Watson, K. A. Comparative Molecular-Field Analysis Using Grid Force-Field and Golpe Variable Selection Methods in a Study of Inhibitors of Glycogen-Phosphorylase-B. *J. Med. Chem.* **1994**, *37*, 2589–2601.
- (6) Verma, J.; Khekar, V. M.; Coutinho, E. C. 3D-QSAR in drug design--a review. *Curr. Top. Med. Chem.* **10**, 95–115.
- (7) Sippl, W. 3D-QSAR - Applications, Recent Advances, and Limitations. In *Recent Advances in QSAR Studies*; Puzyn, T., Leszczynski, J., Cronin, M. T., Eds.; Springer: Netherlands: Vol. 8, pp 103–125.
- (8) Team, R. D. C. *The R Foundation for Statistical Computing*. <http://www.r-project.org/> (accessed month day, year).
- (9) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.
- (10) The University of California, S. F. U. UCSF chimera. <http://www.cgl.ucsf.edu/chimera/> (accessed month day, year).
- (11) Python Molecular Viewer (PMV). Molecular Graphics Laboratory, T. S. R. I. <http://mglttools.scripps.edu/> (accessed month day, year).
- (12) AutoDockTools (ADT). Molecular Graphics Laboratory, T. S. R. I. <http://mglttools.scripps.edu/> (accessed month day, year).
- (13) Willighagen, E. *genalg: R Based Genetic Algorithm*, 0.1.1.; 2005.
- (14) Cho, S. J.; Tropsha, A. Cross-validated R2-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results. *J. Med. Chem.* **1995**, *38*, 1060–1066.
- (15) Cho, S. J.; Tropsha, A.; Suffness, M.; Cheng, Y. C.; Lee, K. H. Antitumor agents. 163. Three-dimensional quantitative structure-activity relationship study of 4'-O-demethylepipodophyllotoxin analogs

- using the modified CoMFA/q2-GRS approach. *J. Med. Chem.* **1996**, *39*, 1383–1395.
- (16) Cruciani, G.; Clementi, S.; Pastor, M. GOLPE-guided region selection. *Perspect. Drug Discovery Des.* **1998**, *12–14*, 71–86.
- (17) Peng, Y.; Keenan, S. M.; Zhang, Q.; Kholodovych, V.; Welsh, W. J. 3D-QSAR comparative molecular field analysis on opioid receptor antagonists: pooling data from different studies. *J. Med. Chem.* **2005**, *48*, 1620–1629.
- (18) Musmuca, I.; Caroli, A.; Mai, A.; Kaushik-Basu, N.; Arora, P.; Ragno, R. Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NSSB polymerase inhibitors. *J. Chem. Inf. Model.* **2010**, *50*, 662–676.
- (19) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J., III. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735–743.
- (20) Clark, M.; Cramer, R. D. The Probability of Chance Correlation Using Partial Least-Squares (PLS). *Quant. Struct.-Act. Relat.* **1993**, *12*, 137–145.
- (21) Cruciani, G. *Molecular interaction fields: applications in drug discovery and ADME prediction*; Wiley-VCH: Weinheim, 2006; p xviii, 307 p.
- (22) Wold, S.; Johansson, E.; Cocchi, M. *PLS: Partial Least Squares Projections to Latent Structures in 3D QSAR in Drug Design: Theory Methods and Applications*; ESCOM Science Publishers: 1993.
- (23) Mitchell, T. J. An algorithm for the construction of "D-optimal" experimental designs. *Technometrics* **2000**, *42*, 48–54.
- (24) Box, G. E. P.; Hunter, W. G.; Hunter, J. S.; *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*; John Wiley & Sons: 1978; p 653.
- (25) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680.
- (26) Holland, J. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*; A Bradford Book: 1992.
- (27) Baroni, M.; Clementi, S.; Cruciani, G.; Kettanehworld, N.; Wold, S. D-Optimal Designs in Qsar. *Quant. Struct.-Act. Relat.* **1993**, *12*, 225–231.
- (28) Puzyn, T.; Leszczynski, J.; Cronin, M. T. D. *Recent advances in QSAR studies: methods and applications*; Springer: Dordrecht; New York, p xiv, 423 p.
- (29) Shen, M.; LeTiran, A.; Xiao, Y. D.; Golbraikh, A.; Kohn, H.; Tropsha, A. Quantitative structure-activity relationship analysis of functionalized amino acid anticonvulsant-agents using k nearest neighbor and simulated annealing PLS methods. *J. Med. Chem.* **2002**, *45*, 2811–2823.
- (30) Kubinyi, H.; Folkers, G.; Martin, Y. C. *3D QSAR in drug design*; Kluwer Academic: Dordrecht; Boston, MA, 1998; p v. < 2- >.
- (31) Pastor, M.; Cruciani, G.; Clementi, S. Smart region definition: a new way to improve the predictive ability and interpretability of three-dimensional quantitative structure-activity relationships. *J. Med. Chem.* **1997**, *40*, 1455–1464.
- (32) Cho, S. J.; Tropsha, A. Cross-validated R²-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results. *J. Med. Chem.* **1995**, *38*, 1060–1066.
- (33) Wold, S.; Eriksson, L. *Chemometrics Methods in Molecular Design*; VCH: Weinheim: 1995.
- (34) Crawley, M. J. *The R book*; Wiley: Chichester, England; Hoboken, NJ, 2007; p viii, 942 p.
- (35) Wickham, H. *ggplot2: elegant graphics for data analysis*; Springer: New York, 2009.
- (36) *Jmol: an open-source Java viewer for chemical structures in 3D*.
- (37) Cruciani, G.; Clementi, S.; Baroni, M. Variable Selection in PLS Analysis. In *3D QSAR in Drug Design*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 551–564.
- (38) Tedesco, R.; Shaw, A. N.; Bambal, R.; Chai, D.; Concha, N. O.; Darcy, M. G.; Dhanak, D.; Fitch, D. M.; Gates, A.; Gerhardt, W. G.; Halegoua, D. L.; Han, C.; Hofmann, G. A.; Johnston, V. K.; Kaura, A. C.; Liu, N.; Keenan, R. M.; Lin-Goerke, J.; Sarisky, R. T.; Wiggall, K. J.; Zimmerman, M. N.; Duffy, K. J. 3-(1,1-dioxo-2H-(1,2,4)-benzothiazin-3-yl)-4-hydroxy-2(1H)-quinolinones, potent inhibitors of hepatitis C virus RNA-dependent RNA polymerase. *J. Med. Chem.* **2006**, *49*, 971–983.
- (39) Li, T.; Froeyen, M.; Herdewijn, P. Insight into ligand selectivity in HCV NSSB polymerase: molecular dynamics simulations, free energy decomposition and docking. *J. Mol. Model.* **2010**, *16*, 49–59.
- (40) Ryu, K.; Kim, N. D.; Il Choi, S.; Han, C. K.; Yoon, J. H.; No, K. T.; Kim, K. H.; Seong, B. L. Identification of novel inhibitors of HCV RNA-dependent RNA polymerase by pharmacophore-based virtual screening and in vitro evaluation. *Bioorg. Med. Chem.* **2009**, *17*, 2975–2982.
- (41) Ballante, F.; Musmuca, I.; Patsilnakos, A.; Ragno, R. An Alternative Method for Generating 3-D QSAR Models using Free Software. In *5th Joint Sheffield Conference on Chemoinformatics*; Sheffield, UK, 2010.
- (42) Patsilnakos, A.; Ballante, F.; Musmuca, I.; Ragno, R. 3-D QSAR SERVER – A 3-D QSAR Models Database for Virtual Screening. In *14th Hellenic Symposium on Medicinal Chemistry*; Thessaloniki, Greece, 2010.
- (43) Musmuca, I.; Ballante, F.; Ragno, R. R/AUTOGRID/ADT Combination As An Alternative To Build 3-D QSAR Models. Methodologies And Applications. In *18th European Symposium on Quantitative Structure-Activity Relationships*; Rhodes, Greece, 2010.

Comprehensive model of wild-type and mutant HIV-1 reverse transcriptases

Flavio Ballante · Ira Musmuca · Garland R. Marshall · Rino Ragno

Received: 13 December 2011 / Accepted: 28 June 2012 / Published online: 26 July 2012
© Springer Science+Business Media B.V. 2012

Abstract An enhanced version of COMBINE that uses both ligand-based and structure-based alignment of ligands has been used to build a comprehensive 3-D QSAR model of wild-type HIV-1 reverse transcriptase and drug-resistant mutants. The COMBINER model focused on 7 different RT enzymes complexed with just two HIV-RT inhibitors, niverapine (NVP) and efavirenz (EFV); therefore, 14 inhibitor/enzyme complexes comprised the training set. An external test set of chiral 2-(alkyl/aryl)amino-6-benzylpyrimidin-4(3H)-ones (DABOs) was used to test predictability. The COMBINER model MC4, although developed using only two inhibitors, predicted the experimental activities of the test set with an acceptable average absolute error of prediction (0.89 pK_i). Most notably, the model was able to correctly predict the right eudismic ratio for two *R/S* pairs of DABO derivatives. The enhanced COMBINER approach was developed using only software freely available to academics.

Keywords 3-D-QSAR · HIV-1 reverse transcriptase · Drug resistance · NNRTI · RT mutants · Molecular modeling · COMBINER · DABO inhibitors · PLS

F. Ballante · I. Musmuca · G. R. Marshall · R. Ragno
Rome Center for Molecular Design, Dipartimento di Chimica e
Tecnologie del Farmaco, Sapienza Università di Roma, P. le A.
Moro 5, 00185 Rome, Italy
e-mail: rino.ragno@uniroma1.it

G. R. Marshall (✉)
Department of Biochemistry and Molecular Biophysics,
Washington University School of Medicine, St. Louis, MO
63110, USA
e-mail: garlandm@gmail.com

Introduction

When faced with an abundance of data from diverse sources of structure-activity studies, for example, SAR data on multiple drug-resistant mutants of a therapeutically important enzyme, choosing the best paradigm for an integrative analysis is difficult. An example common to most HIV antivirals is an abundance of crystal structures of diverse inhibitor complexes with both wild-type and mutant enzymes. The design of the next generation of inhibitors with enhanced ability to withstand the impact of known mutations should emerge from a 3-D QSAR analysis that integrates common features found in the diverse structure-activity studies within the structural 3-D framework of the enzyme. HIV-1 reverse transcriptase (HIV-RT) continues to be of therapeutic interest in the ongoing effort to provide AIDS therapeutics that have improved efficacy against multiple drug-resistant mutants, and provides a case in point for application of modern 3-D QSAR techniques.

The paradigm chosen for modeling of HIV-RT and several drug-resistant mutants was an enhanced 3-D QSAR version of the original COMBINE approaches of Gago et al. [1–3]. This Roman reincarnation, referred to in this manuscript hereafter as COMBINER to avoid confusion, used both ligand-based and structure-based alignment of inhibitors as pioneered by Musmuca et al. [4] in a previous study of hepatitis C virus NS5B polymerase inhibitors. The basic premise is that all experimental information is relevant if an appropriate frame of reference can be found, in this case, the 3-D-structure of the parent enzyme. An additional goal was the exclusive use of software readily available to the academic community to allow testing and validation of this approach by others. In a study by the Mai group, Rotili et al. [5] generated data for novel HIV-RT

inhibitors that was used as an external test set for the current iteration of the COMBINER model.

Methodology

Molecular modeling, COMBINER, and docking calculations

All molecular modeling calculations were performed on a 6-blade (8 Intel-Xeon E5520 2.27 GHz CPU with 24 GB DDR3 RAM each) cluster (48 CPU total) running the Debian GNU/Linux 5.03 operating system. The experimental activities of EFV and NVP reported by Rotili et al. [5] were performed as previously described [6, 7]. To build the non-experimental complexes, a cross-docking procedure previously described [4] was used with the AutodockVina program. Docking assessments were checked for either Autodock 4.2 or AutodockVina 1.1, root-mean-square-deviation (RMSD) errors are reported in Table 3.

All complexes were arbitrary superimposed using a reference template 1vrt, chosen for its superior crystallographic resolution ($R = 2.2 \text{ \AA}$). Superimpositions of the HIV-RT complexes were made with Chimera [8] using the command-line implementation of MatchMaker [9]. Prior to any minimization, all crystallographic waters were discarded by a procedure previously described [10–12]. Hydrogen atoms were added using the tleap module of AMBER [13]. Protonation states at pH 7.4 were selected, i.e., lysines, arginines, aspartates, and glutamates were assumed to be in the ionized form and parameters were calculated by means of the Antechamber module of AMBER. The complexes were solvated (SOLVATEOCT command) in a box extending 10 \AA with water molecules (TIP3 model) and neutralized with Na^+ and Cl^- ions. The solvated complexes were then refined by a single-point minimization using the Sander module of AMBER. The minimized complexes were realigned with MatchMaker using the same reference complex while maintaining the coordinates (experimental alignments) into ligands (key) and proteins (lock) and were used to obtain the energy-deconvolution matrix to develop the COMBINER models. Using Autogrid4 [14], three contributing energetic fields were calculated: the electrostatic (ELE), the steric (STE) and the desolvation (DRY). As HIV-RT is composed of 1,000 residues, 1,000 COMBINE descriptors were calculated for each field. Seven combination of the field were examined (ELE, STE, DRY, ELE + STE, ELE + DRY, STE + DRY and ELE + STE + DRY). The PLS algorithm as implemented in the R [15] environment, was used with an in-house script to compute all statistical calculations and cross-validations (Table 2).

Results and discussion

COMBINER model

To build the COMBINER model, training set selection was driven by both the availability of co-crystal structures and homogeneous inhibition data from the Mai lab. From a literature search, 14 complexes (characterized by 7 different HIV-RT wild-type and mutant enzymes) were selected as a training set using complexes with only two HIV-RT inhibitors, NVP and EFV, for which inhibition constants were available as previously tested by our collaborators [4].

As reported in Table 1, the training set was composed of NVP and EFV in complex with seven different HIV-RT enzymes (WT, L100I, K103N, V106A, V179D, Y181I, Y188L). Of the 14 complexes, structural data were experimentally available from the PDB for only five (WT/EFV: 1fk9 [16], K103N/EFV: 1fko [16], WT/NVP: 1vrt [17], L100I/NVP: 1s1u [18], and K103N/NVP: 1fkp [16]). The other nine complexes (L100I/EFV, V106A/NVP, V106A/EFV, V179D/NVP, V179D/EFV, Y181I/NVP, Y181I/EFV, Y188L/NVP and Y188L/EFV) were directly modeled using side-chain structural information retrieved from other complexes present in the PDB and using the BUILD module of UCSF Chimera.

Different from the original COMBINE protocol, COMBINER used the Autogrid module of the AutoDock 4 suite [14] to compute the energy interactions between the inhibitors and each amino-acid residue of the enzyme in a complex. The ligand/residues/energy deconvolution matrix was directly obtained by the sum of the interaction energies between all ligand atoms and those composing each amino acid residue in HIV-RT. The complexes were optimized by a short energy minimization followed by docking experiments conducted with AutoDockVina [19]. From the Autogrid application, three kinds of interaction contributions were calculated: the STE, the ELE and the desolvation (DRY) ones. HIV-1 RT is a heterodimer with a subunit of 560 residues (p66) and a second subunit (p51) of 440 residues. Therefore, for each contribution, a total of 1,000 interactions were computed, and modeled using the PLS algorithm implemented in the R [20] environment. Considering all possible combination of contributions, seven different COMBINER models were independently derived (CM1–CM7, Table 2). From data reported in Table 2, all seven COMBINER models were highly robust and endowed with good predictive power. Among the seven models, CM1 and CM4 (Fig. 1) exhibited the best statistical-value profiles (compare r^2 , q^2 and SDEP values in Table 2).

As discussed by Gago et al. [2, 3] and common to other 3-D QSAR studies [21, 22], COMBINE-like models have to be analyzed by means of PLS coefficients and activity contribution (interaction energies multiplied by the PLS

Table 1 Structures, anti-HIV-RT activities (μM) of nevirapine (NVP) and efavirenz (EFV) used to build the COMBINER models

RT	NVP	EFV
WT	0.4	0.03
L100I	9.0	0.12
K103N	7.0	0.16
V106A	10.0	0.04
V179D	2.0	0.10
Y181I	36.0	0.15
Y188L	18.0	0.38

Table 2 Statistical coefficients of the COMBINER models

CM	Model	r^2	SDEC	q_{LOO}^2	SDEP _{LOO}	q_{LS05}^2	SDEP _{LS05}	q_{LS02}^2	SDEP _{LS02}
1	DRY	0.91	0.31	0.82	0.43	0.79	0.46	0.63	0.58
2	ELE	0.80	0.45	0.51	0.71	0.49	0.72	0.37	0.79
3	STE	0.81	0.44	0.69	0.57	0.65	0.60	0.52	0.68
4	DRY_STE	0.88	0.35	0.78	0.48	0.75	0.50	0.61	0.61
5	ELE_STE	0.82	0.43	0.58	0.66	0.53	0.69	0.44	0.75
6	DRY_ELE	0.89	0.34	0.66	0.59	0.63	0.62	0.48	0.70
7	DRY_ELE_STE	0.86	0.38	0.66	0.59	0.62	0.62	0.50	0.70

CM COMBINER model number, r^2 conventional squared-correlation coefficient, SDEC standard error of calculation, q^2 cross-validation coefficient, LOO leave-one-out, SDEP standard error of prediction, LS05 and LS02 leave-some-out using 5 and 2 groups respectively

coefficients) plots. While PLS coefficients indicated which residues contributed most to the COMBINE relationships (general indication), the activity contributions provided the real pK_i contribution for each inhibitor/residue pair to the enhancement or decrease of the given inhibitor activity starting from a constant threshold value (intercept). Further indications of significance can be inferred from the PLS coefficients weighted by the standard deviation values (PLS \times SD) to give the overall importance of each amino-acid residue in the COMBINER model. In Figs. 2 and 3 are reported the PLS coefficients, the PLS \times SD and activity-contribution histograms for CM1 and CM4 models, respectively.

Regarding the desolvation energy (DRY), from Figs. 2A and 3A, residues Leu100 (Ile100), Lys101, Lys103 (Asn103), Val106 (Ala106), Val179 (Asp179), Tyr181 (Ile181), Tyr188 (Leu188), Trp229, Leu234 and Tyr318

are mainly involved in defining either model CM1 or model CM4. As suggested by Wesson and Eisenberg [23], DRY is proportional to the change in the surface area that is available to water, therefore, the DRY energies are an estimation of the hydrophobic effect similar to the DRY probe in the Goodford GRID program [24]. The DRY interactions have only positive values; therefore, multiplication of the PLS value by the standard deviation of a certain residue can be interpreted in the same way as the 3-D-QSAR CoMFA [25] plots in which positive PLS Coeff \times SD values are directly correlated with enhanced activity and negative values correlate with decreased biological affinities (Fig. 2B). In Fig. 2B, residues Leu100 (Ile100), Lys101 and Tyr188 (Leu188) have the highest PLS Coeff \times SD values and, therefore, interaction with these residues are desirable, while low negative PLS Coeff \times SD values are associated with residues Trp229 and

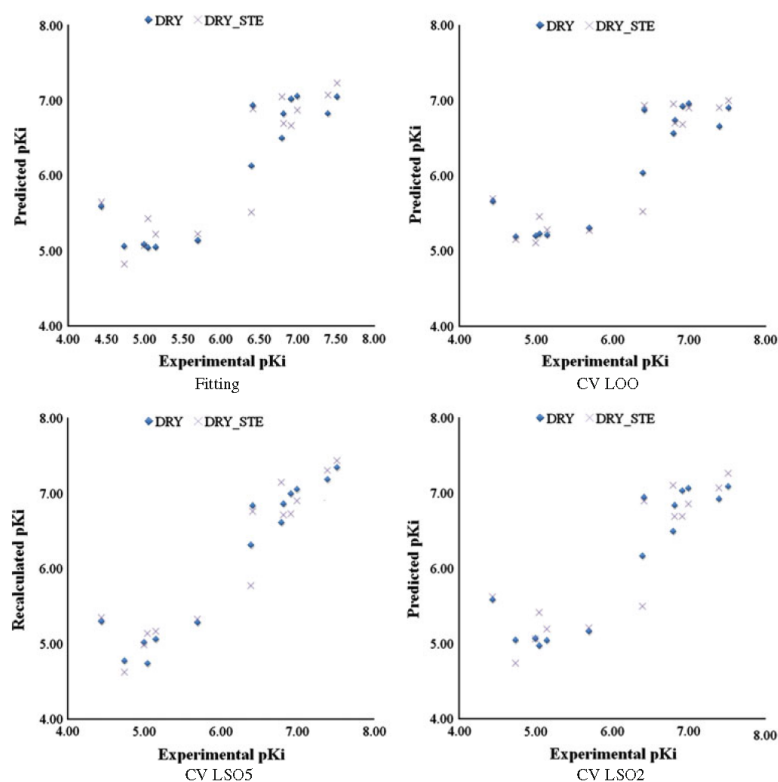


Fig. 1 Fitting and cross-validation plots (*LOO*, *LSO5* and *LSO2*) recalculate/experimental and predicted/experimental pK_i for COMBINER models CM1 and CM4

Leu234 meaning that the interaction with these residues should be minimized. Observing Fig. 3a, in model CM4, residues Leu100 (Ile100), Lys101 and Tyr188 (Leu188) are more sensitive to STE interactions, in agreement with the above. On the other hand, investigation on the energy of interaction on the STE field revealed that almost only negative values are present, in agreement with the fact that the 14 complexes were generated by means of docking experiments with van der Waals and hydrogen-bonding optimized. Thus the significance of the PLS Coef*StDev bars of histogram in Fig. 3B relative to the STE fields have inverse signification to those of the DRY fields. Although some redundancy occurs in the Autogrid-field calculations, the fact that the charge of the atom is incorporated in the calculation of desolvation interactions and that the STE fields is the sum of the interactions of the residue atoms, thus containing also hydrogen-bonding calculations, the DRY and the STE field together contain most of the ELE interactions. Similar analyses were also done for the ELE

(CM2), STE (CM3), ELE_STE (CM5), DRY_ELE (CM6) and the triple field containing COMBINER model CM7. In all COMBINER models containing the ELE field merged with other fields, its contribution to the description was almost negligible. As a matter of fact, the CM2 models (only ELE) had lower statistical coefficients, thus, indicating a lower correlation between the biological activities and ELE interactions. In the multifield models (CM4–CM7), therefore, the PLS code correctly recognized this low correlation and contribution of the ELE field was essentially eliminated. Since the models were obtained using single point RT-mutated forms, interesting sources of data are the activity-contribution plot of Figs. 2C and 3C. These plots reported the product of each residue field by the respective PLS coefficients. The sum of all these products and the intercept values for each complex returns the fitted values of the COMBINER models (Fig. 8). Due to the similar profile of the DRY field in both CM1 and CM4 models, only the DRY_STE double-field model is

Fig. 2 PLS coefficients **A**, PLS \times SD **B** and activity-contribution histograms **C** for the DRY model CM1. Only bars with values higher than 0.001 and lower than -0.001 are shown

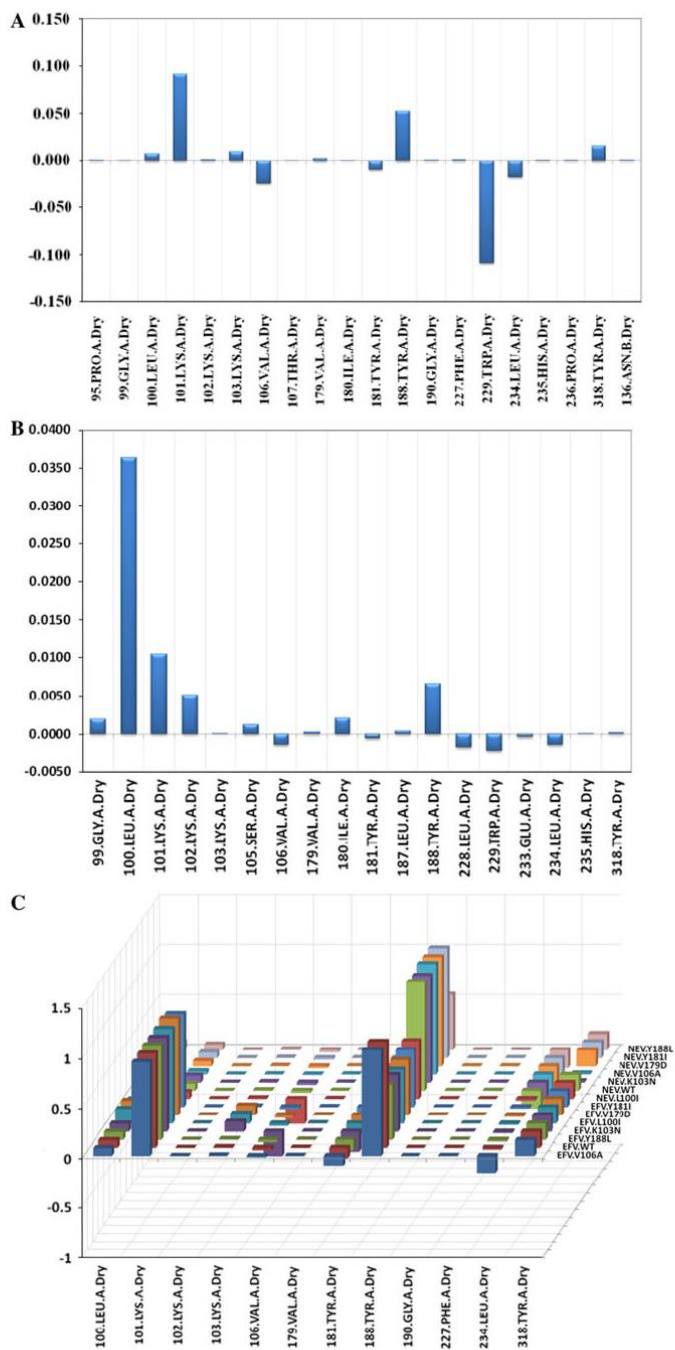
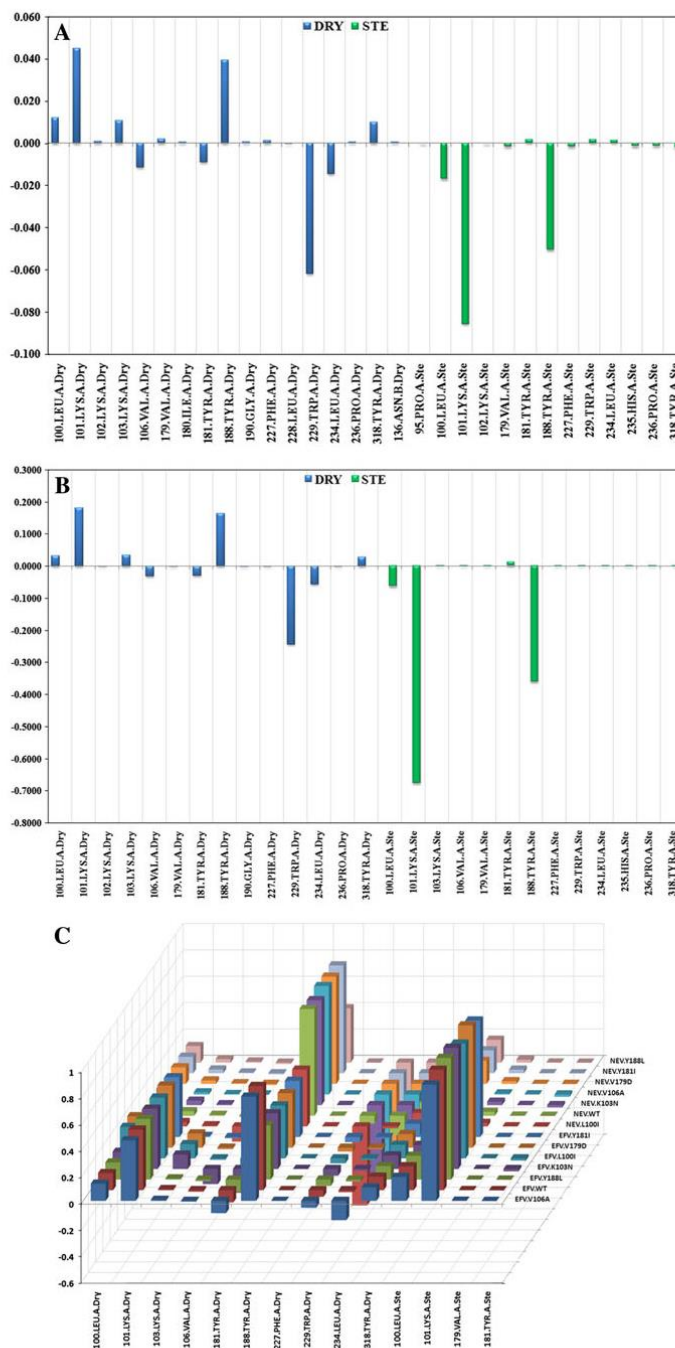


Fig. 3 PLS coefficients
A, PLS \times SD **B** and activity-
 contribution histograms **C** for
 DRY_STE model CM4. Only
bars with values higher than
 0.001 and lower than -0.001
 are shown



considered for future comments. It could be argued that all statistics of the DRY model are slightly better or comparable to those of the DRY_STE model. It was decided, nevertheless, to focus only on the DRY_STE model so to have a more complete description of the ligand/enzyme interactions. Analyses of activity-contribution plots confirmed that the amino-acids mutations were directly and indirectly responsible for the different activity profiles of EFV and NVP. Any description of the detailed interaction network is far too complicated; after analysis of the CM1 and CM4 models plots reported in Figs. 2 and 3, a

schematic view (Fig. 4) on the direct influence to the NVP and EFV anti-RT activities by their surrounding residues (and their mutations) is presented.

COMBINER predictions

The reported COMBINER model CM4 was used to rationalize the role of mutation on the activity profile of (*R*)- and (*S*)-MC1501, and of (*R*)- and (*S*)-MC2082 reported by Rotili et al. [5]. The binding modes of the four DABO derivatives (Fig. 5) were analyzed by the means of the

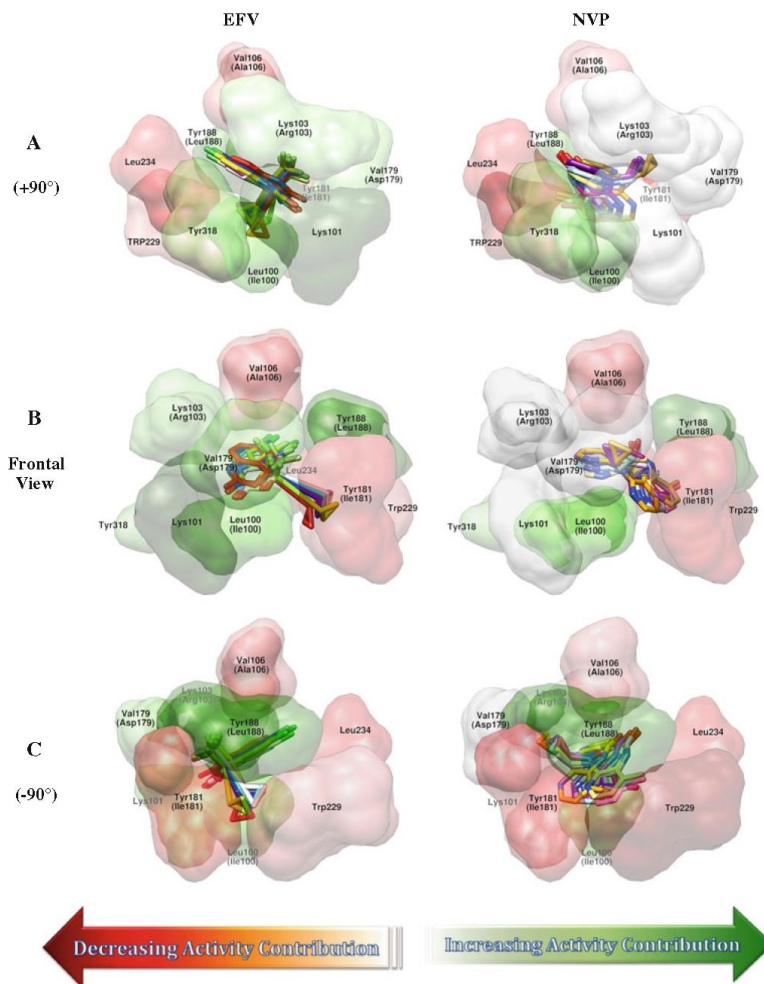


Fig. 4 Efavirenz (*left column*) and nevirapine (*right column*) with the surrounding residues surfaces as in the experimental complexes. The surfaces are colored by activity contribution. Here are three orthogonal views of the complexes (rotated along the *X* axes by $\pm 90^\circ$)

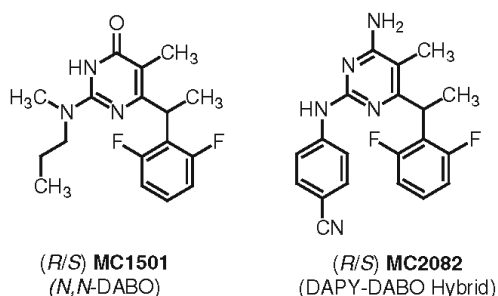


Fig. 5 Structures of racemic HIV-RT inhibitors resolved by Rotili et al. [5] used to validate CM4

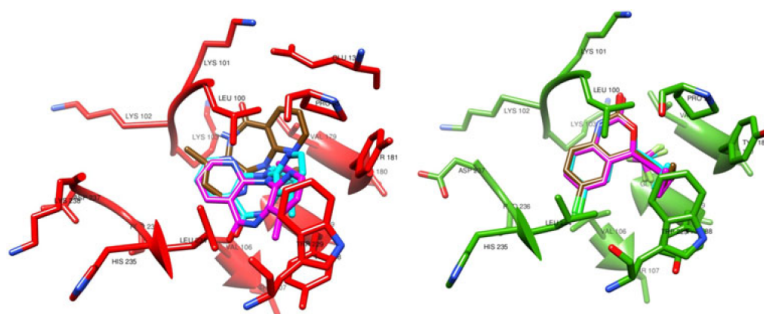
Vina program [19] which proved more reliable, as shown in Table 3 and Fig. 6, than Autodock [14] in reproducing the EFV- and NVP-experimental binding modes. In redocking Vina was more reliable than Autodock in reproducing the binding mode of both NVP and EFV starting from the experimental conformation of the ligands [4]. In view of these results and the fact that Vina was 10-times faster than Autodock, Vina was selected for docking experiments.

Figure 7 shows the binding modes of the DABO derivatives (structure of test set compounds are available in Appendix 1 and in the publication by Rotili et al. [5]) with the WT and the mutated HIV-RTs used in this study. Similarly to

Table 3 Docking assessment: root-mean-square deviations (RMSDs) displayed by the Vina and Autodock docking programs

PDB code	Mutation	Ligand	Vina		Autodock
			Exp	Mod	Exp
1fk9	WT	EFV	0.33	0.41	0.29
1fko	K103N	EFV	0.35	0.43	0.59
1fkp	K103N	NVP	0.53	0.81	3.41
1s1u	L100I	NVP	0.26	0.48	3.52
1vrt	WT	NVP	0.51	0.86	3.53

Fig. 6 Docking assessments: comparison of redocking by Vina and Autodock. In cyan are reported the experimental conformations in the 1vrt and 1fko complexes; in magenta are those redocked with Vina and in brown those obtained with Autodock. In red is shown HIV-RT in the 1vrt (nevirapine) complex and in green HIV-RT for 1fko (efavirenz)



previous studies [10–12], the *R*-conformations display an overall binding profile similar for either MC1501 or MC2082. In the *S*-configurations, the methyl at the C6-benzylic position (highlighted in red) prevented similar interactions [12]. The (*R*)-MC2082 binding mode is comparable with that of TMC278 (rilpivirine) [26], a recently reported DAPY derivative now undergoing clinical trials [27].

Figure 8 displays the (*R*)-MC2082 binding modes overlapped with the experimental complexes of etravirine and TMC278 in wild-type and mutated HIV-RTs.

Once the binding modes of MC compounds were calculated, the COMBINER model CM4 was readily applied. As reported in Table 4, the COMBINER model, although developed with only two different HIV-RT inhibitors, predicted the experimental MC activities with a surprisingly acceptable average absolute-error-of-prediction (0.89 pK_i). The CM4 model percentage of prediction error ranged between 61.6 and 0.9 % with an average error of 14.3 % comparable to those experimentally reported by Rotili et al. [5] (37.5, 1.5 and 16.2 %, respectively).

Most notably, the COMBINER CM4-model correctly predicted the right eudismic ratio for the two *R/S* pairs of MC derivatives.

Application of the COMBINER CM4-model to the test set of Rotili et al. [5] (MC compounds) helped the interpretation of the calculated activity contributions (Fig. 9) directly highlighting difference between MC1501 and MC2082 upon binding to HIV-RTs. Figure 9 shows that the activity contributions associated to the interactions of the most active MC enantiomers (stereoisomers *R*) with residues Lys101 were mainly responsible for the higher activities of (*R*)-MC2082 versus the (*R*)-MC1501 with an average increase of activities of about 0.29 and 0.19 of pK_i units for the hydrophobic and STE fields, respectively.

Comparing the activity contributions of *R*- and *S*-enantiomers of MC1501, the hydrophobic effect of residue Lys101 became negligible, while that from Trp229 increased with an average contribution of 0.24 pK_i units. In comparison, the Lys101-related STE contribution was

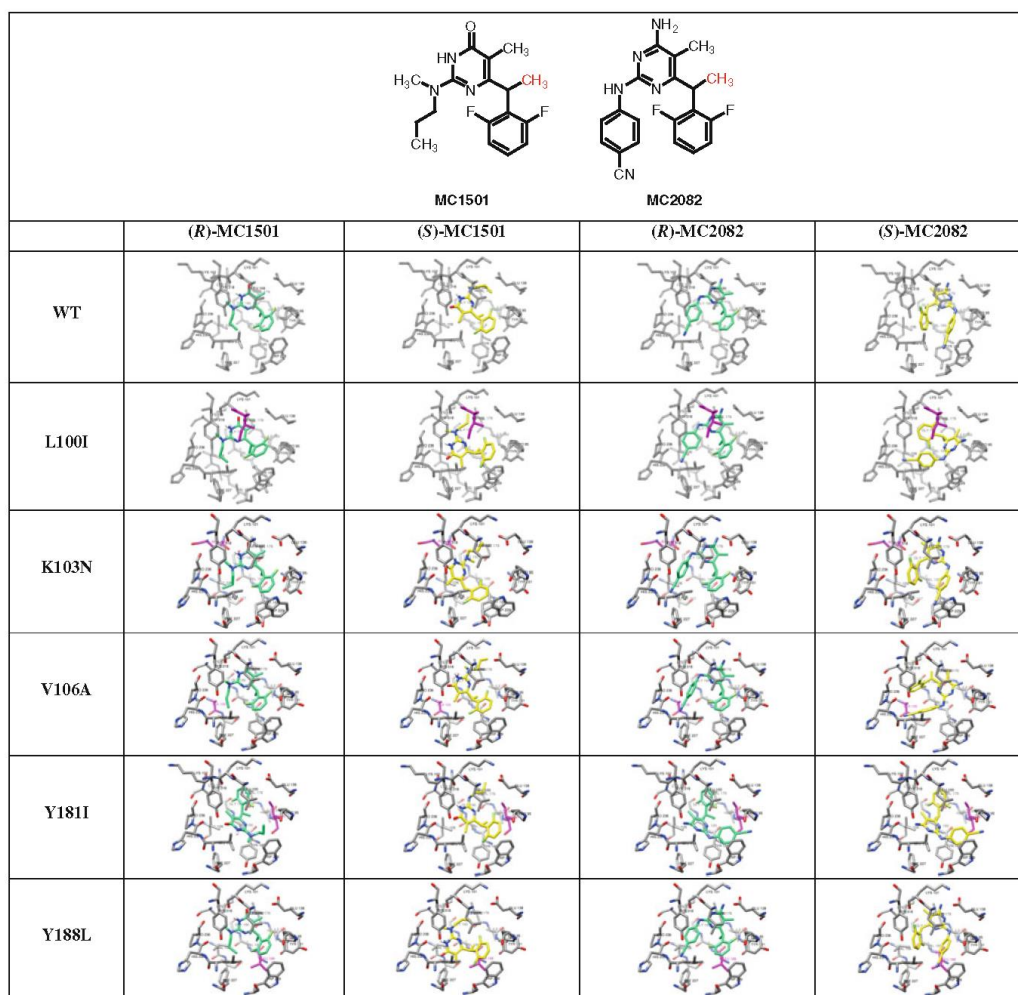


Fig. 7 Vina-proposed binding modes for the MC1501 and MC2082 enantiomers in six different HIV-RT proteins. The molecular structures are shown with the C6-methyl group highlighted in *red* at the top of the figure

more than doubled (see Tables 5, 6). In the case of MC2082 *R*- and *S*-enantiomers, the activity contribution of Lys101 was only reduced by 32 % (0.17), Trp229 increased to 0.16, and the Lys101 STE contribution increased more than 5 times (1.05).

Single-point mutations within the COMBINER CM4 model showed that residue 188 demonstrated a key role in modulating the interactions of the ligands both in its wild-type (Tyr188) and in the Leu188 mutation. Interestingly, for another mutating residue, residue 188 seemed to offset loss of interactions as a result of the other residue mutation, most

remarkably in the case of the more active compounds (*R*)-MC1501 and (*R*)-MC2082. Comparing the activity-contribution profile of (*R*)-MC2082 docked into wild-type HIV-RT and in the V106A-mutated protein, the only values changed drastically were those associated with Tyr188. A possible explanation for this might be that incoming missing interactions for the (*R*)-MC2082/Val106 → (*R*)-MC2082/Ala106 replacement are filled readily by the augmented (*R*)-MC2082/Tyr188 interactions (compare Tyr188 positions in Fig. 3), but this remains speculative without further analyses.

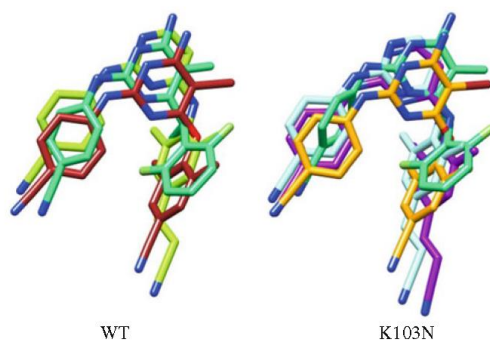


Fig. 8 Binding modes of (*R*)-MC2082 overlapped with etravirine and TMC278. On the *left* side are shown (*R*)-MC2082 in green, etravirine (3mec) in brown and TMC278 (2zd1) in light green, all bound to wild-type HIV-RT. On the *right* side are shown (*R*)-MC2082 (green) binding mode in K103N-mutated RT overlapped with etravirine (orange) that was co-crystallized with K103N HIV-RT, TMC278 (light blue) in the K103N–Y181C double mutant (3bgr) and in the L100I-K103 double mutant (purple, 2ze2)

Finally, Tables 5 and 6 clearly demonstrated that most mutations force the ligands to primarily re-adapt their interaction network around the two non-mutating Lys101 and Trp229 residues, supplying alternate ways of hydrogen-bond and hydrophobic anchor points with which ligands interact upon complex formation.

Fig. 9 Activity-contribution histograms calculated for the test MC compounds. Only bars with values higher than 0.001 and lower than -0.001 are shown

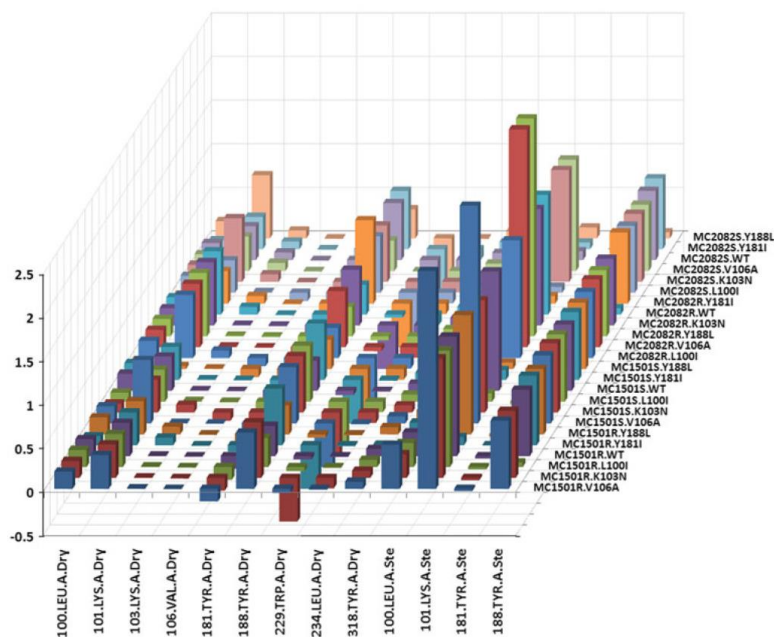


Table 4 Experimental and COMBINER CM4-model predicted activities of MC compounds of Rotili et al. [5]

	MC1501		MC2082		MC1501		MC2082	
					<i>R</i>		<i>S</i>	
	Exp	Pred	Exp	Pred	Exp	Pred	Exp	Pred
WT	8.70	7.46	6.93	7.20	6.81	7.21	4.52	5.77
V106A	8.52	9.19	6.45	5.78	9.52	9.43	6.62	7.51
K103N	7.02	7.17	6.01	7.52	8.52	9.11	7.19	7.52
L100I	7.02	6.69	4.40	7.11	8.10	7.49	6.74	6.03
Y188L	6.71	7.51	4.40	5.11	8.10	7.09	4.40	5.95
Y181I	6.35	6.05	4.40	6.12	6.12	6.25	6.29	5.48

Conclusions

The COMBINER approach integrates multiple sources of SAR information to build a self-consistent model of the amino-acid residues in both wild-type and mutant enzymes responsible for molecular recognition and discrimination. As with all such underdetermined 3-D QSAR models, predictability is the only real means of selection one model over another. This study on HIV-RT used a minimal set of inhibitor complexes to extract possible models for HIV-RT variants that rationalize the experimentally observed inhibitory activity of a novel set of compounds described

Table 5 COMBINEr CM4-model predicted activity contributions of MC1501 with average values higher than 0.01 absolute pK_i values

Field	Dry									Ste			
	100	K101	103	106	181	188	T229	L234	Y318	100	K101	181	188
(R)-MC1501.WT	0.19	0.38	0.01	-0.01	-0.15	0.35	-0.04	-0.01	0.08	0.27	1.37	-0.03	0.76
(R)-MC1501.L100I	0.20	0.38	0.01	-0.01	-0.15	0.33	-0.05	-0.01	0.08	0.28	1.34	-0.03	0.05
(R)-MC1501.K103N	0.19	0.38	0.01	-0.01	-0.15	0.64	-0.51	-0.12	0.08	0.27	1.37	-0.03	0.77
(R)-MC1501.V106A	0.20	0.39	0.01	0.00	0.14	0.65	0.05	0.01	0.08	0.51	2.55	0.03	0.78
(R)-MC1501.Y181I	0.10	0.38	0.09	-0.01	-0.07	0.65	-0.51	-0.01	0.01	0.26	0.10	0.00	0.80
(R)-MC1501.Y188L	0.19	0.38	0.09	-0.01	-0.15	0.33	-0.04	-0.01	0.08	0.27	1.37	-0.03	0.75
Average	0.18	0.38	0.03	-0.01	-0.13	0.49	-0.20	-0.03	0.07	0.31	1.35	-0.02	0.65
SD	0.04	0.00	0.04	0.00	0.03	0.17	0.24	0.04	0.03	0.10	0.78	0.01	0.29
Max	0.20	0.39	0.09	0.00	-0.07	0.65	-0.04	-0.01	0.08	0.51	2.55	0.00	0.80
Min	0.10	0.38	0.01	-0.01	-0.15	0.33	-0.51	-0.12	0.01	0.26	0.10	-0.03	0.05
Range	0.09	0.01	0.08	0.00	0.07	0.32	0.47	0.11	0.08	0.25	2.45	0.02	0.74
(S)-MC1501.WT	0.20	0.39	0.01	0.00	0.08	0.35	0.05	0.01	0.08	0.28	1.37	0.03	0.76
(S)-MC1501.L100I	0.10	0.37	0.01	-0.01	-0.15	0.64	-0.52	-0.12	0.08	0.26	0.10	-0.03	0.77
(S)-MC1501.K103N	0.10	0.38	0.09	-0.09	-0.08	0.65	-0.51	-0.12	0.08	0.26	1.29	0.00	0.80
(S)-MC1501.V106A	0.20	0.73	0.01	0.00	-0.08	0.65	-0.50	-0.01	0.08	0.51	2.58	0.00	0.78
(S)-MC1501.Y181I	0.19	0.38	0.01	0.01	0.08	0.65	0.52	0.01	0.08	0.27	0.11	0.00	0.78
(S)-MC1501.Y188L	0.10	0.03	0.01	-0.09	-0.08	0.34	-0.52	-0.12	0.08	0.26	0.07	0.00	0.76
Average	0.15	0.38	0.02	-0.03	-0.09	0.54	-0.44	-0.07	0.08	0.31	0.92	-0.01	0.78
SD	0.05	0.22	0.03	0.05	0.03	0.16	0.19	0.06	0.00	0.10	1.01	0.01	0.01
Max	0.20	0.73	0.09	0.00	-0.08	0.65	-0.05	-0.01	0.08	0.51	2.58	0.00	0.80
Min	0.10	0.03	0.01	-0.09	-0.15	0.34	-0.52	-0.12	0.08	0.26	0.07	-0.03	0.76
Range	0.09	0.69	0.08	0.09	0.07	0.31	0.47	0.11	0.00	0.25	2.50	0.02	0.04
R_{vsS}^a	0.03	0.00	0.01	0.03	-0.05	-0.05	0.24	0.04	-0.01	0.00	0.43	-0.01	-0.12

^a Differences between (R)-MC1501 and (S)-MC1501 activity-contribution averages. In bold are highlighted values cited in the interpretations of prediction reported in the text

Table 6 COMBINEr CM4-model predicted activity contributions of MC2082 with average values higher than 0.01 absolute pK_i values

Field	Dry									Ste			
	100	K101	103	106	181	188	T229	L234	Y318	100	K101	181	188
(R)-MC2082.WT	0.20	0.73	0.09	-0.01	-0.15	0.34	-0.03	-0.12	0.09	0.28	1.37	-0.03	0.07
(R)-MC2082.L100I	0.20	0.73	0.09	-0.09	-0.15	0.35	-0.51	-0.12	0.08	0.28	1.35	-0.03	0.76
(R)-MC2082.K103N	0.20	0.73	0.01	-0.01	-0.15	0.64	-0.51	-0.12	0.08	0.28	1.34	-0.03	0.76
(R)-MC2082.V106A	0.20	0.73	0.01	-0.01	-0.15	0.64	-0.05	-0.12	0.08	0.51	2.54	-0.03	0.78
(R)-MC2082.Y181I	0.10	0.37	0.09	-0.01	-0.08	0.96	-0.54	-0.12	0.08	0.26	0.09	-0.03	0.82
(R)-MC2082.Y188L	0.20	0.73	0.01	-0.01	-0.15	0.34	-0.04	-0.12	0.08	0.51	2.55	-0.03	0.76
Average	0.18	0.67	0.05	0.02	0.14	0.54	0.28	0.12	0.08	0.35	1.54	0.03	0.66
SD	0.04	0.14	0.04	0.03	0.03	0.25	0.26	0.00	0.00	0.12	0.92	0.00	0.29
Max	0.20	0.73	0.09	-0.01	-0.08	0.96	-0.03	-0.12	0.09	0.51	2.55	-0.03	0.82
Min	0.10	0.37	0.01	-0.09	-0.15	0.34	-0.54	-0.12	0.08	0.26	0.09	-0.03	0.07
Range	0.10	0.35	0.08	0.09	0.07	0.62	0.51	0.00	0.00	0.25	2.46	0.00	0.74
(S)-MC2082.WT	0.20	0.39	0.09	-0.01	-0.15	0.65	-0.52	-0.12	0.08	0.27	1.10	-0.03	0.79
(S)-MC2082.L100I	0.19	0.37	0.01	-0.09	-0.15	0.65	-0.51	-0.12	0.08	0.26	0.07	-0.03	0.77
(S)-MC2082.K103N	0.20	0.73	0.09	-0.01	-0.15	0.65	-0.52	-0.12	0.08	0.27	1.28	-0.03	0.78
(S)-MC2082.V106A	0.19	0.40	0.09	0.01	0.08	0.35	0.01	0.12	0.09	0.27	1.28	0.00	0.76
(S)-MC2082.Y181I	0.10	0.37	0.09	-0.01	-0.08	0.67	-0.55	-0.12	0.08	0.26	0.08	0.00	0.81

Table 6 continued

Field	Dry									Ste			
	100	K101	103	106	181	188	T229	L234	Y318	100	K101	181	188
(<i>S</i>)-MC2082.Y188L	0.20	0.72	0.09	-0.01	-0.15	0.33	-0.50	-0.01	0.08	0.27	0.12	-0.03	0.07
Average	0.18	0.50	0.08	-0.02	-0.12	0.55	-0.44	-0.10	0.08	0.27	0.49	-0.02	0.66
SD	0.04	0.18	0.03	0.04	0.04	0.16	0.21	0.04	0.00	0.01	0.61	0.01	0.29
Max	0.20	0.73	0.09	-0.01	-0.08	0.67	-0.01	-0.01	0.09	0.27	1.28	0.00	0.81
Min	0.10	0.37	0.01	-0.09	-0.15	0.33	-0.55	-0.12	0.08	0.26	0.07	-0.03	0.07
Range	0.09	0.36	0.09	0.09	0.07	0.34	0.54	0.11	0.00	0.01	1.22	0.02	0.74
RvsS ^a	0.00	0.17	-0.03	0.00	-0.01	-0.01	0.16	-0.02	0.00	0.08	1.05	-0.01	-0.01

^a Differences between (*R*)-MC2082 and (*S*)-MC2082 activity-contribution averages. In bold are highlighted the values cited in the interpretations of predictions reported in the text

by Rotili et al. including the relative activity of two different sets of stereoisomers. Obviously, prediction of novel inhibitors and their activities against HIV-RT is a logical next step to validate the utility of the COMBINER approach. Extension to similar problems, such as prediction of isoform selectivity of the eleven zinc-based histone deacetylase inhibitors, is underway.

Acknowledgments The authors thank the research group (Rotili et al. [5]) of Prof. Antonello Mai for sharing their data about the separation and activity of their DABO derivatives prior to publication. In addition, Garland R. Marshall acknowledges financial support from the Dipartimento di Chimica e Tecnologie del Farmaco, Facoltà di Farmacia e Medicina, Sapienza Università di Roma, which made his visiting professorship in Rome feasible.

References

- Lozano JJ, Pastor M, Cruciani G, Gaedt K, Centeno NB, Gago F, Sanz F (2000) 3-D-QSAR methods on the basis of ligand-receptor complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 ligands. *J Comput Aided Mol Des* 14:341–353
- Perez C, Pastor M, Ortiz AR, Gago F (1998) Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design. *J Med Chem* 41:836–852
- Rodriguez-Barrios F, Gago F (2004) Chemometrical identification of mutations in HIV-1 reverse transcriptase conferring resistance or enhanced sensitivity to arylsulfonylbenzotrioles. *J Am Chem Soc* 126:2718–2719
- Musmuca I, Caroli A, Mai A, Kaushik-Basu N, Arora P, Ragno R (2010) Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors. *J Chem Inform Model* 50:662–676
- Rotili D, Samuele A, Tarantino D, Ragno R, Musmuca I, Ballante F, Botta G, Morera L, Pierini M, Cirilli R, Nawrozkij MB, Gonzalez E, Clotet B, Artico M, Este JA, Maga G, Mai A (2012) 2-(Alkyl/aryl)amino-6-benzylpyrimidin-4(3H)-ones as inhibitors of wild-type and mutant HIV-1: enantioselectivity studies. *J Med Chem* 55:3558–3562
- Cancio R, Mai A, Rotili D, Artico M, Sbardella G, Clotet-Codina I, Este JA, Crespan E, Zanolli S, Hubscher U, Spadari S, Maga G (2007) Slow-, tight-binding HIV-1 reverse transcriptase non-nucleoside inhibitors highly active against drug-resistant mutants. *ChemMedChem* 2:445–448
- Samuele A, Facchini M, Rotili D, Mai A, Artico M, Armand-Ugon M, Este JA, Maga G (2008) Substrate-induced stable enzyme-inhibitor complex formation allows tight binding of novel 2-aminopyrimidin-4(3H)-ones to drug-resistant HIV-1 reverse transcriptase mutants. *ChemMedChem* 3:1412–1418
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612
- Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE (2006) Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics* 7:339
- Mai A, Sbardella G, Artico M, Ragno R, Massa S, Novellino E, Greco G, Lavecchia A, Musiu C, La Colla M, Murgioni C, La Colla P, Loddo R (2001) Structure-based design, synthesis, and biological evaluation of conformationally restricted novel 2-alkylthio-6-[1-(2,6-difluorophenyl)alkyl]-3,4-dihydro-5-alkylpyrimidin-4(3H)-ones as non-nucleoside inhibitors of HIV-1 reverse transcriptase. *J Med Chem* 44:2544–2554
- Quaglia M, Mai A, Sbardella G, Artico M, Ragno R, Massa S, del Piano D, Setzu G, Doratiotto S, Cotichini V (2001) Chiral resolution and molecular modeling investigation of rac-2-cyclopentylthio-6-[1-(2,6-difluorophenyl)ethyl]-3,4-dihydro-5-methylpyrimidin-4(3H)-one (MC-1047), a potent anti-HIV-1 reverse transcriptase agent of the DABO class. *Chirality* 13:75–80
- Ragno R, Mai A, Sbardella G, Artico M, Massa S, Musiu C, Mura M, Marturana F, Cadeddu A, La Colla P (2004) Computer-aided design, synthesis, and anti-HIV-1 activity in vitro of 2-alkylamino-6-[1-(2,6-difluorophenyl)alkyl]-3,4-dihydro-5-alkylpyrimidin-4(3H)-ones as novel potent non-nucleoside reverse transcriptase inhibitors, also active against the Y181C variant. *J Med Chem* 47:928–934
- Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26:1668–1688
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock and AutoDockTools: automated docking with selective receptor flexibility. *J Comput Chem* 30:2785–2791

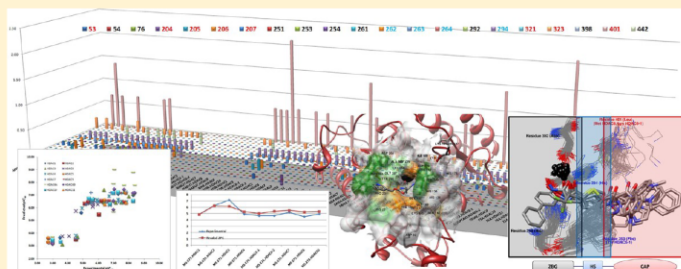
15. Mevik B-H, Wehrens R (2007) The pls package: principal component and partial least squares regression in R. *J Stat Softw* 18(2):1–24
16. Ren J, Milton J, Weaver KL, Short SA, Stuart DI, Stammers DK (2000) Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase. *Structure* 8:1089–1094
17. Ren J, Esnouf R, Garman E, Somers D, Ross C, Kirby I, Keeling J, Darby G, Jones Y, Stuart D et al (1995) High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat Struct Biol* 2:293–302
18. Ren J, Nichols CE, Chamberlain PP, Weaver KL, Short SA, Stammers DK (2004) Crystal structures of HIV-1 reverse transcriptases mutated at codons 100, 106 and 108 and mechanisms of resistance to non-nucleoside inhibitors. *J Mol Biol* 336:569–578
19. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461
20. R-Development-Core-Team R: a language and environment for statistical computing. <http://www.R-project.org>
21. Ballante F, Ragno R (2012) 3-D QSAutogrid/R: an alternative procedure to build 3-D QSAR models. *Methodologies and applications. J Chem Inf Model* 52:1674–1685
22. Baroni M, Costantino G, Cruciani G, Riganelli D, Valigi R, Clementi S (1993) Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3-D-QSAR problems. *Quant Struct Activ Relatsh* 12:9–20
23. Wesson L, Eisenberg D (1992) Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci* 1:227–235
24. Goodford PJ (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* 28:849–857
25. Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 110:5959–5967
26. Azijn H, Tirry I, Vingerhoets J, de Bethune MP, Kraus G, Boven K, Jochmans D, Van Craenenbroeck E, Picchio G, Rimsky LT (2010) TMC278, a next-generation nonnucleoside reverse transcriptase inhibitor (NNRTI), active against wild-type and NNRTI-resistant HIV-1. *Antimicrob Agents Chemother* 54:718–727
27. Macarthur RD (2011) Clinical trial report: TMC278 (rilpivirine) versus efavirenz as initial therapy in treatment-naive, HIV-1-infected patients. *Curr Infect Dis Rep* 13:1–3

Histone Deacetylase Inhibitors: Structure-Based Modeling and Isoform-Selectivity Prediction

Laura Silvestri,[†] Flavio Ballante,[†] Antonello Mai,[‡] Garland R. Marshall,^{†,§} and Rino Ragno^{*†}

[†]Rome Center for Molecular Design Dipartimento di Chimica e Tecnologie del Farmaco, Facoltà di Farmacia e Medicina, [‡]Istituto Pasteur—Fondazione Cenci Bolognetti Dipartimento di Chimica e Tecnologie del Farmaco, Facoltà di Farmacia e Medicina, Sapienza Università di Roma, P.le A. Moro 5, 00185 Rome, Italy

S Supporting Information



ABSTRACT: An enhanced version of comparative binding energy (COMBINE) analysis, named COMBINER, based on both ligand-based and structure-based alignments has been used to build several 3-D QSAR models for the eleven human zinc-based histone deacetylases (HDACs). When faced with an abundance of data from diverse structure–activity sources, choosing the best paradigm for an integrative analysis is difficult. A common example from studies on enzyme–inhibitors is the abundance of crystal structures characterized by diverse ligands complexed with different enzyme isoforms. A novel comprehensive tool for data mining on such inhomogeneous set of structure–activity data was developed based on the original approach of Ortiz, Gago, and Wade, and applied to predict HDAC inhibitors’ isoform selectivity. The COMBINER approach (apart from the AMBER programs) has been developed to use only software freely available to academics.

1. INTRODUCTION

Histone deacetylases (HDACs) are a family of protein modifying-enzymes found in bacteria, fungi, plants and animals. In the human, 18 different isoforms have been identified and divided into 4 classes according to size, cellular localization, number of active sites and homology with yeast deacetylases.¹ This paper focuses on the eleven zinc-based HDACs isoforms of class I and II found in humans. Class I, that includes HDAC-1, -2, -3, and -8, is related to yeast RPD3, shares nuclear localization with the exception of HDAC3, and has ubiquitous expression. Instead, class II shows domains with similarity to yeast Hda1 and can be further divided into class IIa, which includes HDAC-4, -5, -7, and -9, and class IIb (HDAC-6 and -10) that contains two catalytic sites. HDAC-3 and members of class II have been shown to shuttle between the cytoplasm and nucleus and have tissue-specific expression. HDAC-11 is the only member of class IV. HDAC classes I, II, and IV are zinc-dependent proteases; unlike those of class III, called sirtuins, which require NAD⁺ as cofactor.

HDACs play a key role in epigenetics—controlling gene expression involved in all aspects of biology—cell proliferation, chromosome remodeling, and gene transcription.² They regulate the acetylated state of histone proteins removing the acetyl

moiety from the ϵ -amino group of lysine residues on the N-terminal extension of the core histones, this leads to changes in the structure of histones and therefore modifies the accessibility of transcription enzymes with gene-promoter regions. In addition, HDACs dynamically modify the activity of diverse types of nonhistone proteins.³ These include transcription factors, signal-transduction mediators, microtubules, and a molecular chaperone. In particular, distinct HDACs class I and II are overexpressed in several types of cancer. For these reasons, HDAC inhibitors (HDACIs) have been developed and approved for the treatment of cutaneous T-cell lymphoma: Merck’s Zolinza (SAHA) and Celgene’s Istodax (Romidepsin, FK228).⁴ More recently, HDACIs have emerged as potential therapeutics for the stimulation of viral expression from infected cells in the hope of eradication of HIV infection.⁵

HDAC inhibitors are classified according to their chemical structure as follows: short-chain fatty acids, hydroxamic acids, benzamides, ketones, and cyclic peptides with a pendant functional group. Each HDACI shows variability in its ability

Received: March 26, 2012

Published: July 4, 2012

to inhibit particular isoforms. Unfortunately, as for SAHA and TSA, the majority of HDACs inhibit most HDAC isoforms nonspecifically. Others, such as MS-275, a benzamide, are more selective for class I. Selective HDAC inhibitors, which affect either a single HDAC isoform or only a few isoforms within a single class, would be ideal molecular scalpels to help elucidate the individual functions of each HDAC isoform in the complexity of epigenetics.

This paper focuses on a predictive tool for the anti-HDAC activity and selectivity prediction of potential new inhibitors to enhance isoform specificity. To this aim, structure-based three-dimensional structure–activity relationship (3-D QSAR) models were derived by a comparative binding energy (COMBINER) analysis on a series of inhibitors for which biological activities against the 11 human zinc-based HDACs isoforms were available.

2. OVERVIEW

Modified COMBINE Approach (COMBINER). The comparative binding energy (COMBINE) approach is a structure-based 3-D QSAR method that uses a series of receptor–ligand complexes to quantify interaction energies by molecular mechanics.⁶ The fundamental idea of a COMBINE analysis is that a simple expression for the differences in binding affinity of a series of related ligand–receptor complexes can be derived by using multivariate statistics to correlate experimental data on binding affinities with per residue ligand–receptor interactions, computed from 3-D structures. The basis of the COMBINE method is the assumption that the protein–receptor binding free energy, ΔG , can be approximated by a weighted sum of n terms, ΔU , each describing the change in property u upon binding as described by the following equation:

$$\Delta G = \sum_{i=1}^n w_i \Delta u_i + C$$

From this expression, biological activities may be derived by assuming that these quantities are linear functions of ΔG . The expression is derived by analyzing the interaction of a set of ligands with experimentally known binding affinities for a target receptor.^{6b}

In order to apply this approach to predict the selective inhibition of HDAC isozymes, a modified protocol, called COMBINER,⁷ (Figure 1) used the AutoDock's AutoGrid engine to compute the components of the ligand–residues interaction energies for each ligand/enzyme complex. The PLS (*partial least squares for latent variables*) paradigm, as implemented in the R⁸ environment, was used to derive robust, predictive COMBINER models. Although the original COMBINE (gCOMBINE)⁹ was available, it was decided to develop COMBINER because it allows direct calculation of ligand/enzyme per residue interaction from docking results without further complex parametrization as required in the original COMBINE.

Training Set. Nine experimental 3-D structures of HDAC-2, -4, -7, and -8 cocrystallized with different ligands were retrieved from the Protein Data Bank¹⁰ (Table 1). The remaining HDAC isoforms whose experimental structures were not experimentally available (HDAC-1, -3, -5, -6-1, -6-2, -9, -10, and -11) were built by homology modeling. In the case of HDAC-6, both the histone- and tubuline-catalytic domains were built (histones: HDAC-6-1 and tubulin: HDAC-6-2) with the same experimental inhibitory activities assigned to each complex.

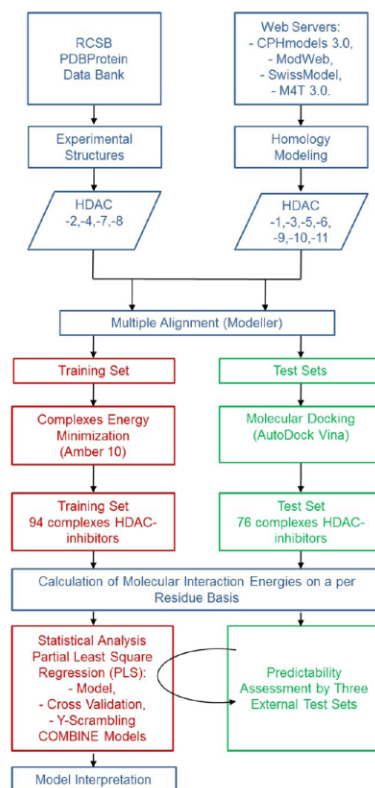


Figure 1. Flowchart of the COMBINER model generation. The red contoured boxes identify the model preparation, while the green contoured boxes refer to the test set. The blue contoured boxes refer to either training set or test set.

In addition to cocrystallized inhibitors, other compounds (Table 2) reported simultaneously from the same laboratory by Blackwell et al.¹³ were selected. The data set composed by 15 different inhibitors and 12 HDAC isoforms was reduced from the theoretical number of 180 to 94 due to lacking of complete isozyme-inhibitory data. Therefore, the final training set summarized in Table 3 comprised 39 complexes derived with crystallized structures, built according to structural similarity of modeled inhibitors with cocrystallized compounds, and 55 complexes derived with homology models. The latter are generated according to the web-servers used for producing the homology models (see the Experimental Section).

The training sets complexes were energy minimized with Amber 10²¹ and multiply aligned using Modeller²² to establish structure-based residue equivalence. This alignment provided the structural basis for computing the molecular-interaction fields with a corresponding per-residue basis for all enzyme isoforms. Because different isoforms of HDACs show structural diversity in terms of amino-acid sequences and differed in numbers of amino acids (multitarget study), all HDAC residues were renumbered in an arbitrarily way: the same numbering were assigned to those residues showing spatial superimposition;

Table 1. PDB Codes, Ligand Names, Chemical Structures, and HDAC Inhibitory Activities of Complexes Downloaded from Protein Data Bank^a

PDB code	HDAC Class	HDAC Number	Ligand structure	IUPAC name	IC ₅₀ (μM)
3MAX ¹¹	I	2		<i>N</i> -(4-aminobiphenyl-3-yl)benzamide (LLX)	0.9 ¹¹
3F07 ¹²	I	8		(2E)- <i>N</i> -hydroxy-3-[1-methyl-4-(phenylacetyl)-1H-pyrrol-2-yl]prop-2-enamide (APHA8)	2.8 ¹³
1T64 ¹⁴	I	8		7-(4-(Dimethylamino)phenyl)- <i>N</i> -hydroxy-4,6-dimethyl-7-oxo-2,4-heptadienamide (TSA)	1.1 ¹³
1T67 ¹⁴	I	8		4-dimethylamino- <i>n</i> -(6-hydroxycarbamoylethyl)benzamide- <i>n</i> -hydroxy-7-(4-dimethylaminobenzoyl)aminoheptanamide (MS-344)	0.249 ¹⁵
1T69 ¹⁴	I	8		octanedioic acid hydroxyamidephenylamide (SAHA)	2.2 ¹³
1W22 ¹⁴	I	8		<i>N</i> -hydroxy-4-(methyl[(5-pyridin-2-yl)thiophen-2-yl]sulfonyl)amino)benzamide (NHB)	0.175 ¹⁶
2VQM ¹⁷	II a	4		<i>N</i> -hydroxy-5-[(3-phenyl-5,6-dihydroimidazo[1,2-a]pyrazin-7(8H)-yl)carbonyl]thiophene-2-carboxamide (HA3)	0.978 ¹⁷
2VQJ ¹⁷	II a	4		2,2,2-trifluoro-1-[5-[(3-phenyl-5,6-dihydroimidazo[1,2-a]pyrazin-7(8H)-yl)carbonyl]thiophen-2-yl]ethane-1,1-diol (TFMK)	0.367 ¹⁷
3C0Z ¹⁸	II a	7		octanedioic acid hydroxyamidephenylamide (SAHA)	0.05 ¹³
3C10 ¹⁸	II a	7		7-(4-(Dimethylamino)phenyl)- <i>N</i> -hydroxy-4,6-dimethyl-7-oxo-2,4-heptadienamide (TSA)	0.014 ¹³

^aIC₅₀s were all evaluated in a similar way using a fluorescently labeled acetylated peptide as substrate.



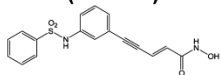
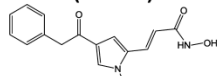
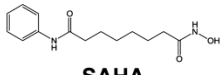
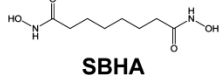
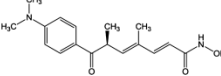
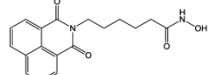
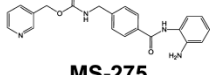
conversely, a “ghost” residue was attributed in the regions which presented structural diversity (see the Supporting Information for the multiple sequence alignment of all proteins, Supporting Info File 6). In this way, a total of 571 amino-acid residues, 12-fragmented HDAC isoform structures, were obtained. The calculation of the ligand/residues was conducted similarly as previously reported.⁷ The calculated molecular descriptors were imported in R²³ to generate structure-based 3-D QSAR models. The purpose of training-set complex minimization was to generate not only 94 optimized complexes, but also to have several conformation for each HDAC useful in the subsequent preparation of test-set complexes by ligand cross-docking (see below).

Each derived COMBINER model was subjected to internal (cross-validation) and external (test-set) assessments. Cross-validation was done using both the leave-one-out (LOO) and random five groups leave-some-out (R5G-LSO) techniques. For external validation, a series of molecules with known inhibitory activity against HDAC isozymes was selected as an external test set for the model’s predictability assessment.

External Test Sets for the COMBINER Model Validation.

Three different test sets were used for external validation. The first one (modeled test set, MTS) contained a series of molecules, docked with AutoDockVina,²⁴ that showed inhibitory activity against several HDAC isoforms (Table 4).

Table 2. Training Set—Chemical Structures and HDACs Inhibitory Activities^a

HDAC	CLASS	I					IIa			IIb		IV
		Number	1	2	3	8	4	5	7	9	6	10
Chemical Structures and IDs	 valproic acid (VALP)	1000	1000	226.08	228.85	-	2000 ¹⁹	-	2000 ¹⁹	1000	1000	-
	 Butyrate (NABUT)	319	28.9	22.5	85.6	30	2000 ¹⁹	30	2000 ¹⁹	1000	292	-
	 Oxamflatin (OXAM)	0.05	0.2	0.01	2.2	0.03	-	0.03	-	0.09	0.05	-
	 APHA8	3.7	7.4	0.42	2.8	3.1	-	3.1	-	0.1	4.2	-
	 SAHA	0.1	0.44	0.02	2.2	0.05	0.378 ²⁰	0.05	0.316 ²⁰	0.02	0.1	0.362 ²⁰
	 SBHA	2.1	4.6	0.41	3.7	1.4	-	1.3	-	0.1	2.3	-
	 TSA	0.005	0.021	0.005	1.1	0.014	0.0165 ²⁰	0.014	0.0381 ²⁰	0.005	0.005	0.0152 ²⁰
	 SCRIPTAID (SCRIP)	0.17	0.64	0.03	2.3	0.2	-	0.16	-	0.004	0.17	-
	 MS-275	13	0.51	0.07	30	12	-	6.2	-	21	11.5	-

^aIC₅₀s (expressed in micromolar) were all evaluated in a similar way using a fluorescent-labeled acetylated peptide as substrate.

The second test set was comprised of a series of cocrystallized complexes structures (crystal test set, CTS) containing two HDAC-8 complexes (not available from the PDB during model development) and four bacterial HDAC homologues (Table 5). The third test set was also modeled, using largazole (a cyclotetrapeptide-containing HDAC inhibitor, largazole test set, LTS) whose crystal structure with HDAC-8 was reported,³¹ but whose inhibitory activity was available only for four HDAC isoforms (Table 6). For LTS, largazole was docked with HDAC-1, HDAC-2, HDAC-3, and HDAC-6-1. The bacterial HDACs

complexes with hydroxamic acids were available from the PDB (Table 5).

3. RESULTS AND DISCUSSION

COMBINER Models—Overall Analysis. All final models contained 94-inhibitor/enzyme complexes spanning an activity range, expressed as pIC₅₀, between 2.7 (NABUT against HDAC-5) and 8.4 (SCRIPTAID against HDAC-6). The statistical results of the final models are summarized in Table 7. Genetic algorithm (GA) variable-selection was applied, but provided little

Table 3. Training Set Composition: Inhibitors' Names, Corresponding HDAC Used in the Complex, Information on Source of Protein Structure, and pIC₅₀ Values

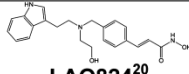
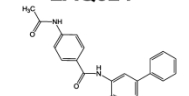
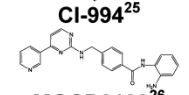
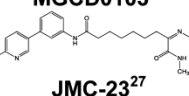
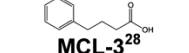
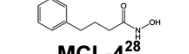
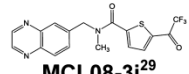
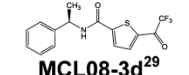
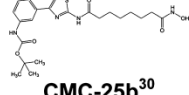
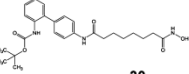
no.	compound name	HDAC isoform	protein source	pIC ₅₀	no.	compound name	HDAC isoform	protein source	pIC ₅₀
1	VALP	HDAC1	ModWeb	3.00	48	SAHA	HDAC6-1	SwissModel	7.70
2	NABUT	HDAC1	ModWeb	3.50	49	TSA	HDAC6-1	SwissModel	8.30
3	MS-275	HDAC1	M4T	4.89	50	SCRIP	HDAC6-1	SwissModel	8.40
4	APHA8	HDAC1	SwissModel	5.43	51	NABUT	HDAC6-2	CPH	3.00
5	SBHA	HDAC1	CPH	5.68	52	VALP	HDAC6-2	ModWeb	3.00
6	SCRIP	HDAC1	ModWeb	6.77	53	APHA8	HDAC6-2	CPH	7.00
7	SAHA	HDAC1	M4T	7.00	54	MS-275	HDAC6-2	M4T	7.00
8	OXAM	HDAC1	ModWeb	7.30	55	SBHA	HDAC6-2	ModWeb	7.00
9	TSA	HDAC1	CPH	8.30	56	OXAM	HDAC6-2	CPH	7.05
10	VALP	HDAC2	Crystal	3.00	57	SAHA	HDAC6-2	CPH	7.70
11	NABUT	HDAC2	Crystal	4.54	58	TSA	HDAC6-2	M4T	8.30
12	APHA8	HDAC2	Crystal	5.13	59	SCRIP/TAID	HDAC6-2	M4T	8.40
13	SBHA	HDAC2	Crystal	5.34	60	NABUT	HDAC7	Crystal	4.52
14	LLX	HDAC2	Crystal	6.05	61	MS-275	HDAC7	Crystal	5.21
15	SCRIP	HDAC2	Crystal	6.19	62	APHA8	HDAC7	Crystal	5.51
16	MS-275	HDAC2	Crystal	6.29	63	SBHA	HDAC7	Crystal	5.89
17	SAHA	HDAC2	Crystal	6.36	64	SCRIP	HDAC7	Crystal	6.80
18	OXAM	HDAC2	Crystal	6.70	65	SAHA	HDAC7	Crystal	7.30
19	TSA	HDAC2	Crystal	7.68	66	OXAM	HDAC7	Crystal	7.52
20	VALP	HDAC3	CPH	3.65	67	TSA	HDAC7	Crystal	7.85
21	NABUT	HDAC3	SwissModel	4.65	68	VALP	HDAC8	Crystal	3.64
22	APHA8	HDAC3	CPH	6.38	69	NABUT	HDAC8	Crystal	4.07
23	SBHA	HDAC3	SwissModel	6.39	70	MS-275	HDAC8	Crystal	4.52
24	MS-275	HDAC3	CPH	7.16	71	SBHA	HDAC8	Crystal	5.43
25	SCRIP	HDAC3	SwissModel	7.52	72	APHA8	HDAC8	Crystal	5.55
26	SAHA	HDAC3	CPH	7.70	73	SCRIP	HDAC8	Crystal	5.64
27	OXAM	HDAC3	SwissModel	8.00	74	OXAM	HDAC8	Crystal	5.66
28	TSA	HDAC3	SwissModel	8.30	75	SAHA	HDAC8	Crystal	5.66
29	NABUT	HDAC4	Crystal	4.52	76	TSA	HDAC8	Crystal	5.96
30	MS-275	HDAC4	Crystal	4.92	77	MS344	HDAC8	Crystal	6.60
31	APHA8	HDAC4	Crystal	5.51	78	NHB	HDAC8	Crystal	6.76
32	SBHA	HDAC4	Crystal	5.89	79	NABUT	HDAC9	ModWeb	2.70
33	HA3	HDAC4	Crystal	6.01	80	VALP	HDAC9	CPH	2.70
34	TFMK	HDAC4	Crystal	6.44	81	SAHA	HDAC9	ModWeb	6.50
35	SCRIP	HDAC4	Crystal	6.70	82	TSA	HDAC9	ModWeb	7.42
36	SAHA	HDAC4	Crystal	7.30	83	VALP	HDAC10	M4T	3.00
37	OXAM	HDAC4	Crystal	7.52	84	NABUT	HDAC10	CPH	3.54
38	TSA	HDAC4	Crystal	7.85	85	MS-275	HDAC10	ModWeb	4.94
39	NABUT	HDAC5	ModWeb	2.70	86	APHA8	HDAC10	SwissModel	5.38
40	VALP	HDAC5	ModWeb	2.70	87	SBHA	HDAC10	M4T	5.64
41	SAHA	HDAC5	CPH	6.42	88	SCRIP	HDAC10	ModWeb	6.77
42	TSA	HDAC5	CPH	7.80	89	SAHA	HDAC10	ModWeb	7.00
43	NABUT	HDAC6-1	M4T	3.00	90	OXAM	HDAC10	CPH	7.30
44	VALP	HDAC6-1	CPH	3.00	91	TSA	HDAC10	CPH	8.30
45	APHA8	HDAC6-1	SwissModel	7.00	92	SAHA	HDAC11	ModWeb	6.44
46	MS-275	HDAC6-1	ModWeb	7.00	93	TSA	HDAC11	ModWeb	7.82
47	SBHA	HDAC6-1	CPH	7.00	94	SAHA	HDAC6-1	SwissModel	7.70

improvement in either descriptive or predictive performance; hence, the non-GA-optimized models were used.

Structure-activity relationships of the various HDAC inhibitors have previously been described in other studies.³⁶ Crystal structures of receptor-ligand complexes have been analyzed qualitatively or by comparison of bound ligands.³⁷ COMBINER analysis permits quantification of structure-activity relationships through the electrostatic (Coulombic) and van der Waals interaction energies as well as additional parameters, such as solvation energy. Distinguished from the original COMBINE procedure of Ortiz,^{6b} COMBINER computes enzyme/ligand

interactions using the AutoGrid program based on the AMBER united-atom force field and chosen for its simpler molecular format (PDBQT). The data in Table 7 refer to the monoprobes fields (ELE, STE, DRY) and the multiprobes ones: electrostatic-steric (ELE+STE), electrostatic-desolvation (ELE+DRY), and electrostatic-steric-desolvation (ELE+STE+DRY). The reported statistical coefficients allowed estimates of goodness and robustness of each model. Results indicated the ELE+DRY model as the best. In fact, the overall generated model showed the highest conventional squared correlation coefficient (r^2) and lowest standard deviation error of calculation (SDEC) values:

Table 4. MTS Chemical Structures and Reported HDAC Inhibitory Activities^a

HDAC	Class	I				IIa				IIb		IV	
		Number	1	2	3	8	4	5	7	9	6	10	11
Chemical Structures and IDs	 LAQ824 ²⁰		0.00323	0.01570	0.01050	0.00384	0.00582	0.00558	0.00611	0.00824	0.00593	0.00841	0.00558
	 CI-994 ²⁵		0.41	-	0.75	100	-	-	-	-	100	-	-
	 MGCD0103 ²⁶		0.15	0.29	1.66	-	-	-	-	-	-	-	0.59
	 JMC-23 ²⁷		19.3	69.7	1.99	100	58.9	21.0	29.7	13.3	93.5	23.1	34.1
	 MCL-3 ²⁸		64	65	260	93	2000	2000	2000	2000	240	-	-
	 MCL-4 ²⁸		0.6	0.6	2	4	140	25	150	430	0.5	-	-
	 MCL08-3i ²⁹		0.58	-	0.67	-	0.098	-	-	-	0.089	-	-
	 MCL08-3d ²⁹		0.32	-	0.23	-	0.076	-	-	-	0.36	-	-
	 CMC-25b ³⁰		0.004	0.021	0.002	2.58	-	-	-	-	0.0002	0.002	-
	 CMC-7f ³⁰		0.057	0.074	0.018	1.72	-	-	-	-	0.011	0.083	-

^aIC₅₀ expressed in micromolar.

0.80 and 0.73, respectively (Figure 2A), comparable to those reported by Wade et al. in a similar application.³⁸ To assess the models' internal predictive power and robustness, two validation methods were used as follows: cross-validation (CV, internal validation) and Y-scrambling. LOO and RSG-LSO methods were chosen for cross-validation, obtaining for both q^2 values of 0.76 for the ELE+DRY probe, using only 2 principal components (Figure 2B). These results suggested good internal predictability (CV) of the model. Furthermore, SDEP (standard deviation error-of-prediction) provided an estimation of model internal predictivity by means of cross-validation; values less than 1 are generally considered indexes of good predictions. Upon further inspection, a high level of inverse correlation between the DRY and STE fields was found; more than 84 out of 94 complexes

(~90%) showed a correlation coefficient between -0.60 and -0.99 , rationalizing the similar statistical coefficients among models 4, 5, and 7 (Table 7). Therefore, the DRY field may be interpreted here as a probable estimation of steric interactions as well.

The charts in Figure 2 highlight the results of Table 7 and show linearity between experimental and recalculated/predicted data, expressed as pIC_{50} . Two views of experimental versus the RSG-LSO cross-validation predictions, indicating with different symbols each inhibitor and each HDAC isoform, are shown in Figure 3. This double representation emphasizes how the COMBINER model retains the correlation within various subgroups, either considering all the training-set inhibitors versus each HDAC (correlation of anti-HDAC inhibitors

Table 5. CTS: PDB Codes, Ligand Names, Chemical Structures, and HDAC Inhibitory Activities

PDB code	Receptor name	Ligand structure	Ligand name	IC ₅₀ (μM)
3SFF ³²	HDAC8		(2R)-2-amino-3-(3-chlorophenyl)-1-[4-(2,5-difluorobenzoyl)piperazin-1-yl]propan-1-one (ODI)	0.2
3SFH ³²	HDAC8		(2R)-2-amino-3-(2,4-dichlorophenyl)-1-(1,3-dihydro-2H-isoindol-2-yl)propan-1-one (1DI)	0.09
1C3R ³³	HDLP		TSA	0.4
2GH6 ³⁴	HDAH		9,9,9-trifluoro-8-oxo-n-phenylnonanamide(CF3)	11.19
1ZZ1 ³⁵	HDAH		SAHA	0.95
1ZZ3 ³⁵	HDAH		3-cyclopentyl-n-hydroxypropanamide (3YP)	0.29

Table 6. LTS: PDB Code, Ligand Name, Chemical Structure, and HDAC Inhibitory Activities

PDB code	Ligand structure	Ligand name	IC ₅₀ (μM)				
			HDAC 1	HDAC 2	HDAC 3	HDAC 6	HDAC 8
3RQD ³ 1		Largazolethiol	0.0012	0.0035	0.0034	0.049	-

Table 7. Statistical Results of the COMBINER Models

#	Field	PC ^a	r ^{2b}	SDEC ^c	q ² _{RSG-LSO} ^d	SDEP _{RSG-LSO} ^e	q ² _{LOO} ^f	SDEP _{LOO} ^g	scrambled q ^{2h} % positive values	Max. value
1	ELE	2	0.69	0.91	0.67	0.94	0.68	0.93	5	0.07
2	STE	2	0.27	1.40	0.14	1.52	0.15	1.51	n.d.	n.d.
3	DRY	2	0.46	1.21	0.34	1.33	0.36	1.32	n.d.	n.d.
4	ELE+STE	2	0.74	0.84	0.68	0.93	0.68	0.93	2	0.05
5	ELE+DRY	2	0.80	0.73	0.76	0.81	0.76	0.81	6	0.08
6	STE+DRY	3	0.54	1.11	0.33	1.34	0.35	1.33	n.d.	n.d.
7	ELE+DRY+STE	2	0.77	0.78	0.72	0.87	0.72	0.87	4	0.04

^aNumber of principal components used. ^bConventional squared correlation coefficient. ^cStandard deviation error of recalculation. ^dCross-validation coefficient for the RSG-LSO method. ^eStandard deviation error of prediction for the RSG-LSO cross-validation. ^fCross-validation coefficient for the LOO method. ^gStandard deviation error of prediction for the LOO cross-validation. ^hPercentage of positive cross-validation coefficients for the scrambling method and its maximum values.

potency, Figure 3) or considering the each inhibitor binding into different HDAC isoforms (correlation of selectivity, Figure 3). This latter consideration is consistent and supported the fact that the LOO and RSG-LSO cross-validation q^2 's showed the same values. Furthermore, to check for methodological self-consistency, reduced COMBINER models built for several inhibitors against each HDAC isoform (inhibition potencies) and for each

inhibitor against several HDAC isoforms (selectivity issue) revealed the existence of relationships with r^2 ranging from 0.7 to 0.8.

Finally, both robustness and absence-of-chance correlation of the COMBINER models listed in Table 7 were checked by random scrambling (Y-scrambling). Through this approach, a random reassignment of inhibitory activity to compounds of the

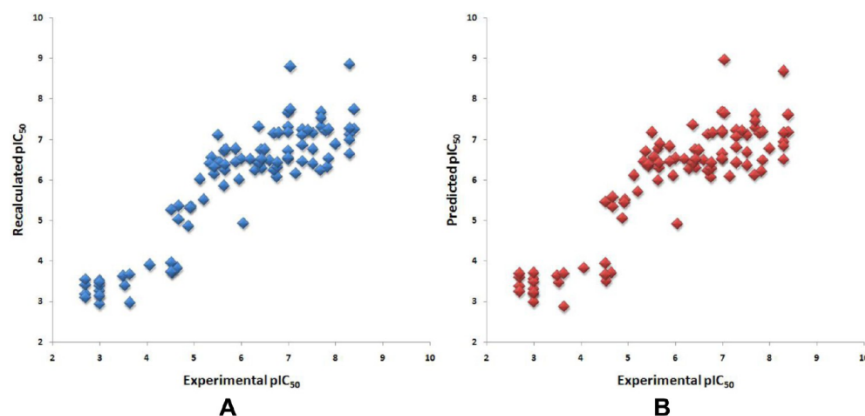


Figure 2. Fitting (A) and RSG-LSO cross-validation (B) plots, for the ELE+DRY model (Table 7).

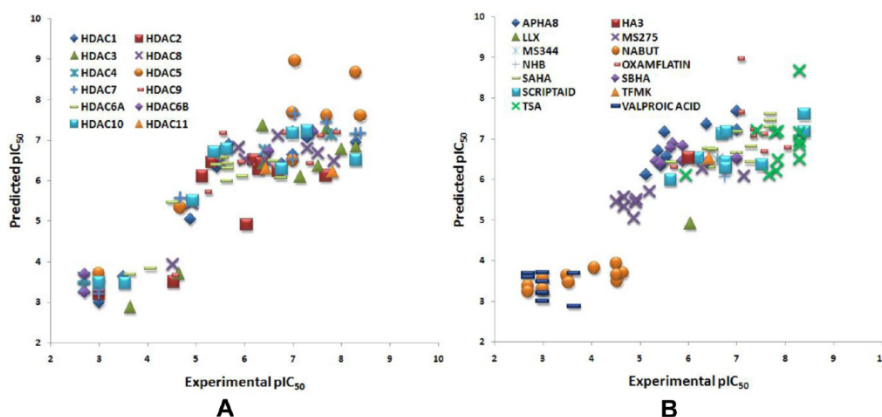


Figure 3. RSG-LSO cross-validation predictions depicted by HDAC isoforms (A) and by inhibitor (B).

data set was achieved to generate numerous data sets; for each scrambled data set, a RSG-LSO cross-validation was run. One-hundred Y-scrambling runs were examined; their analysis revealed that only 6% of all Y vectors had a correlation with the original Y values with maximum scrambled q^2 of only 0.08 in the case of ELE+DRY probe. Regarding the other models, in the case of ELE and ELE+STE+DRY, a chance correlation of 4% and 5% with a q^2 maximum value of 0.04 and 0.07 were observed, respectively. The ELE+STE probe showed a chance correlation of 2% with a q^2 maximum value of 0.05. These correlations appear random and excluded possible correlations between the original Y vector and the scrambled Y vectors. For the best model (ELE+DRY) in 100 random scrambled models, the number of positive q^2 values were only 6 leading to a probability of chance correlation lower than 1% with a q^2 value of 0.1, quite acceptable results considering the cross-validation coefficients of 0.76 of the model. Cross-validation runs using the most stringent leave-half-out method confirmed the robustness of the models (data not shown).

ELE and DRY Model Interpretation. Interpretation of COMBINER models can identify the residues relevant for differences in activity and quantify their relative importance. To this aim, the PLS-coefficients (Figure 4) and activity-contribution plots (Figure 5) are useful. The former provides a global view and gives information on all of the training set. The sign and the magnitude of PLS coefficient of an energy term multiplied by the corresponding energy term (field) shows the influence of the corresponding residue on ligand binding.^{6c} Interpretation of the PLS coefficients can lead, however, to possible misconceptions. A positive PLS coefficient for an attractive, negative energy term indicates a term that contributes favorably to binding affinity (resulting in a more negative ΔG value). A positive PLS coefficient for a repulsive, positive energy term indicates a term that is unfavorable for binding affinity (resulting in a more positive ΔG value). On the other hand, a negative PLS coefficient will result in an energy term favoring binding when the energy term is positive (repulsive) and disfavoring binding when the energy term is negative (attractive).³⁸ The PLS-coefficient plot is shown in Figure 4A.

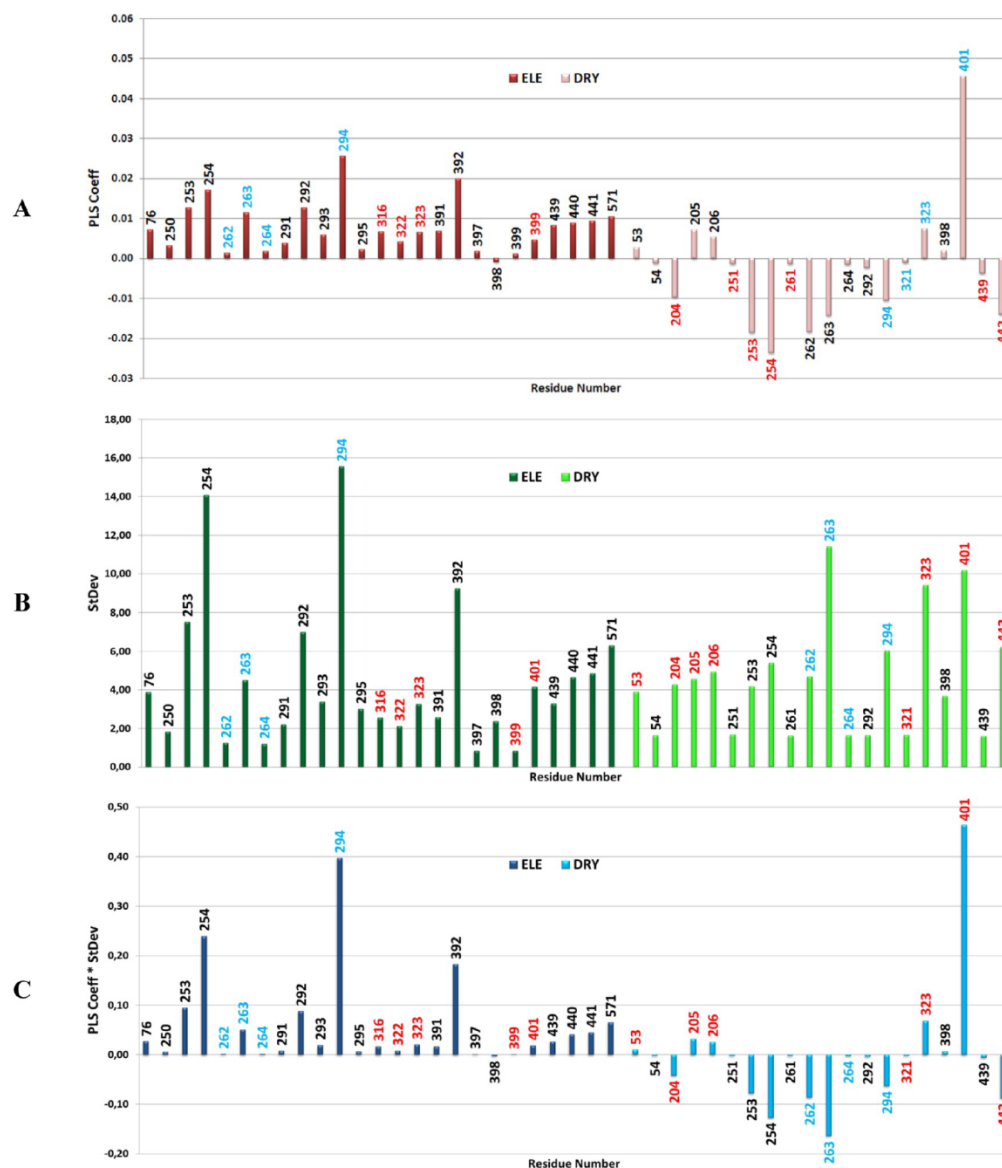


Figure 4. PLS Coeff (A), StDev (B), and PLS Coeff \times StDev plots for the ELE+DRY COMBINER model (C). The residue were selected using a PLS-coefficient threshold value of 0.001. Residue numbers are color-coded according to Table 8. The residue numbers reported correspond to those in Supporting Information File 6.

By multiplying the PLS coefficients with the field values, the activity-contribution plots were obtained for each training-set compound (see below). As can be seen (Table 8 and Figure 4), the COMBINER model can explain isoform selectivity considering only 34 residues of the enzymes (Table 8) even

though all residues of the eleven HDAC isoforms with a PLS coefficient greater than 0.001 were included in the analyses.

To analyze the significance of the fields (ELE and DRY) and the contribution for each ligand–residue interaction, the residues were color-coded in Table 8. The residues located in the rim region are colored red, while the residues forming the central

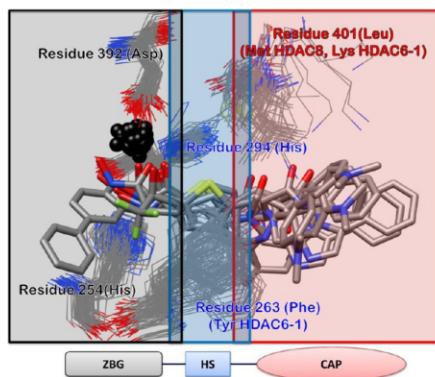


Figure 5. Four most important residues (MIRs, see text) from the COMBINER model analysis. The labels and regions are color-coded: in red are the residues in the HDAC's rim region; in blue are those forming the central tube channel; and in black are those in the proximity of the catalytic Zn ion. The zinc binding region (black line box), the connection region (blue line box), and the CAP region (red line box) are also highlighted to recall the HDAC pharmacophore model depicted at the bottom. ZBG: Zn-binding group. HS: hydrophobic spacer. CAP: hydrophobic capping group.

channel are blue, and those in proximity to the catalytic Zn ion are black (Supporting Information File 2). In Figure 4B are reported the ligand/residue/interactions standard deviations (StDev) used to produce the PLS Coeff \times StDev plot (Figure 4C) in which the PLS coefficients are weighted so that the global importance of the interactions can be understood similar to a standard 3-D QSAR model.³⁹ The variables reported in Figure 4 and Table 8 are significant for the model; however, the most important residues that modulate the inhibitory activities are as follows: 254 (His for all the HDACs, in the Zn-binding site), 294 (His for all the HDACs, either in the Zn- or tube-binding sites), and 392 (Asp for all the HDACs, in the Zn-binding site) mainly for the ELE field while 263 (Tyr for HDAC-6-1 and Phe for all the others in the tube-binding sites) and 401 (Met for HDAC-8, Lys for HDAC-6-1 and Leu for all the others, in the rim-binding site) for the DRY field (Figure 5). Residue 254 has also some negative modulating factor in the DRY field. These five residues account for 95% of the explained variance (\sim 80%) of the model indicating that interactions of ligands with these four residues are of major importance in determining the inhibitor potencies (coarse tuning, Figure 6). Fine tuning of both potency and selectivity result from other contributions, and therefore, each isoform needs to be inspected individually.

Regarding the importance of the overall interactions, the sums for either the ELE or DRY activity contributions for each training-set complex are shown in Figure 7. While the DRY field contribution mostly modulates the activities (bigger red bars on bulkier compounds), the ELE contribution becomes more important in modulating the low activities of the smaller inhibitors (bigger blue bars on short fatty acid inhibitors), NABUT and VALPROIC ACID (VA), due to missing interactions with residue 401 and others at the enzymes' rims (Figure 8). Indeed, the COMBINER model correctly indicates that NABUT and VA miss residue 401's contributions so activity contributions from other main residues (254, 294, and 392) of the

ELE field) are highly negative ranging from -0.27 to -1.02 and from -0.14 to -1.02 for NABUT and VA, respectively.

Field ELE. All residues selected having PLS Coeff higher than 0.001, except for 398, showed positive values, indicating that all the electrostatic interaction are attractive (Figure 4A). Indeed the PLS Coeff \times StDev plot clearly indicates that all electrostatic interactions are positively contributing to the model. In particular, plots in Figure 4 show that the ELE field is definitively more important in the inner part (black-labeled residues) of the HDACs catalytic domains than for residues forming the channel (blue-labeled residues in Figure 4) and those at the entrance rim (red-labeled residues in Figure 4) where only four and five out of 27 residues displayed PLS coefficients higher than the chosen threshold value.

In the outer part of the enzymes, the five selected residues (Figure 4) do not show appreciable activity contributions highlighting that these parts are not associated with high variation in ligand/enzyme electrostatic interactions. Detectable negative values relate to small compounds (NABUT and VA) for which the model correctly records the missing contribution.

Regarding the channel-forming residues, 294 (at the edge between the channel and the bottom of the HDAC-binding sites) displayed the highest values in all three plots of Figure 4. Indeed, this residue (a conserved histidine for all HDACs) is primarily involved in modulating the potency between small inhibitors (NABUT and VA) and channel-filling inhibitors (i.e., SAHA and TSA). For NABUT and VA, diminished interactions with residue 294 account for 0.8 to 1.0 decrement in activity. To some extent, the fact that either NABUT or VA are carboxylic acids indicates that higher negative charge (NABUT and VA were modeled as carboxylates, thus bearing a discrete negative formal charge) in proximity to residue 294 is unproductive. Analogous to a CoMFA analysis, the high PLS Coeff \times StDev values for residue 294 represent a blue polyhedron, placed in the same space of 294, indicating that an enhanced negative charge decreases the overall activity, while a positive-charged group (or a less negative one) is preferred to maintain the activity (the maximum contribution associated with 294 is lower than 0.01). Among the other channel-forming selected residues, 262 (always a Gly), 263 (mostly a Phe), and 264 (always a Cys), the most interesting is residue 263 involved in modulating the activity decrement for small compounds, in particular for VA.

Most of the ELE-selected residues (18 out of 27) are in the deep part of the channels around the catalytic Zn. Of particular interest are residues, involved in HDAC catalytic process conserved among the 12 isoforms, as follows: residues 253 (His), 254 (His), 292 (Asp), 392 (Asp), and 571 (Zn). In general the activity contribution associated with these five residues modulates the activity decrement for carboxylate-based zinc-binding groups. As examples, residues 253 (SAHA in HDAC-1) and 254 (SAHA in HDAC-3, HDAC-4, and HDAC-6-2; and SBHA in HDAC-4 and HDAC-8) are associated with a positive activity contribution of about 0.1.

Field DRY. The DRY field gives a rough estimation of steric interactions. Between ELE and DRY selected residues about 35% of these are shared (12 out of 34) in significance, nevertheless, for the DRY field a totally different and more complicated scenario can be observed on the relative importance of each residue. In general, the most important modulating interaction relates to 401Leu, replaced by Met in HDAC-8 or by Lys in HDAC-6-1 (Table 8). Upon deeper inspection (not considering the small-molecule complexes, NABUT, VA, and NHB), only 27 of 94 activities are modulated by residue 401 with activity contribu-

Table 8. List of the Most Important Residues to Interpret the COMBINER Model⁴²

N. of residuals	53	54	76	204	205	206	250	251	253	254	261	
HDAC1	HIS28	PRO29	ARG34	GLU98	-	-	GLY138	LEU139	HIS140	HIS141	SER148	
HDAC2	HIE22	PRO23	ARG28	GLU92	-	-	GLY132	LEU133	HIE134	HIS135	SER142	
HDAC3	HIS22	PRO23	ARG28	ASP92	-	-	GLY132	LEU133	HIS134	HIS135	SER142	
HDAC8	-	-	ARG37	TYR100	-	-	GLY140	TRP141	HIS142	HIS143	MET151	
HDAC4	-	-	ARG32	-	-	-	PRO151	GLY152	HIE153	HIE154	MET840	
HDAC5	HIS704	PRO705	ARG710	-	-	-	PRO830	GLY831	HIS832	HIS833	ASP137	
HDAC7	HIE27	PRO28	ARG33	-	-	-	PRO153	GLY154	HIE155	HIE156	CYS137	
HDAC9	-	-	ARG660	-	-	-	PRO780	GLY781	HIS782	HIS783	MET163	
HDAC6-1	PHE19	PRO20	ARG25	THR84	TYR85	-	PRO127	GLY128	HIS129	HIS130	SER150	
HDAC6-2	HIS19	PRO20	ARG25	-	-	PHE85	PRO127	GLY128	HIS129	HIS130	MET790	
HDAC10	GLU24	ILE25	ARG30	-	-	-	PRO132	GLY133	HIS134	HIS135	ASN142	
Class IV	HDAC11	HIS35	PRO36	LYS41	PRO102	-	GLY140	PHE141	HIS142	HIS143	GLY150	
N. of residuals	262	263	264	291	292	293	294	295	316	321	322	
HDAC1	GLY148	PHE150	CYS151	ILE175	ASP176	ILE177	HIS178	HIS179	LYS200	GLU203	TYR204	
HDAC2	GLY143	PHE144	CYS145	ILE169	ASP170	ILE171	HIE172	HIE173	LYS194	TYR198	-	
HDAC3	GLY143	PHE144	CYS145	ILE169	ASP170	ILE171	HIS172	HIS173	LYS194	ASN197	TYR198	
HDAC8	GLY151	PHE152	CYS153	LEU177	ASP178	LEU179	HIE180	HIS181	LYS202	GLY206	PHE207	
HDAC4	GLY162	PHE163	CYS164	TRP190	ASP191	VAL192	HIE193	HIE194	ARG215	ASN220	PHE221	
HDAC5	GLY841	PHE842	CYS843	TRP869	ASP870	ILE871	HIS872	HIS873	ARG894	ASN899	PHE900	
HDAC7	GLY164	PHE165	CYS166	TRP192	ASP193	VAL194	HIE195	HIE196	ARG217	ASN222	PHE223	
HDAC9	GLY791	PHE792	CYS793	LEU819	ASP820	VAL821	HIS822	HIS823	ARG844	ASN849	PHE850	
HDAC6-1	GLY138	TYR139	CYS140	TRP166	ASP167	VAL168	HIS169	HIS170	ARG191	ARG196	PHE197	
HDAC6-2	GLY138	PHE139	CYS140	TRP167	ASP168	VAL169	HIS170	HIS171	ARG192	THR197	PHE198	
HDAC10	GLY143	PHE144	CYS145	TRP171	ASP172	VAL173	HIS174	HIS175	ARG196	ARG201	PHE202	
Class IV	HDAC11	GLY151	PHE152	CYS153	LEU180	ASP181	ALA182	HIS183	GLN184	ASN205	ILE208	TYR209
N. of residuals	323	391	392	397	398	399	401	439	440	441	442	
HDAC1	-	SER263	ASP264	ASP269	ARG270	-	LEU271	GLY300	GLY301	GLY302	TYR303	
HDAC2	PHE199	ALA257	ASP258	ASP263	ARG264	-	LEU265	GLY294	GLY295	GLY296	TYR297	
HDAC3	PHE199	ALA258	ASP259	ASP264	ARG265	-	LEU266	GLY295	GLY296	GLY297	TYR298	
HDAC8	PHE208	ALA266	ASP267	ASP272	PRO273	-	MET274	GLY303	GLY304	GLY305	TYR306	
HDAC4	PHE222	PHE284	ASP285	HIE290	PRO291	THR292	LEU294	GLU324	GLY325	GLY326	HIE327	
HDAC5	PHE901	PHE963	ASP964	HIS969	LEU970	SER971	LEU973	GLU1003	GLY1004	GLY1005	HIS1006	
HDAC7	PHE224	PHE286	ASP287	HIE292	PRO293	ALA294	LEU296	GLU326	GLY327	GLY328	HIE329	
HDAC9	PHE851	PHE913	ASP914	HIS919	THR920	PRO921	LEU923	GLU953	GLY954	GLY955	HIS956	
HDAC6-1	TRP198	PHE259	ASP260	ASP265	PRO266	-	LYS267	GLU297	GLY298	GLY299	TYR300	
HDAC6-2	PHE199	PHE260	ASP261	ASP266	PRO267	-	LEU268	GLU298	GLY299	GLY300	TYR301	
HDAC10	TRP203	PHE264	ASP265	ASP270	PRO271	GLU272	-	GLU302	GLY303	GLY304	TYR305	
Class IV	HDAC11	-	THR260	ASP261	ASP266	ARG267	-	LEU268	SER301	GLY302	GLY303	TYR304

⁴²The labels are color-coded: in red are the residues in the HDAC's rim region; in blue are those forming the central tube channel; and in black are those in the proximity of the catalytic Zn. The residues labels were color-coded according to the reported pharmacophoric model.¹ The residues were selected using a PLS-coefficient threshold value of 0.001. See Supporting Information File 001 for 3-D graphical disposition of the listed residues in each HDAC isoform.

tions ranging between 0.7 and 2.13 (Supporting Information File 1 Figure SI-1).

Without considering the contribution of residue 401, it is evident from the plot in Figure 8B that the other 10 residues play a major role in modulating the overall biological activities (Supporting Information File 1 Figure SI-2, Table SI-1). Seven out of 10 residues (204, 253, 254, 262, 263, 294, and 442) are related to negative modulating values, while the other three (205, 206, and 323) are positive modulators. Residue 263 (Tyr for HDAC-6-1 and Phe for the others) located in the wall of the channel shows the largest range with larger negative values. No

specific pattern is detected for residue 263 in modulating regarding the different enzyme classes or inhibitor structures (Supporting Information File 1 Figure SI-3). The small inhibitor NABUT is not influenced by residue 263, likely due to the fact that there are no direct contacts. Residue 442 (His for Class IIa and Tyr for the others) located in the bottom of the binding sites shows the largest range with larger negative values associated mainly with class I complexes, with particular reference to HDAC-8 (Supporting Information File 1 Figure SI-4) thus suggesting that interaction with this residues might be used to selectively avoid inhibition of HDAC-8.

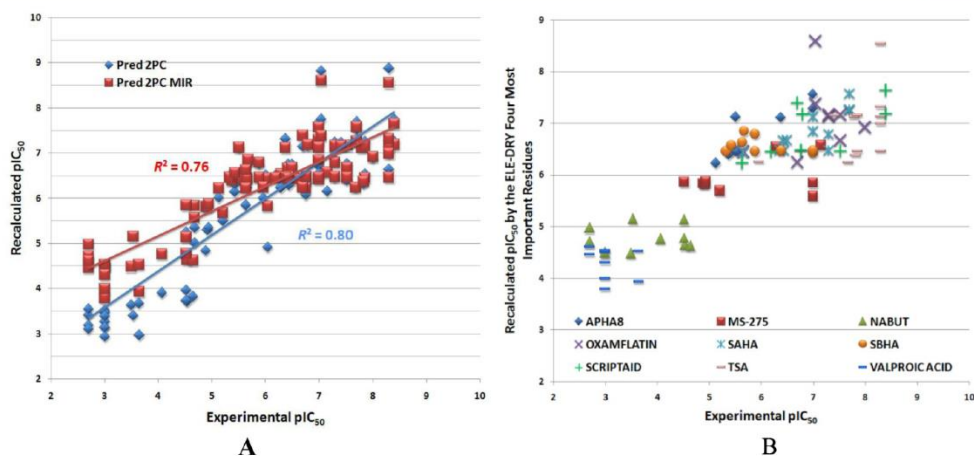


Figure 6. Comparison between the cross-validation predictions for the full model (blue squares) and with only the four most-important residues (MIRs). The coarse tuning of the relationships by the MIRs is indicated by the red squares in panel A. The differences between the red and blue squares indicate the importance of fine-tuning determined by relatively minor interactions. In B are reported the MIR predictions classified by inhibitor type. For comparison purposes, only inhibitors for which isozyme profiles of inhibition data were available are shown.

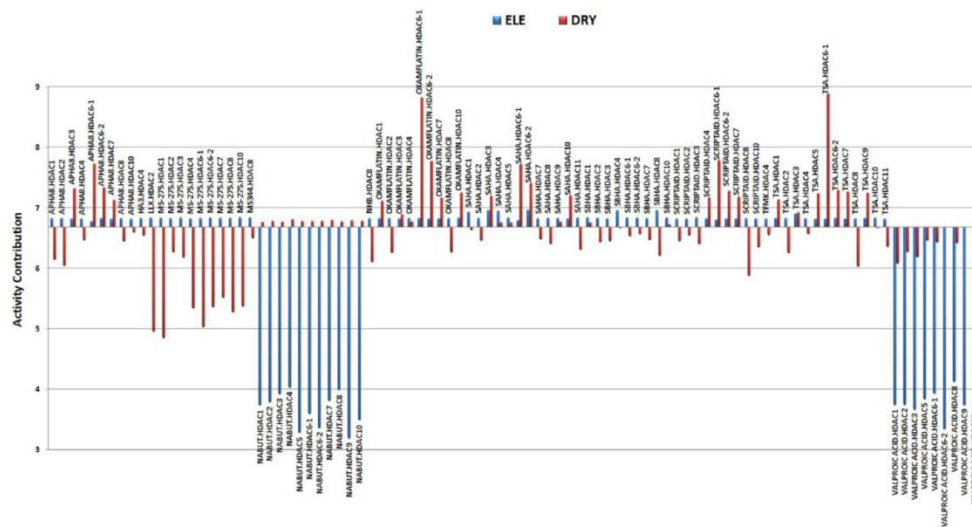


Figure 7. ELE and DRY total-activity contributions. The constant (PLS intercept) of the COMBINER equation takes the value of 6.68. The sum of ELE and DRY contribution is obtained by the algebraic sum of all per-residue contributions.

Residue 254 (His in the zinc-binding region) is second with the higher StDev value and from Supporting Information File 1 Figure SI-5 clearly negatively modulates mainly nonhydroxamate inhibitors making complexes (LLX, MS-275, and VA) consistently with that reported for the ELE field. Residue 204 (of various nature present on the rim of 6 out of 12 HDACs) and 294 (His, a channel-forming residue) are also negative-modulating residues, but the associated low standard deviation indicates that no selectivity can be attributed to the DRY

interactions (Supporting Information File 1 Figure SI-6 and 7); residue 204 seems to specifically modulate the inhibitory activity for HDAC-8 complexes (Supporting Information File 1 Figure SI-7). Considering the high correlation between DRY and STE, interactions with residues 263 and 294 are of crucial importance for optimal fitting of inhibitors in the HDAC channels.

Among the three DRY positive-modulating residues, 323, an aromatic side-chain-bearing residue missing in HDAC-1 and HDAC-11, shows the highest maximum-activity contribution

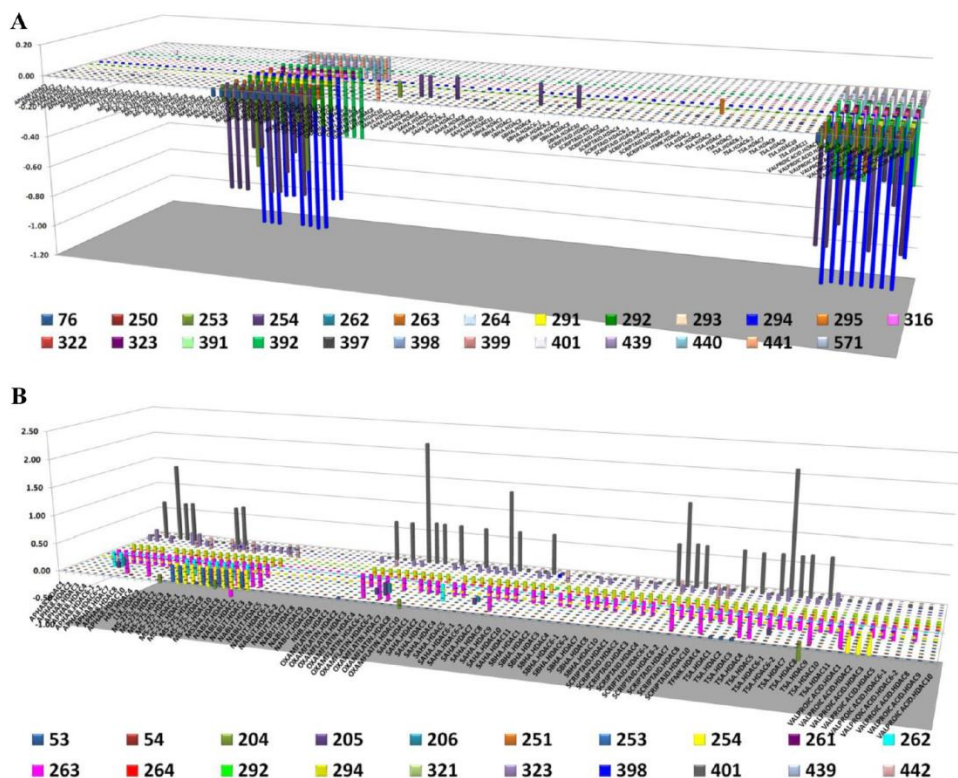


Figure 8. Per-residue activity-contribution plots for the ELE (A) and DRY (B) fields.

and larger variability; maximum-activity contributions occur with APHA8 and TSA binding to either class I or class II enzymes (Supporting Information File 1 Figure SI-8). The other highly positively contributing residue 205 is peculiar for HDAC-6-1 (Tyr85) and thus uniquely modulates inhibition of this enzyme.

Analysis of Interactions Contributing to Isoform Selectivity. Interaction- and activity-contribution analyses suggest that useful insight into structural determinants exists for both HDAC isoforms and their inhibitors to help optimize isoform-specific inhibitors using the derived COMBINER model. Derivation of rules to guide the structural basis for isoform selectivity required single analysis for each specific isoform model. For nine of the inhibitors used in the training set (Table 2), at least 9 out of 12 isoform-inhibition profiles were available (Supporting Information File 1 Table SI-2). In Supporting Information File 3 are reported the recalculated activity profiles for each of the nine inhibitors of Table 2 showing the models sensitivity to HDAC-isoform inhibition by different compounds. To illustrate the COMBINER model's potential use, two inhibitors were selected seeking potential structure determinants for isoform selectivity. Among the training set, analysis on the activity range indicated MS-275 and SCRIPTAID as good examples. From Supporting Information File 1 Table SI-2, MS-275 and SCRIPTAID display large variability, and from Table 2, MS-275 results partially selective for class I HDACs (particularly

for HDAC-3 $IC_{50} = 0.07 \mu\text{M}$ and HDAC-2 $IC_{50} = 0.5 \mu\text{M}$), while SCRIPTAID is partially selective for class II displaying submicromolar activities against these enzymes.

MS-275. This inhibitor is specifically selective for class I HDAC-3 over class IIa HDAC-4 and comparison of data belonging to the relative complexes shows how the model helps rationalize the higher activity of MS-275 for HDAC-3 versus HDAC-4. As shown in Figure 9, it is possible to indicate, either numerically or graphically, the residues responsible for this activity difference. Considering electrostatic interactions, it is evident that, as already above highlighted, there is very low correlation with activity, and only gray or light blue surfaces can be observed in Figures 10C and E (see Figure 9 footnotes for color-coding). On the other hand, the DRY field seems very sensitive as shown in Figure 9D and F; there is a high color variation clearly indicating those residues responsible for the higher activity of MS-275 against HDAC-3 (Phe199 and Arg265 are dark green). Other green-colored residues are also located around the rim, for example, Leu266. A few residues are colored yellow, residue 263 (Phe144 in Figure 9D) indicating that MS-275 anti-HDAC-3 activity could be improved by optimizing the interactions in the enzyme channel. Going to the MS-275/HDAC-4 complex, many DRY surfaces have turned from green to yellow thus highlighting that residue 263 (HDAC4-Phe163)

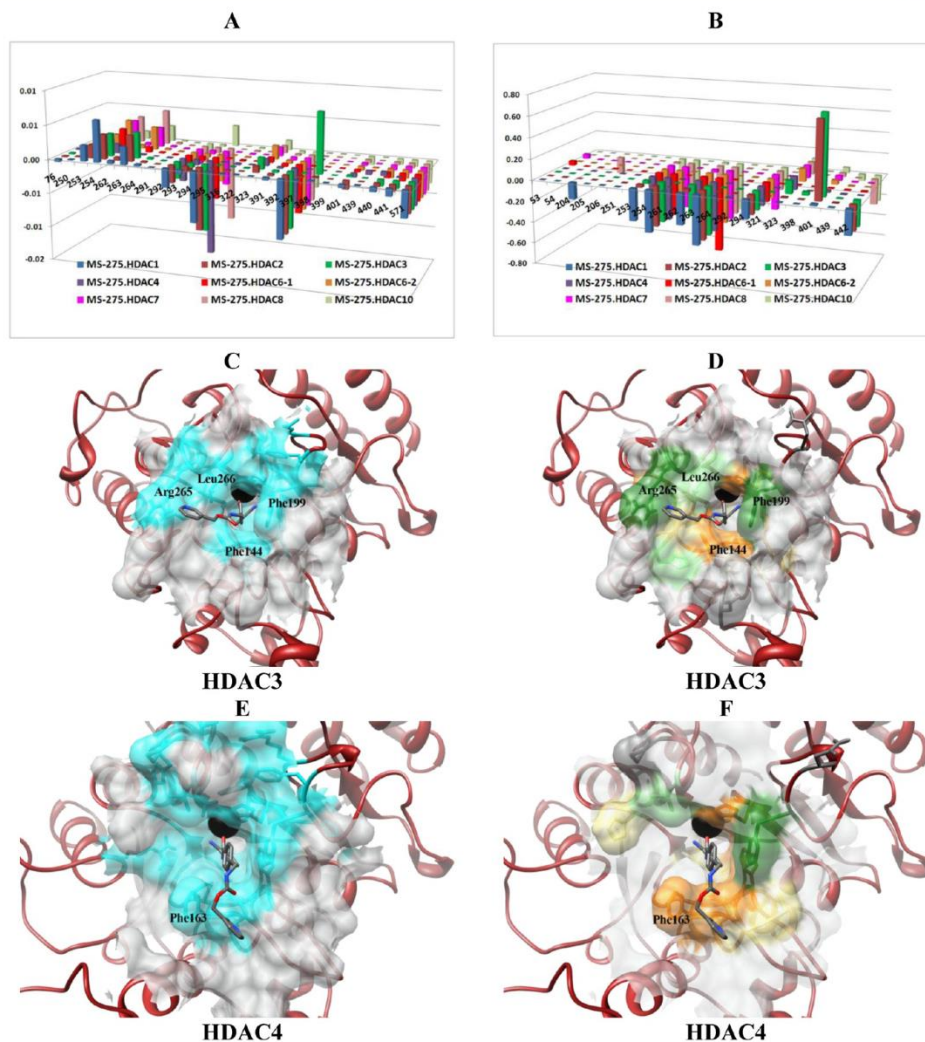


Figure 9. Activity contributions (A and B) for MS-275 and their graphical representations (C–F). A, C, and E in the left side, account for the ELE field. The DRY is depicted on the right (B, D, and F). Residue surfaces are color-coded: for ELE blue-based surfaces indicate a positive contribution (light blue if the contribution is less than 50% of maximum contribution for a given residue; dark blue indicate areas with higher contributions); red-based surfaces indicate negative contributions (light red for absolute contribution less than 50% of the corresponding residue; dark red for higher percentage of negative contribution). For the DRY field, positive contributions are indicated in green (dark green for contribution higher than 50% of the maximum activity contribution; light green for less contribution); yellow colors are used to indicate negative DRY contributions (dark yellow for absolute contribution higher than 50% of the maximum activity contribution; light yellow for low negative contributions). Dark gray surfaces indicate zero contribution, while light gray are residues with PLS coefficients lower than 0.001 (see above in the model interpretation section). Only residues cited in the text are labeled.

plays a major role in decreasing activity with many residues showing zero activity contribution.

SCRIPTAID. SCRIPTAID was chosen as a selective class II inhibitor. Similarly to MS-275, the electrostatic interactions differentiated when comparing the activity contributions of HDAC-6 and HDAC-8 (Figure 10). Indeed, Figure 10A clearly indicates that the ELE contributions are below 0.02. So

analogously to MS-275, DRY terms help rationalize the inhibitory activities of SCRIPTAID with HDAC-6 and HDAC-8. Most differences are located in the rim zone. Specifically, Lys267 in HDAC-6 is responsible of a strong positive contribution, while Met261, its counterpart in HDAC-8, displays a much smaller contribution.

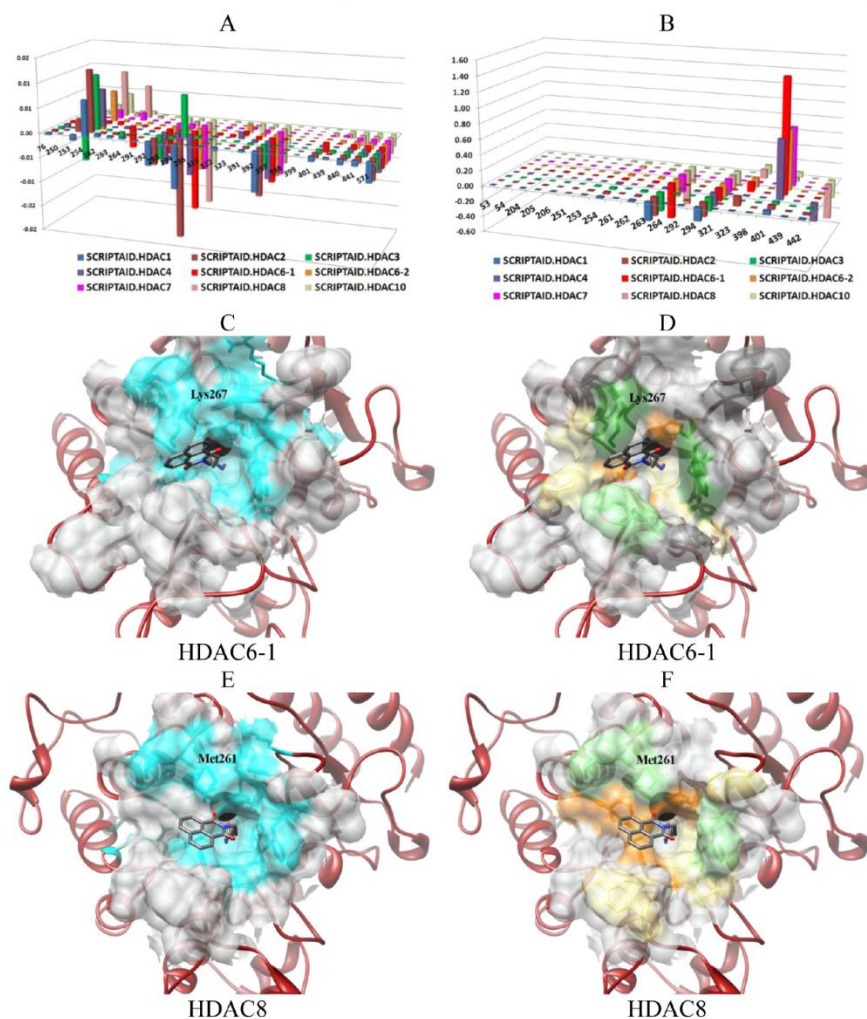


Figure 10. Activity contributions (A and B) for SCRIPTAID and their graphical representations (C–F). A, C, and E account for the ELE field. The DRY field is depicted on the right (B, D, and F). Residue surfaces are color-coded: for ELE, blue-based surfaces indicate positive contributions (light blue if the contribution is less than 50% of maximum contribution for a given residue; dark blue indicate areas with higher contributions); red-based surfaces indicate negative contributions (light red for absolute contributions less than 50% of the corresponding residue; dark red for higher percentage of negative contributions). For the DRY field, positive contributions are indicated in green (dark green for contribution higher than 50% of the maximum activity contribution; light green for less contribution); yellow colors are used to indicate negative DRY contribution (dark yellow for absolute contribution higher than 50% of the maximum activity absolute contribution; light yellow for low negative contributions). Dark gray surfaces indicate zero contributions, while light gray are residue with PLS coefficients lower than 0.001 (see above in the model interpretation section). Activity contribution plots and associated graphical for all the training set are reported in Supporting Information Files 4 and 5. Only residues cited in the text are labeled.

Docking Assessment. X-ray structures of HDAC-inhibitors were used to evaluate the ability of a docking program to predict the correct geometry of protein–ligand complex (redocking). To this aim, two different docking programs were tested: AutoDock Ver. 4.2 and AutoDockVina Ver. 1.1. Docking results were assessed with RMSD (root-mean-square deviation) of the predicted ligand configuration versus the crystal structure. Table

9 and 10 show RMSD values for best docked (the lowest energy docked conformation of the first cluster generated), best cluster (the lowest energy docked conformation of the most populated cluster), and best fit (the lowest energy conformation of the cluster showing the lowest RMSD value),⁴⁰ obtained with the two programs. In all cases, AutoDockVina was found to be more accurate displaying a docking accuracy (DA) of 75% for the best

Table 9. Redocking Results (RMSD) with AutoDock Program

complex name	best docked	best cluster	best fit
LLX.HDAC2	0.48	0.48	0.48
HA3.HDAC4	5.25	4.76	4.4
TMFK.HDAC4	3.46	5.75	3.46
SAHA.HDAC7	10.36	10.36	2.18
TSA.HDAC7	6.06	6.06	1.4
APHA.HDAC8	5.4	2.26	2.26
SAHA.HDAC8	5.84	7.29	4.1
TSA.HDAC8	5.1	5.52	1.45
DA %	12.5	18.75	50

Table 10. Redocking Results (RMSD) with AutoDockVina Program

complex name	best docked	best cluster	best fit
LLX.HDAC2	0.24	0.24	0.24
HA3.HDAC4	3.87	2.34	1.93
TMFK.HDAC4	4.02	1.9	1.46
SAHA.HDAC7	2.45	2.45	1.88
TSA.HDAC7	2.19	2.19	1.21
APHA.HDAC8	1.43	1.43	1.43
SAHA.HDAC8	2.49	2.49	1.72
TSA.HDAC8	2.09	1.22	1.22
DA %	50	75	100

cluster poses (Tables 9 and 10). AutoDockVina was able to predict the right binding disposition of all ligands with a RMSD <3 Å. From Tables 9 and 10, the best cluster conformation displayed the lowest RMSD values. For subsequent dockings, therefore, only the AutoDockVina program was used considering

the best cluster conformation as the first choice. Considering the Best Fit pose, AutoDockVina proved to be able to find the correct binding mode with a DA of 100%. Although the Best Fit poses are irrelevant for the docking applicability, they further supported that AutoDockVina is quite good in searching for the right conformation, but the scoring function is not able to select it. For docking, the side-chain flexibility features of AutoDock and AutoDockVina were not used as the results were always worse than in fixed receptor dockings in preliminary docking studies.

Model Predictivity. Once the docking protocols were assessed, a cross-docking approach was applied to the MTS, CTS, and LTS test sets of inhibitors to prepare the HDAC- α complexes.

Modeled Test Set. Regarding the MTS, all minimized HDAC structures were used as templates for docking simulations. Thus, each inhibitor of Table 4 was docked into all receptor binding sites, for a total of 304 individual docking simulations. For each isoform, all poses were collected in a bin and the output poses clustered by means of the AutoDock program. It was found that AutoDockVina had the ability to reproduce the experimental binding modes with modest errors (Table 10); in some cases (not shown), the best cluster conformation was found in a nonactive pose (i.e., the zinc-binding group rotated away from the Zn ion). This clearly indicated the limitations of the docking protocol in selecting the correct poses. In these cases, either the best-docked pose or an arbitrary-chosen conformation on the basis of Zn chelation that mimicking the binding mode of closest-related experimentally bound inhibitor was used. This approach is consistent with the fact that AutoDock Vina proved to be able to find the right binding mode (see comments for the Best Fit pose in the Docking Assessment section). For MTS, a total of 76 HDAC-inhibitors complexes were compiled, and the ELE+DRY

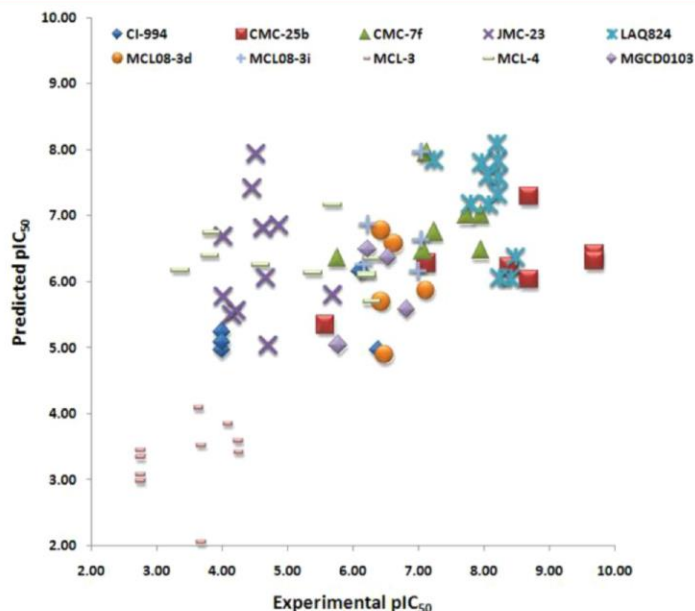
Figure 11. Experimental/predicted pIC_{50} for the MTS.

Table 11. Experimental/Predicted pIC_{50} for the CTS Test Set^a

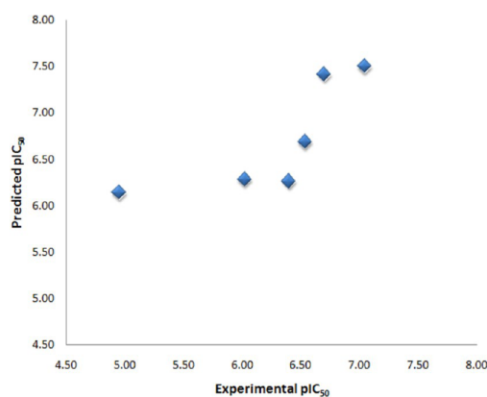
PDB code	HDAC	molecule name	experimental	PC1	PC2	PC3	PC4	PC5
3SFH ³²	HDAC8	1DI	7.05	8.81	7.50	7.34	7.19	7.08
3SFH ³²	HDAC8	0DI	6.70	8.90	7.41	7.21	6.90	6.96
1ZZ3 ³⁵	HDAH	3YP	6.54	6.46	6.69	6.39	6.48	6.34
2GH6 ³⁴	HDAH	CF3	4.95	6.53	6.14	5.99	6.02	6.05
1ZZ1 ³⁵	HDAH	SAHA	6.02	6.72	6.28	6.01	5.82	5.76
1C3R ³³	HDLP	TSA	6.40	6.72	6.26	6.45	6.58	6.76

^aPDB codes are used for indicating the structure of inhibitor from Table 5. The predicted values at different principal components (PCs) are reported.

COMBINER model was used to predict inhibitors activities. Figure 11 and Supporting Information File 1 Table SI-3 show the pIC_{50} predicted for the MTS external test set and statistical results (SDEP_{ext} and AAEP). The model showed a good external predictivity with SDEP of 1.41 for the optimal two principal components. Figure 11 reveals that JMC-23 and MCL-4 are the worst predicted compounds. JMC-23 contains an oxime amide as a ZBG (Zn binding group) that can be interpreted as a modified version of the efficient hydroxamate moiety. As reported by Botta et al.,²⁷ this compound is a poor pan-HDAC inhibitor; the COMBINER model fails to predict correctly 5 out of 11 activities. Regarding MCL-4, this is the hydroxamate version of MCL-3, while the latter is recognized as a very poor inhibitor with the correct trend, MCL-4 is highly over predicted in HDAC-4, HDAC-5, HDAC-7, and HDAC-9 complexes. Nevertheless the average pIC_{50} value for MCL-4 (exp = 5.18, pred = 6.31) was correctly calculated to be higher than that for MCL-3 (exp = 3.40, pred = 3.33).

Comparisons of predictions for single HDAC isoforms reveal that complexes of HDAC-2 and HDAC-3 were the best predicted with an average absolute error of prediction (AAEP) of 0.53 and 0.65, respectively. Complexes related with HDAC-7, HDAC-9, HDAC-10, and HDAC-11 showed the highest AAEP values. For HDAC-9, HDAC-10, and HDAC-11, the worst predictions were associated with a lower number of complexes in the training set. In general, the model was able to reproduce the activity of class I HDACs better than class II. Regarding HDAC-10 and HDAC-11, the smaller amount of experimental data in the training set was the probable cause for the failed activity-trend predictions (Supporting Information File 1 Figure SI-10, Panels K and L). Notably the external SDEP value confirmed that the model at 2 PCs was indeed the most predictive as correctly indicated by the cross-validation runs (Supporting Information File 1 Table SI-3). The application of the COMBINER model to the MTS proved the ability of the model in predicting the relative potency and the correct activity trend of a given series of inhibitors for 10 out of 12 HDAC isoforms (Supporting Information File 1 Table SI-3 and Figure SI-10) even when the binding conformations of the test set inhibitors were obtained from docking. Furthermore the lowest SDEP_{ext} and AAEP values obtained from the MTS analysis fully supported the optimal number of PCs as indicated by cross-validation.

Crystal Test Set. The CTS was compiled using only experimental bound inhibitors. The usefulness of this test set was twofold. First, from Table 11, the training-set model-binding conformations were confirmed to be self-consistent with only 2 PCs (Figure 12), the COMBINER model predicted the correct trend and activity potencies with an AAEP values of only 0.71 (not shown). Second, the inclusion of bacterial HDACs (HDAH and HDLP) indicates that the derived COMBINER model might be used to predict activities against nonhuman HDACs,

Figure 12. Experimental/predicted pIC_{50} for the CTS.

potentially useful in the search for antiparasitic, antifungal, and antibacterial therapeutics.

Largazole Test Set. Finally the third test set comprised a cyclotetrapeptide-like inhibitor (largazole).³¹ In this case the model was tested for its predictive ability against a class of inhibitor (peptide-like) totally different from those included in the training set. To some extent, the COMBINER model was able to recognize the relative potency of largazole for HDAC-1, HDAC-2, and HDAC-6-1; while for HDAC-3, the predicted pIC_{50} was underestimated, indicating that further modeling of this class of inhibitor is needed (Table 12 and Figure 13). As a

Table 12. Experimental/Predicted pIC_{50} for the LTS Test Set^a

	exp	PC1	PC2	PC3	PC4	PC5
HDAC-1	8.92	6.98	7.64	8.03	7.88	8.09
HDAC-2	8.46	6.94	7.72	7.59	7.23	7.33
HDAC-3	8.47	6.80	6.73	6.97	6.80	6.86
HDAC-6-1	7.31	7.12	6.47	6.26	5.77	6.35

^aThe predicted values at different principal components (PC) are reported.

matter of fact, the docking approach used did not allowed flexibility of the largazole cyclic headgroup; thus, full ligand flexible docking and smaller error of prediction should be expected with enhanced docking and inclusion of more inhibitors that interact with the headgroup region.

CONCLUSION

A structure-based 3-D QSAR model using comparative binding-energy analysis that focused on the selectivity of the 11 human

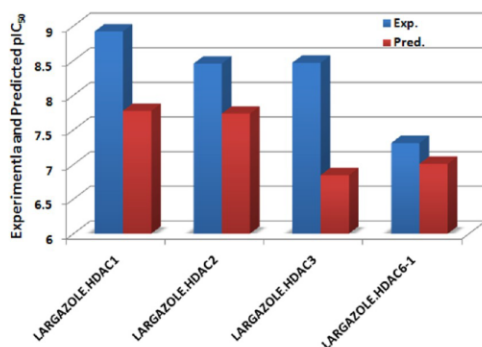


Figure 13. LTS predictions at two PCs. The X-axis represents HDACs complexes with largazole, and the y-axis represents biological activity values measured as pIC_{50} .

zinc-based histone deacetylase isoforms has been developed through a modified protocol called COMBINER. The derived COMBINER model shows good statistical coefficients, was predictive for the compounds in the test sets, and robust to cross-validation while omitting multiple data. The model was able to rationalize the different activity profiles of the HDAC inhibitors studied. This model should provide a useful tool for the *a priori* prediction of activity of compounds yet to be synthesized in order to improve their selectivity profiles. The role of dynamic acetylation in epigenetics and other signaling pathways³ provides strong motivation for the development of molecular scalpels, specific inhibitors of histone deacetylases, to dissect the complexities of epigenetic control of gene expression and other signaling pathways. The COMBINER model should prove useful in this endeavor.

EXPERIMENTAL SECTION

All molecular graphics images were produced using UCSF Chimera package (www.cgl.ucsf.edu/chimera/) from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, on a 3 Ghz AMD CPU-equipped, IBM-compatible workstation using the Debian 5.0 version of the Linux operating system. For all calculations, a Beowulf cluster of 12 quadcore Xeon CPUs was used.

Complex Preparation. Inhibitor Structures. All ligands were generated with Chemaxon Marvin molecular mechanics software (<http://www.chemaxon.com/>) and used without further optimization. The protonation and tautomer states were assigned considering a physiological pH and the more common tautomer according to basic organic chemistry and structural information reported in the corresponding ligand referenced papers.

HDAC Homology Models. Those HDAC isoforms whose experimental structures were not available (HDAC-1, -3, -5, -6-1, -6-2, -9, -10, and -11), were built by homology modeling using four automated web servers:

- CPHmodels-3.0 Server⁴¹ (<http://www.cbs.dtu.dk/services/CPHmodels/>),
- M4T Server ver.3.0⁴² (<http://manaslu.aecom.yu.edu/M4T/>),
- SwissModel⁴³ (<http://swissmodel.expasy.org/>),

- ModWeb Server⁴⁴ (<http://modbase.compbio.ucsf.edu/ModWeb20-html/modweb.html>).

Using several protein conformations for each HDAC isoform was done to include some target flexibility in the subsequent training set, test set cross-docking runs. For each HDAC isoform, four homology models were generated. All inhibitors were modeled into each of the four homology models, and the resulting complexes energy minimized to supply four complexes for each inhibitor leading to 220 complexes. The servers were used with their default parameters and in a total automatic way to avoid human intervention and to allow maximum reproducibility.

To compile the final training set of 94 complexes (see the Training Set section above), one homology complex per inhibitor was chosen using the preliminary COMBINER models derived with only crystallized HDAC complexes, (data not shown). For each inhibitor, the HDAC/inhibitor complex whose predicted pIC_{50} s had the best fit to the experimental pIC_{50} s for that isoform was selected and utilized in the final training set (Table 13).

Complex Minimization. Training set complexes were submitted to a single-point minimization using a protocol described earlier.⁴⁰ Briefly, the minimization protocol was applied as follows. (1) ANTECHAMBER with AM1-BCC charges was used to determine missing ligand parameters; (2) the tLeap module was used to solvate the complexes with water molecules in a octahedral box extending 10 Å and to neutralize them with Na^+ and Cl^- ions; (3) the structures were minimized with the Amber 2003 force field by energy minimization with the SANDER modules: 1000 steps of steepest-descent energy minimization followed by 4000 steps of conjugate-gradient energy minimization, with a nonbonded cutoff of 5 Å. Trials for longer nonbonded cutoff values were done without substantial differences; therefore, 5 Å was chosen for faster calculations. The Zn ion was treated as nonbonded, similarly as in several other applications where HDACs were reported.

COMBINER. Ligand/Residues Interactions. The calculation of the ligand/residue interactions was conducted similarly as previously reported.⁷ The AutoGrid module of AutoDock was used with its default setting to compute the interaction energies between each amino-acid residue of the enzymes and an inhibitor. AutoGrid used the united-atom AMBER force field and returned an energy value combining Lennard-Jones (LJ) and hydrogen-bonding (HB) energies between a target and each atom type (probe). The electrostatic interactions were calculated using a distant-dependent Coulombic function, and finally, a third score for hydrophobic interactions was also estimated. In its original use, AutoGrid calculated the interaction energies of a probe atom that was placed on a regularly spaced grid in which a molecular target (the protein) or a portion of it was buried. In this way AutoGrid returns what is called the molecular interaction field (MIF) of a given target, where at each grid point estimates the interactions value for LJ and HB (STE), electrostatic (ELE) and desolvation (DRY), saved in three distinct map files. In the COMBINER approach, the target was the inhibitor in the complex and the STE, ELE, and DRY interactions were calculated using a grid box centered, at each step, on each atom of the protein (the probe). To the grid is given a step size, so that the whole complex was contained within it, and thus only one value was returned (the center) for each field. The interaction energy for each amino acid of the enzyme was simply obtained by summing all the values for all residue atoms. The

Table 13. Predicted pIC₅₀ for the Modeled Complexes Inserted in the Final Training Set

HDAC	complex name	homology server	pIC ₅₀ exp	pIC ₅₀ pred
HDAC-1	APHA8/HDAC1	SwissModel	5.432	6.13
	MS-275/HDAC1	M4T	4.886	5.2
	SAHA/HDAC1	M4T	7	6.69
	SBHA/HDAC1	CPH	5.678	6.61
	TSA/HDAC1	CPH	8.301	6.78
	OXAMFLATIN/HDAC1	ModWeb	7.301	6.92
	NABUT/HDAC1	ModWeb	3.496	3.7
	VALPROICACID/HDAC1	ModWeb	3	3.2
	SCRIPTAID/HDAC1	ModWeb	6.77	6.2
	HDAC-3	APHA8/HDAC3	CPH	6.377
MS-275/HDAC3		CPH	7.155	6.4
SAHA/HDAC3		CPH	7.699	6.92
SBHA/HDAC3		SwissModel	6.387	6.2
TSA/HDAC3		SwissModel	8.301	6.64
OXAMFLATIN/HDAC3		SwissModel	8	6.43
NABUT/HDAC3		SwissModel	4.648	4.34
VALPROICACID/HDAC3		CPH	3.646	3.2
SCRIPTAID/HDAC3		SwissModel	7.523	6.17
HDAC-5		SAHA/HDAC5	CPH	6.423
	TSA/HDAC5	CPH	7.796	6.97
	NABUT/HDAC5	ModWeb.2	2.699	4.19
	VALPROICACID/HDAC5	ModWeb.3	2.699	3.43
	HDAC-6-1	APHA8/HDAC6-1	SwissModel	7
MS-275/HDAC6-1		ModWeb.1	4.678	5.32
SAHA/HDAC6-1		SwissModel	7.699	6.77
SBHA/HDAC6-1		CPH	7	6.25
TSA/HDAC6-1		SwissModel	8.301	7.62
OXAMFLATIN/HDAC6-1		SwissModel	7.046	7.68
NABUT/HDAC6-1		M4T	3	3.65
VALPROICACID/HDAC6-1		CPH	3	3.13
SCRIPTAID/HDAC6-1		SwissModel	8.398	7.63
HDAC-6-2		APHA8/HDAC6-2	CPH	7
	MS-275/HDAC6-2	M4T	4.678	5.68
	SAHA/HDAC6-2	CPH	7.699	6.44
	SBHA/HDAC6-2	ModWeb.1	7	6.2
	TSA/HDAC6-2	M4T	8.301	7.02
	OXAMFLATIN/HDAC6-2	CPH	7.046	7.1
	NABUT/HDAC6-2	CPH	3	4.7
	VALPROICACID/HDAC6-2	ModWeb.1	3	3.84
	SCRIPTAID/HDAC6-2	M4T	8.398	7.13
	HDAC-9	SAHA/HDAC9	ModWeb.1	6.5
TSA/HDAC9		ModWeb.1	7.419	7
NABUT/HDAC9		ModWeb.1	2.699	4.03
VALPROICACID.HDAC9		CPH	2.699	4.05
HDAC-10		APHA8/HDAC10	SwissModel	5.377
	MS-275/HDAC10	ModWeb.1	4.939	5.67
	SAHA/HDAC10	ModWeb.1	7	6.96
	SBHA/HDAC10	M4T	5.638	6.6
	TSA/HDAC10	CPH	8.301	6.21
	OXAMFLATIN/HDAC10	CPH	7.301	6.8
	NABUT/HDAC10	CPH	3.535	4.1
	VALPROICACID/HDAC10	M4T	3	4.25
	SCRIPTAID/HDAC10	ModWeb.2	6.77	6.23
	HDAC-11	SAHA/HDAC11	ModWeb.3	6.441
TSA/HDAC11		ModWeb.1	7.824	5.64

calculations were performed in a box with dimensions of 70 Å × 128 Å × 74 Å. This procedure allowed the decomposition of the enzymes/inhibitor interactions energies into three main contributions (fields) as follows: steric, electrostatic, and hydrophobic. The default parameters for Zn in AutoGrid were used and no attempts to include intramolecular terms were done.

Statistical Analysis. All statistical calculations were performed with R, a free software environment for statistical computing and graphics. For the final training set, seven different combinations of the fields previously calculated were tried: the single fields (STE, ELE, and DRY) and the multifield ELE+STE, ELE+DRY, STE+DRY, and ELE+STE+DRY.

PLS. All calculations were conducted using the PLS and cross-validation features of the PLS package described by Mevik.⁴⁵ An in-house R script was compiled to import the data and carry out all calculations.

BUW. Furthermore, in the case of multiple probes, a scaling procedure, called Block Unscaled Weights (BUW), was applied as data pretreatment. This procedure enforces the same importance to each interaction type within the model, normalizing the energy distribution of the X-variables as described by Kastenholz et al.⁴⁶ BUW coefficients are reported in Supporting Information File 1 Table SI-4.

Molecular Docking. AutoDock Settings. The AutoDock-Tools package was employed to generate the docking input files and to analyze the docking results. A grid box size of 57 × 44 × 53 with a spacing of 0.375 Å between the grid points was implemented. A total of 100 runs were generated by using the genetic algorithm, while the remaining run parameters were maintained at their default setting. A cluster analysis was carried out using 2 Å as the RMSD tolerance.

AutoDockVina Settings. The same AutoDock grid box was used for its calculations. The docking simulations were carried out with an energy range of 10 kcal/mol and exhaustiveness of 100. The output comprised 20 different conformations for every receptor considered. Although Vina does not include any clustering of the output poses, the clustering feature of the AutoDock program was used to inspect the conformation families using a clustering tolerance set at 2 Å.

■ ASSOCIATED CONTENT

📄 Supporting Information

Figures SI-1–9, Tables SI-1–4, and supporting files. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rino.ragno@uniroma1.it.

Notes

The authors declare no competing financial interest.

[§]Visiting Professor from the Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, United States.

■ ACKNOWLEDGMENTS

G.R.M. thanks the Facoltà di Farmacia e Medicina at Sapienza University for supporting his visiting professorship during the fall of 2011 as well as their outstanding hospitality. This study was supported by grants from Italian Ministry of University and Research (MIUR Grant 2008ZTN724_003 and FIRB Grant RBF10ZJQT).

REFERENCES

- (1) Mai, A.; Massa, S.; Rotili, D.; Cerbara, I.; Valente, S.; Pezzi, R.; Simeonia, S.; Ragno, R. Histone Deacetylation in Epigenetics: An Attractive Target for Anticancer Therapy. *Med. Res. Rev.* **2005**, *25* (3), 261–309.
- (2) Hu, E.; Dul, E.; Sung, C. M.; Chen, Z.; Kirkpatrick, R.; Zhang, G. F.; Johanson, K.; Liu, R.; Lago, A.; Hofmann, G.; Macarron, R.; de los Frailes, M.; Perez, P.; Krawiec, J.; Winkler, J.; Jaye, M. Identification of novel isoform-selective inhibitors within class I histone deacetylases. *J. Pharmacol. Exp. Ther.* **2003**, *307*, 720–728.
- (3) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **2009**, *325* (5942), 834–40.
- (4) Zain, J.; Kaminetzky, D.; O'Connor, O. A. Emerging role of epigenetic therapies in cutaneous T-cell lymphomas. *Expert. Rev. Hematol.* **2010**, *3* (2), 187–203.
- (5) (a) Savarino, A.; Mai, A.; Norelli, S.; El Daker, S.; Valente, S.; Rotili, D.; Altucci, L.; Palamara, A. T.; Garaci, E. "Shock and kill" effects of class I-selective histone deacetylase inhibitors in combination with the glutathione synthesis inhibitor buthionine sulfoximine in cell line models for HIV-1 quiescence. *Retrovirology* **2009**, *6*, 52. (b) Choudhary, S. K.; Margolis, D. M. Curing HIV: Pharmacologic Approaches to Target HIV-1 Latency. *Ann. Rev. Pharmacol. Toxicol.* **2011**, *51* (1), 397–418. (c) Matalon, S.; Rasmussen, T. A.; Dinarello, C. A. Histone deacetylase inhibitors for purging HIV-1 from the latent reservoir. *Mol. Med.* **2011**, *17* (5–6), 466–72.
- (6) (a) Ortiz, A. R.; Pastor, M.; Palomer, A.; Cruciani, G.; Gago, F.; Wade, R. C. Reliability of comparative molecular field analysis models: effects of data scaling and variable selection using a set of human synovial fluid phospholipase A2 inhibitors. *J. Med. Chem.* **1997**, *40* (7), 1136–1148. (b) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem.* **1995**, *38*, 2681–2691. (c) Perez, C.; Pastor, M.; Ortiz, A. R.; Gago, F. Comparative Binding Energy Analysis of HIV-1 Protease Inhibitors: Incorporation of Solvent Effects and Validation as a Powerful Tool in Receptor-Based Drug Design. *J. Med. Chem.* **1998**, *41* (6), 836–852. (d) Lozano, J. J.; Pastor, M.; Cruciani, G.; Gaedt, K.; Centeno, N. B.; Gago, F.; Sanz, F. 3D-QSAR methods on the basis of ligand-receptor complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 ligands. *J. Comput.-Aided Mol. Des.* **2000**, *14* (4), 341–353.
- (7) Ballante, F.; Musmuca, L.; Marshall, G. R.; Ragno, R. Comprehensive Models of Wild-Type and Mutant HIV-1 Reverse Transcriptases. *J. Comput.-Aided Mol. Design* **2012**, in press.
- (8) Team, R. D. C. *The R Foundation for Statistical Computing*. <http://www.r-project.org/>.
- (9) Gil-Redondo, R.; Klett, J.; Gago, F.; Morreale, A. gCOMBINE: A graphical user interface to perform structure-based comparative binding energy (COMBINE) analysis on a set of ligand-receptor complexes. *Proteins* **2010**, *78* (1), 162–72.
- (10) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112* (3), 535–42.
- (11) Bressi, J. C.; Jennings, A. J.; Skene, R.; Wu, Y.; Melkus, R.; De Jong, R.; O'Connell, S.; Grimshaw, C. E.; Navre, M.; Gangloff, A. R. Exploration of the HDAC2 foot pocket: Synthesis and SAR of substituted N-(2-aminophenyl)benzamides. *Bioorg. Med. Chem. Lett.* **2010**, *20* (10), 3142–3145.
- (12) Dowling, D. P.; Gantt, S. L.; Gattis, S. G.; Fierke, C. A.; Christianson, D. W. Structural studies of human histone deacetylase 8 and its site-specific variants complexed with substrate and inhibitors. *Biochemistry* **2008**, *47* (51), 13554–63.
- (13) Blackwell, L.; Norris, J.; Suto, C. M.; Janzen, W. P. The use of diversity profiling to characterize chemical modulators of the histone deacetylases. *Life Sci.* **2008**, *82* (21–22), 1050–1058.
- (14) Somoza, J. R.; Skene, R. J.; Katz, B. A.; Mol, C.; Ho, J. D.; Jennings, A. J.; Luong, C.; Arvai, A.; Buggy, J. J.; Chi, E.; Tang, J.; Sang, B. C.; Verner, E.; Wynands, R.; Leahy, E. M.; Dougan, D. R.; Snell, G.; Navre, M.; Knuth, M. W.; Swanson, R. V.; McRee, D. E.; Tari, L. W. Structural snapshots of human HDAC8 provide insights into the class I histone deacetylases. *Structure* **2004**, *12* (7), 1325–34.
- (15) Ortore, G.; Di Colo, F.; Martinelli, A. Docking of hydroxamic acids into HDAC1 and HDAC8: a rationalization of activity trends and selectivities. *J. Chem. Inf. Model.* **2009**, *49* (12), 2152–2161.
- (16) Vannini, A.; Volpari, C.; Filocamo, G.; Casavola, E. C.; Brunetti, M.; Renzoni, D.; Chakravarty, P.; Paolini, C.; De Francesco, R.; Gallinari, P.; Steinkuhler, C.; Di Marco, S. Crystal structure of a eukaryotic zinc-dependent histone deacetylase, human HDAC8, complexed with a hydroxamic acid inhibitor. *Proc. Natl. Acad. Sci. USA* **2004**, *101* (42), 15064–9.
- (17) Bottomley, M. J.; Lo Surdo, P.; Di Giovine, P.; Cirillo, A.; Scarpelli, R.; Ferrigno, F.; Jones, P.; Neddermann, P.; De Francesco, R.; Steinkuhler, C.; Gallinari, P.; Carfi, A. Structural and functional analysis of the human HDAC4 catalytic domain reveals a regulatory structural zinc-binding domain. *J. Biol. Chem.* **2008**, *283* (39), 26694–704.
- (18) Schuetz, A.; Min, J.; Allali-Hassani, A.; Schapira, M.; Shuen, M.; Loppnau, P.; Mazitschek, R.; Kwiatkowski, N. P.; Lewis, T. A.; Maglathlin, R. L.; McLean, T. H.; Bochkarev, A.; Plotnikov, A. N.; Vedadi, M.; Arrowsmith, C. H. Human HDAC7 harbors a class IIa histone deacetylase-specific zinc binding motif and cryptic deacetylase activity. *J. Biol. Chem.* **2008**, *283* (17), 11355–63.
- (19) Fass, D. M.; Shah, R.; Ghosh, B.; Hennig, K.; Norton, S.; Zhao, W. N.; Reis, S. A.; Klein, P. S.; Mazitschek, R.; Maglathlin, R. L.; Lewis, T. A.; Haggarty, S. J. Effect of Inhibiting Histone Deacetylase with Short-Chain Carboxylic Acids and Their Hydroxamic Acid Analogs on Vertebrate Development and Neuronal Chromatin. *ACS Med. Chem. Lett.* **2011**, *2* (1), 39–42.
- (20) Hanesian, S.; Auzzas, L.; Larsson, A.; Zhang, J.; Giannini, G.; Gallo, G.; Ciacci, A.; Cabri, W. Vorinostat-Like Molecules as Structural, Stereochemical, and Pharmacological Tools. *ACS Med. Chem. Lett.* **2010**, *1* (2), 70–74.
- (21) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–88.
- (22) Fiser, A.; Sali, A. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* **2003**, *374*, 461–491.
- (23) Ballante, F.; Ragno, R. 3-D QSAutogrid/R: an alternative procedure to build 3-D QSAR models. Methodologies and applications. *J. Chem. Inf. Model.* **2012**, *52*, 1674–1685.
- (24) Trotter, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461.
- (25) Beckers, T.; Burkhardt, C.; Wieland, H.; Gimmich, P.; Ciossek, T.; Maier, T.; Sanders, K. Distinct pharmacological properties of second generation HDAC inhibitors with the benzamide or hydroxamate head group. *Int. J. Cancer* **2007**, *121* (5), 1138–1148.
- (26) Zhou, N.; Moradei, O.; Raeppl, S.; Leit, S.; Frechette, S.; Gaudette, F.; Paquin, L.; Bernstein, N.; Bouchain, G.; Vaisburg, A.; Jin, Z.; Gillespie, J.; Wang, J.; Fournel, M.; Yan, P. T.; Trachy-Bourget, M. C.; Kalita, A.; Lu, A.; Rahil, J.; MacLeod, A. R.; Li, Z.; Besterman, J. M.; Delorme, D. Discovery of N-(2-aminophenyl)-4-[(4-pyridin-3-ylpyrimidin-2-ylamino)methyl]benzamide (MGCD0103), an orally active histone deacetylase inhibitor. *J. Med. Chem.* **2008**, *51* (14), 4072–4075.
- (27) Botta, C. B.; Cabri, W.; Cini, E.; De Cesare, L.; Fattorusso, C.; Giannini, G.; Persico, M.; Petrella, A.; Rondinelli, F.; Rodriguez, M.; Russo, A.; Taddei, M. Oxime Amides as a Novel Zinc Binding Group in Histone Deacetylase Inhibitors: Synthesis, Biological Activity, and Computational Evaluation. *J. Med. Chem.* **2011**, *54* (7), 2165–2182.
- (28) Fass, D. M.; Shah, R.; Ghosh, B.; Hennig, K.; Norton, S.; Zhao, W. N.; Reis, S. A.; Klein, P. S.; Mazitschek, R.; Maglathlin, R. L.; Lewis, T. A.; Haggarty, S. J. Effect of Inhibiting Histone Deacetylase with Short-Chain Carboxylic Acids and Their Hydroxamic Acid Analogs on Vertebrate Development and Neuronal Chromatin. *ACS Med. Chem. Lett.* **2010**, *2* (1), 39–42.

- (29) (a) Bottomley, M. J.; Lo Surdo, P.; Di Giovine, P.; Cirillo, A.; Scarpelli, R.; Ferrigno, F.; Jones, P.; Neddermann, P.; De Francesco, R.; Steinkuhler, C.; Gallinari, P.; Carfi, A. Structural and functional analysis of the human HDAC4 catalytic domain reveals a regulatory structural zinc-binding domain. *J. Biol. Chem.* **2008**, *283* (39), 26694–704. (b) Jones, P.; Bottomley, M. J.; Carfi, A.; Cecchetti, O.; Ferrigno, F.; Lo Surdo, P.; Ontoria, J. M.; Rowley, M.; Scarpelli, R.; Schultz-Fademrecht, C.; Steinkuhler, C. 2-Trifluoroacetylthiophenes, a novel series of potent and selective class II histone deacetylase inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18* (11), 3456–3461.
- (30) (a) Kozikowski, A. P.; Chen, Y.; Gaysin, A. M.; Savoy, D. N.; Billadeau, D. D.; Kim, K. H. Chemistry, biology, and QSAR studies of substituted biaryl hydroxamates and mercaptoacetamides as HDAC inhibitors-nanomolar-potency inhibitors of pancreatic cancer cell growth. *ChemMedChem* **2008**, *3* (3), 487–501. (b) Kozikowski, A. P.; Tapadar, S.; Luchini, D. N.; Kim, K. H.; Billadeau, D. D. Use of the nitrile oxide cycloaddition (NOC) reaction for molecular probe generation: a new class of enzyme selective histone deacetylase inhibitors (HDACIs) showing picomolar activity at HDAC6. *J. Med. Chem.* **2008**, *51* (15), 4370–4373.
- (31) Cole, K. E.; Dowling, D. P.; Boone, M. A.; Phillips, A. J.; Christianson, D. W. Structural basis of the antiproliferative activity of largazole, a depsipeptide inhibitor of the histone deacetylases. *J. Am. Chem. Soc.* **2011**, *133* (32), 12474–12477.
- (32) Whitehead, L.; Dobler, M. R.; Radetich, B.; Zhu, Y.; Atadja, P. W.; Claiborne, T.; Grob, J. E.; McRiner, A.; Pancost, M. R.; Patnaik, A.; Shao, W.; Shultz, M.; Tichkule, R.; Tommasi, R. A.; Vash, B.; Wang, P.; Stams, T. Human HDAC isoform selectivity achieved via exploitation of the acetate release channel with structurally unique small molecule inhibitors. *Bioorg. Med. Chem.* **2011**, *19* (15), 4626–4634.
- (33) Finnin, M. S.; Donigian, J. R.; Cohen, A.; Richon, V. M.; Rifkind, R. A.; Marks, P. A.; Breslow, R.; Pavletich, N. P. Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* **1999**, *401* (6749), 188–93.
- (34) Nielsen, T. K.; Hildmann, C.; Riestler, D.; Wegener, D.; Schwienhorst, A.; Ficner, R. Complex structure of a bacterial class 2 histone deacetylase homologue with a trifluoromethylketone inhibitor. *Acta Crystallogr. Sect. F* **2007**, *63* (Pt 4), 270–3.
- (35) Nielsen, T. K.; Hildmann, C.; Dickmanns, A.; Schwienhorst, A.; Ficner, R. Crystal structure of a bacterial class 2 histone deacetylase homologue. *J. Mol. Biol.* **2005**, *354* (1), 107–120.
- (36) (a) Ragno, R.; Simeoni, S.; Rotili, D.; Caroli, A.; Botta, G.; Brosch, G.; Massa, S.; Mai, A. Class II-selective histone deacetylase inhibitors. Part 2: alignment-independent GRIND 3-D QSAR, homology and docking studies. *Eur. J. Med. Chem.* **2008**, *43* (3), 621–32. (b) Ragno, R.; Simeoni, S.; Valente, S.; Massa, S.; Mai, A. 3-D QSAR studies on histone deacetylase inhibitors. A GOLPE/GRID approach on different series of compounds. *J. Chem. Inf. Model.* **2006**, *46* (3), 1420–30.
- (37) (a) Mai, A.; Massa, S.; Ragno, R.; Cerbara, I.; Jesacher, F.; Loidl, P.; Brosch, G. 3-(4-Aroyl-1-methyl-1H-2-pyrrolyl)-N-hydroxy-2-alkylamides as a new class of synthetic histone deacetylase inhibitors. I. Design, synthesis, biological evaluation, and binding mode studies performed through three different docking procedures. *J. Med. Chem.* **2003**, *46* (4), 512–24. (b) Mai, A.; Massa, S.; Ragno, R.; Esposito, M.; Sbardella, G.; Nocca, G.; Scatena, R.; Jesacher, F.; Loidl, P.; Brosch, G. Binding mode analysis of 3-(4-benzoyl-1-methyl-1H-2-pyrrolyl)-N-hydroxy-2-propenamide: a new synthetic histone deacetylase inhibitor inducing histone hyperacetylation, growth inhibition, and terminal cell differentiation. *J. Med. Chem.* **2002**, *45* (9), 1778–84.
- (38) Henrich, S.; Feierberg, I.; Wang, T.; Blomberg, N.; Wade, R. C. Comparative binding energy analysis for binding affinity and target selectivity prediction. *Proteins* **2010**, *78* (1), 135–153.
- (39) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). I. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110* (18), 5959–5967.
- (40) Musmuca, I.; Caroli, A.; Mai, A.; Kaushik-Basu, N.; Arora, P.; Ragno, R. Combining 3-D Quantitative Structure-Activity Relationship with Ligand Based and Structure Based Alignment Procedures for in Silico Screening of New Hepatitis C Virus NSSB Polymerase Inhibitors. *J. Chem. Inf. Model.* **2010**, *50*, 662–676.
- (41) Nielsen, M.; Lundegaard, C.; Lund, O.; Petersen, T. N. CPHmodels-3.0-remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res.* **2010**, *38* (Web Server issue), W576–581.
- (42) Fernandez-Fuentes, N.; Madrid-Aliste, C. J.; Rai, B. K.; Fajardo, J. E.; Fiser, A. M4T: a comparative protein structure modeling server. *Nucleic Acids Res.* **2007**, *35* (Web Server issue), W363–368.
- (43) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **2006**, *22* (2), 195–201.
- (44) Eswar, N.; John, B.; Mirkovic, N.; Fiser, A.; Ilyin, V. A.; Pieper, U.; Stuart, A. C.; Marti-Renom, M. A.; Madhusudhan, M. S.; Yerkovich, B.; Sali, A. Tools for comparative protein structure modeling and analysis. *Nucleic Acids Res.* **2003**, *31* (13), 3375–3380.
- (45) Mevik, B.-H.; Wehrens, R. The pls Package: Principal Component and Partial Least Squares Regression in R. *J. Stat. Software* **2007**, *18* (2), 1–24.
- (46) Kastenholtz, M. A.; Pastor, M.; Cruciani, G.; Haakma, E. E.; Fox, T. GRID/CPCA: a new computational tool to design selective ligands. *J. Med. Chem.* **2000**, *43* (16), 3033–3044.



Original article

Design, synthesis and biological evaluation of new classes of thieno [3,2-*d*]pyrimidinone and thieno[1,2,3]triazine as inhibitor of vascular endothelial growth factor receptor-2 (VEGFR-2)



Enrico Perspicace^a, Valérie Jouan-Hureauux^d, Rino Ragno^{b,*}, Flavio Ballante^b, Stefania Sartini^c, Concettina La Motta^c, Federico Da Settimo^c, Binbin Chen^a, Gilbert Kirsch^a, Serge Schneider^e, Béatrice Faivre^d, Stéphanie Hesse^{a,**}

^aLaboratoire d'Ingénierie Moléculaire et Biochimie Pharmacologique, UMR CNRS 7565 SRSMC, Institut Jean Barriol, FR CNRS 2843, Université de Lorraine, 1 Boulevard Arago, 57070 Metz, France

^bRome Center for Molecular Design, Dipartimento di Chimica e Tecnologia del Farmaco, Università degli Studi di Roma "La Sapienza", P. le A. Moro 5, 00185 Roma, Italy

^cDipartimento di Farmacia, Università di Pisa, Via Bonanno 6, 56126 Pisa, Italy

^dEA 4421 SiGRETO, Université de Lorraine, Faculté de pharmacie, 5-7 Rue Albert Lebrun, BP80403, 54001 Nancy cedex, France

^eLaboratoire National de Santé, Division de Toxicologie, Université de Luxembourg, 162a Avenue de la Faïencerie, L-1511 Luxembourg, Luxembourg

ARTICLE INFO

Article history:

Received 13 December 2012

Received in revised form

7 March 2013

Accepted 10 March 2013

Available online 19 March 2013

Keywords:

Vascular endothelial growth factor receptor-2 (VEGFR-2)

Anti-angiogenic activity

Structure-based drug design (SBDD)

Ligand-based drug design (LBDD)

3-D QSAR

Thieno[3,2-*d*]pyrimidinone

Thieno[1,2,3]triazine

Endothelial cell tube formation

ABSTRACT

Driven by a multidisciplinary approach combination (Structure-Based (SB) Three-Dimensional Quantitative Structure–Activity Relationships (3-D QSAR), molecular modeling, organic chemistry and various biological evaluations) here is reported the disclosure of new thienopyrimidines **1–3** as inhibitors of KDR activity and human umbilical vein endothelial cell (HUVEC) proliferation. More specifically, compound **2f** represents a new lead compound that inhibits VEGFR-2 and HUVEC at μM concentration. Moreover by the mean of an endothelial cell tube formation *in vitro* model **2f** tartaric acid salt proved to block angiogenesis of HUVEC at μM level.

© 2013 Elsevier Masson SAS. All rights reserved.

1. Introduction

Angiogenesis, the process of new blood vessels formation, creating new capillaries from existing vasculature, is a normal process for organ development. It occurs especially during embryogenesis, in development and homeostasis of adult tissues, in wound repair and in the menstrual cycle [1,2]. It's an essential and highly regulated process under physiological conditions. When a

malfunction in controlling mechanisms of angiogenesis occurs, it may be involved in the development and progression of various diseases such as rheumatoid arthritis [3], inflammation [4], ocular neovascularization [5], psoriasis [6], tumor growth [7] and metastasis [8]. More than twenty different factors are involved in this process, among which the vascular endothelial growth factors (VEGFs) [9–11]. The VEGF family includes VEGF-A (usually named VEGF), VEGF-B, VEGF-C, VEGF-D, VEGF-E, VEGF-F and a structurally related molecule, Placental Growth Factor (PlGF). Three high-affinity VEGF tyrosine kinase receptors have been identified: VEGFR-1 (Flt-1), VEGFR-2 (KDR) and VEGFR-3 (Flt-4). VEGF and its receptors represent one of the best-validated signaling pathways in angiogenesis [12]. Furthermore, disruption of VEGFR-2 signaling has resulted in inhibition of angiogenesis and without new blood vessels

* Corresponding author. Tel.: +39 6 4991 3937; fax: +39 6 4991 3627.

** Corresponding author. Tel.: +33 3 8754 7197; fax: +33 3 8731 5801.

E-mail addresses: rino.ragno@uniroma1.it (R. Ragno), stephanie.hesse@univ-lorraine.fr (S. Hesse).

to supply oxygen, nutrients and catabolic products, tumor cells could not proliferate and thus are likely to remain dormant [13,14]. Therefore, VEGFR-2 has been the principal target of anti-angiogenic therapies [15–18], although additional studies have underlined the importance of signaling through VEGFR-1 [19].

In recent years, many molecules have been described on their ability to inhibit angiogenesis. For example, the recombinant humanized monoclonal antibody bevacizumab (Avastin[®], Genentech) recognizes and blocks VEGF [20], and was the first anti-angiogenic agent to be approved by Food and Drug Administration (FDA). Other inhibitory molecules have been developed targeting the vascular endothelial growth factor receptors (Figs. 1 and 2). Sorafenib (BAY 43-9006 or Nexavar[®]) is a derivative initially selected as an inhibitor of Raf kinase by targeting MAPK pathway but this compound has also a powerful inhibitory action on VEGFR-1, VEGFR-2, and Platelet Derived Growth Factor Receptor- β (PDGFR- β) [21]. Sunitinib (SU 011248 or Sutent[®]) is an oral multi-targeted receptor tyrosine kinase (RTK) which inhibits VEGFR-1, VEGFR-2, PDGFR, KIT, Flt3 and the receptor encoded by the RET proto-oncogene [22]. Today, targeting tumor angiogenesis has become part of daily care for many solid tumors. Indeed, survival gains were achieved in the case of metastatic renal cell carcinoma and hepatocarcinoma for which medical treatment had little or no effect. However, numerous side effects have been reported (skin, cardiovascular) [23,24] to which the clinician had not been confronted with the molecules of conventional chemotherapy. Furthermore, studies have shown the emergence of resistance mechanisms to anti-angiogenic products [25]. Indeed, in some cases, despite initial sensitivity and a dramatic reduction in volume, treated tumors resumed their growth and invasiveness. Within this scenario it is obvious the continuous need to develop new anti-angiogenic agents that can both reduce the side effects and block resistance of tumor cells. In this study we disclose the new thienopyrimidines 1–3 (Fig. 3) as inhibitors of VEGFR-2 designed by a Structure-Based (SB) Three-Dimensional Quantitative Structure–Activity Relationships (3-D QSAR), molecular modeling and biological combined approach.

2. Results and discussion

2.1. Rationale, preliminary screening and design

Quinazoline [27,28] and naphtamide [29] derivatives have attracted great interest as VEGFR-2 inhibitors over the past years.

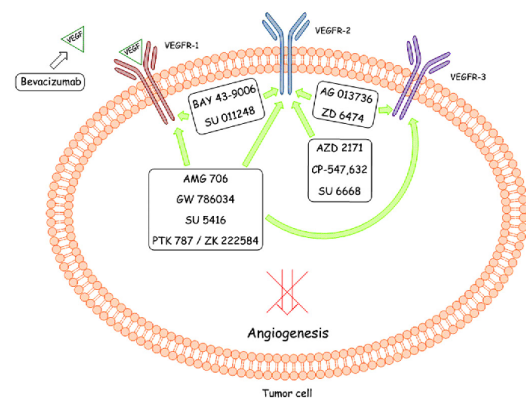


Fig. 1. Inhibition pathways of angiogenesis [26].

Furo[2,3-d]pyrimidines [30], pyridinyltriazines [31] and pyrimidinylindazoles [32] were also reported as VEGFR inhibitors. All compounds were comprehensively studied and many of them were co-crystallized in the ATP-binding site of KDR. Such amount of structural information was used by us to develop a predictive SB 3-D QSAR model [33] and herein used as tools to develop new VEGFR-2 inhibitors.

Recently, some of us were deeply interested in the synthesis of thienopyrimidine scaffolds and their functionalization via palladium-catalyzed cross-coupling or nucleophilic aromatic substitution reaction (S_NAr) [34,35]. Furthermore, we also developed the methodology to access to selenolo- and thiazolo-triazine scaffolds [36,37]. Thirty-five compounds, (4–38), out of all the available ones, were selected as potential VEGFR-2 inhibitors on the basis of their structural similarity to well-known inhibitors from the literature, and thus submitted to a VEGFR-2 inhibition assay, performed as reported in the Experimental section (Tables 1–3) [38]. IC_{50} values were investigated for compounds showing higher % of inhibition (Table 4) by linear regression analysis of the log–dose response curve, which was generated using at least five concentrations of the inhibitor causing an inhibition between 20% and 80%, with three replicates at each concentration.

At the same time, compounds 4–38 were also investigated by our SB 3-D QSAR protocol [33]. As reported in Tables 1–3, only five (17–19, 22, 23) out of the 35 assayed compounds were found totally inactive, while four derivatives (14, 30, 33 and 35) inhibited VEGFR-2 at high percentage (over 73%) and their IC_{50} were determined (Table 4). The SB 3-D QSAR protocol application over predicted compounds 4–38 in the range of 5.82–7.92 pIC_{50} , with an average value of 6.56 (Supplementary material Table SM-1). Besides, the over prediction, the fact that 85% of compound displayed indeed some sort of inhibition activity toward VEGFR-2 (Tables 1–3) proved the calculated data to be in agreement with the experimental. Furthermore, due to the lack of Se parameters, for both dockings and 3-D QSAR calculations the Se containing compounds were modeled as sulfur derivatives. Therefore, some errors of predictions were expected for Se containing derivatives 22–38.

2-D structures, 3-D QSAR predicted pIC_{50} s and docked conformations of compounds 4–38 are reported in Supplementary material Table SM-1. At a first glance, all the compounds bind preferentially in the very first part of the binding site (left sides of figures in Supplementary material Table SM-1) overlapping the thienopyrimidine (4–13), thiazolotriazine (14–21) and selenolotriazine (22–38) scaffolds mainly with the reference structure (PDB ID 2qu5) central benzimidazole. Regarding thienopyrimidines 4–13, the increasing sterical hindrance of the third fused cycle seems to increase the molecules/VEGFR-2 interactions (compare 4, 6, 8 and 11 binding modes with those of 5, 7, 9 and 12 in Supplementary material Table SM-1) in agreement with the lower activities of bicycles 4, 6, 8 and 11 than those of the 5, 7, 9 and 12 (Table 1). In an opposite way, the increase of steric hindrance is detrimental for the thiazolotriazines 14–21 where the smaller derivative 14 shows the higher activity. Actually, within this series, the insertion of different substituents clearly influence both binding modes and activities of the resulting compounds, although no particular pattern can be drawn out from them. Concerning the selenolotriazines 22–38 (modeled as thienotriazines) their binding modes resemble those of thienopyrimidines 4–13. In a whole, this first series of compounds (4–38) was considered as a sort of fragment library for which both the experimental and prediction were used to select the more interesting scaffolds.

A deeper binding mode analysis was then performed on the most active compounds 14, 30, 33 and 35 and their conformations as docked by AutoDock revealed that all the four compounds bind roughly in the same region defined by Val27, Ala45, Lys47, Val93,

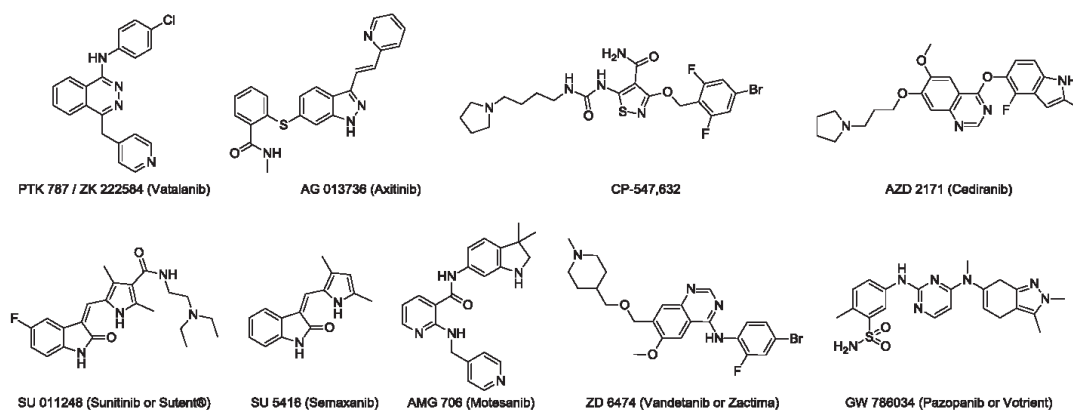


Fig. 2. Examples of anti-angiogenic compounds.

Val95, Leu164 and Phe176 (Fig. 4 and Supplementary material Figure SM-1). While **14**, **30** and **33** shared a common binding mode with the thiazolotriazine/selenolotriazine groups overlapping with the same orientation, on the contrary the least active compound **35** was predicted to bind with the selenolotriazine group rotated of about 180°, likely due to the presence of the morpholino substituent avoiding to maintain the same orientation (Supplementary material Figure SM-1D). The 3-D QSAR models predicted these compounds in the range 0.05–1 μ M (Supplementary material Table SM-2), nevertheless these over-predictions are in agreement with the known limitation of the models in which the training set compounds with activities below higher than 1 μ M were not correctly fitted being over-recalculated by the models themselves (Supplementary material Figure SM-2). Overlapping the 3-D QSAR maps to the most active compound **30** indicated that the introduction of further sterical groups into the ATP binding site would have led to more active compounds. Accordingly, a new series of thienopyrimidines/thienotriazines, **1–3** (Fig. 3), were designed and synthesized, in order to better fill the ATP binding site.

The new designed compounds were promptly modeled and subjected to both molecular docking simulation and 3-D QSAR predictions. Most of the new designed molecules were predicted to be in the 0.01–0.1 μ M range (Supplementary material Table SM-3)

and thus, considering the error prediction observed for the above **14**, **30**, **33** and **35** compounds, we expected these molecules to show biological activities in the range on low micromolar or at least submicromolar. Among the analyzed compounds, **2f** was predicted to be in the low nanomolar range. Binding mode exploration of **2f** revealed that indeed the compound occupies the ATP binding site slightly more than **14**, **30**, **33** and **35**, particularly it seems to block more efficiently the ATP pocket (Fig. 5). Furthermore, in Figure SM-3 (Supplementary material) is clearly visible that the thienopyrimidine core of **2f** binds in a reversed way the same region of **14**, **30** and **33** fused-triazine groups and overlaps the **35**-tolyl substituent. In this way, the benzylindole fragment of **2f** fills-up the ATP binding entrance establishing further sterical interactions that could stabilize the **2f**/VEGFR-2 complex (Supplementary material Figure SM-4). In view of this promising scenario, we promptly designed the synthesis of compounds **1–3** to test their biological activities as potential VEGFR-2 inhibitors.

2.2. Chemistry

The synthesis of compounds **1–3** was designed to be achieved by the key intermediate 5-(indol-3-yl)-3-aminothiophene-2-carboxylate or derivatives. Therefore we have developed a multi-step sequence to lead to 5-(indol-3-yl)-3-aminothiophenes **43a–f**

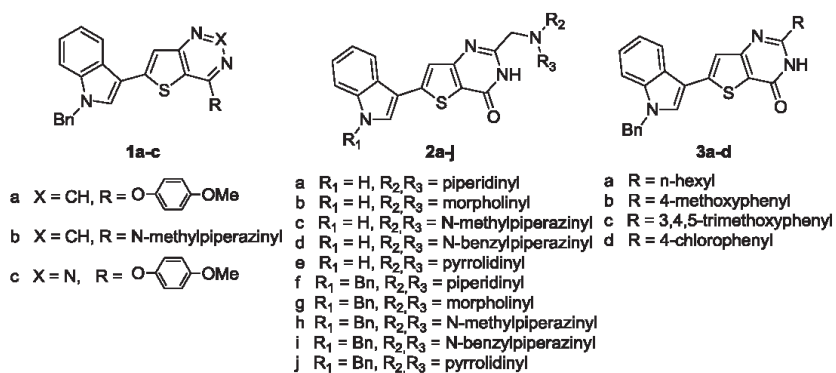
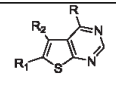
Fig. 3. Designed compounds **1–3**.

Table 1
Structures and VEGFR-2 inhibitory activities of thieno[3,2-d]pyrimidines 4–13.



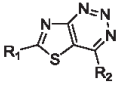
#	R	R ₁	R ₂	% Inhibition @ 200 μM ^a
4	4-MeO-C ₆ H ₄	H	H	25.4 ± 2.0
5	4-MeO-C ₆ H ₄	(CH ₂) ₄	H	47.4 ± 1.8
6	3-Ac-C ₆ H ₄	H	H	13.1 ± 0.6
7	3-Ac-C ₆ H ₄	(CH ₂) ₄	H	34.8 ± 2.9
8	3-Indolyl	H	H	23.3 ± 1.6
9	3-Indolyl	(CH ₂) ₄	H	44.9 ± 3.9
10	4-MeO-C ₆ H ₄ O	H	H	64.2 ± 3.8
11	3-Ac-C ₆ H ₄ O	H	H	21.9 ± 1.7
12	3-Ac-C ₆ H ₄ O	(CH ₂) ₄	H	56.9 ± 4.5
13	4-NH ₂ -C ₆ H ₄ O	H	H	40.1 ± 2.4

^a Values are means ± SEM of two determinations carried out in triplicate.

(Scheme 1). 5-Aryl-3-aminothiophenes are usually synthesized in our lab starting from acetophenones via β-aryl-β-chloroacroleins, oximes, β-aryl-β-chloroacrylonitriles, condensation with sodium sulfide and activated halogen derivatives and finally by a cyclization step to form the thiophene moiety [39]. However, working on 3-acetylindoles needed to modify the reaction conditions for nearly each step of the synthesis. In particular, 3-acetylindoles **39** were first reacted with Vilsmeier–Haack reagent (formed in situ with phosphorus oxychloride and *N,N*-dimethylformamide) to give β-chloroacroleins. The intermediates iminium salts needed to be trapped and isolated as perchlorates **40**. Iminiums salts **40** were then converted to the corresponding oxime **41** [40]. Use of classical conditions for dehydration in β-aryl-β-chloroacrylonitriles (i.e. Ac₂O, PPA, P₂O₅, PCl₅) failed, therefore the expected compounds **42** were obtained in good yields using di-2-pyridinyl thionocarbonate (DPT) [41]. Thiophenes **43** were obtained by treating **42** with ethyl bromoacetate, chloroacetamide or chloroacetamide in presence of sodium sulfide followed by cyclization with sodium ethanolate.

3-Amino-2-cyano-5-indol-3-ylthiophenes **43b** and **43e** were converted into thieno[3,2-d]pyrimidin-4-ones **44a** and **44b** in respectively 84% and 88% by action of formic acid in the presence of a catalytic amount of sulfuric acid (Scheme 2) [34]. Then 4-chlorothienopyrimidines **45a** and **45b** were obtained (45% and 98% yields) using phosphorus oxychloride. Finally, compounds **1a** and **1b** were obtained by S_NAr with *p*-methoxyphenolate and *N*-methylpiperazine, respectively. Intermediate **43e** allowed formation of thienotriazine **46** in 88% yield which was then submitted to S_NAr with *p*-methoxyphenolate to give **1c** in 80% yield [37] (Scheme 2).

Table 2
Structures and VEGFR-2 inhibitory activities of thiazolotriazines 14–21.

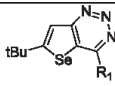


#	R ₁	R ₂	% Inhibition @ 200 μM ^a
14	S-Me	Cl	93.5 ± 7.5
15	NH-Ph	Cl	49.8 ± 2.9
16	NH-Ph	4-Morpholinyl	13.7 ± 1.1
17	1-Pyrrolidinyl	Cl	n.a. ^b
18	1-Pyrrolidinyl	1-Pyrrolidinyl	n.a. ^b
19	1-Pyrrolidinyl	4-Morpholinyl	n.a. ^b
20	4-Morpholinyl	Cl	11.9 ± 0.7
21	4-Morpholinyl	1-Pyrrolidinyl	16.2 ± 1.1

^a Values are means ± SEM of two determinations carried out in triplicate.

^b Not active. No inhibition was observed up to 200 μM of the tested compound.

Table 3
Structures and VEGFR-2 inhibitory activities of selenotriazines 22–38.



#	R ₁	R ₂	% Inhibition @ 200 μM ^a
22	1-Pyrrolidinyl	–	n.a. ^b
23	4-Morpholinyl	–	n.a. ^b
24	Cl	–	37.8 ± 3.2
25	1-Pyrrolidinyl	–	17.0 ± 1.0
26	4-Morpholinyl	–	19.4 ± 1.5
27	Cl	H	60.7 ± 4.8
28	1-Pyrrolidinyl	H	20.8 ± 1.4
29	4-Morpholinyl	H	24.7 ± 1.9
30	Cl	Cl	94.6 ± 6.6
31	1-Pyrrolidinyl	Cl	37.8 ± 1.5
32	4-Morpholinyl	Cl	31.0 ± 1.3
33	Cl	CH ₃	88.0 ± 6.1
34	1-Pyrrolidinyl	CH ₃	73.5 ± 6.6
35	4-Morpholinyl	CH ₃	72.5 ± 3.6
36	Cl	OCH ₃	36.5 ± 2.9
37	1-Pyrrolidinyl	OCH ₃	17.6 ± 1.1
38	4-Morpholinyl	OCH ₃	14.8 ± 1.2

^a Values are means ± SEM of two determinations carried out in triplicate.

^b Not active. No inhibition was observed up to 200 μM of the test compound.

3-Amino-2-carboxamido-5-indol-3-ylthiophenes **43c** and **43f** were reacted successively with chloroacetyl chloride and a secondary amine to give compounds **48a–j** [42]. Cyclization in thienopyrimidinones **2a–j** was realized in basic media in good yields (Scheme 3). Compounds **2a–j** were converted in hydrochlorides.

Compounds **3a–d** were achieved in moderate yields (12–36%) by condensation reactions between 3-amino-2-carboxamido-5-indol-3-ylthiophene **43f** and the proper aldehydes in the presence of hydrochloric acid (Scheme 4).

2.3. Biological activity of designed compounds 1–3

The VEGFR-2 inhibitory activities of the newly synthesized compounds **1–3** were evaluated at 200 μM fixed doses (Table 5) and among the sixteen tested derivatives only four (**2b**, **2e**, **2h** and **2j**) were totally inactive. These experimental results are in agreement with the above reported predictions and interestingly the most active predicted compound **2f** was indeed those with the highest

Table 4
IC₅₀ values of selected compounds, **14**, **30**, **33**, **35**. The SB 3-D QSAR predicted pIC₅₀ is also reported in comparison with the experimental pIC₅₀ value.

#	Exp. IC ₅₀ (μM) ^a	Exp. pIC ₅₀	Pred. pIC ₅₀
14	45.2 ± 2.7	4.3	5.9
30	4.1 ± 0.2	5.4	6.0
33	16.7 ± 1.0	4.8	6.0
35	26.2 ± 1.5	4.6	7.1

^a IC₅₀ values represent the concentration required to produce 50% enzyme inhibition. Values ±SEM are the average from at least two independent dose–response curves.

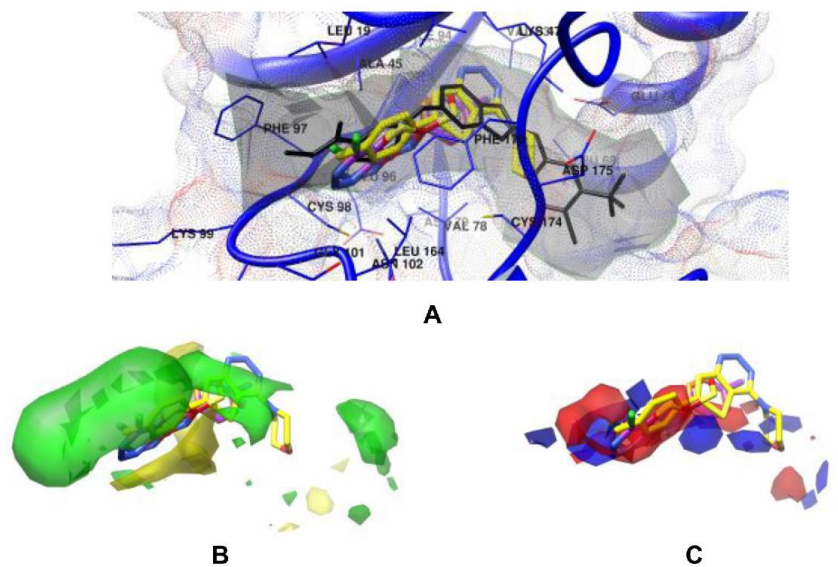


Fig. 4. Docked conformation of **14** (red), **30** (orange), **33** (purple) and **35** (yellow) (A) into VEGFR-2 (PDB ID 2qu5, blue ribbon). As reference the co-crystallized ligand is also displayed in black. The surface of ATP and co-crystallized inhibitor binding site is also shown in light gray. Merged in the steric (B) and electrostatic (C) 3-D QSAR maps. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

percentage of inhibition (99.2%). Both the activity prediction and the fixed dose evaluation were also run on the thiophene intermediates **44b**, **45b** and **48a–j**, but none of those compounds presented higher activity (prediction or experimental result) than the final thienopyrimidine series (Supplementary material Table SM-4).

IC₅₀ values of the most active compounds, **2f** and **3d**, were determined as $2.25 \pm 0.1 \mu\text{M}$ and $15.3 \pm 1.2 \mu\text{M}$, respectively.

To establish further biological activities, endothelial cells were grown in the presence of these molecules. However DMSO, solvent used to dissolve these molecules, was toxic for endothelial cells (Fig. 6). We also noted that the morphology of endothelial cell was affected by DMSO. Their shape was very round when they were incubated with a percentage of DMSO upper to 0.1% whereas they are normally elongated or fusiform. Then, to be

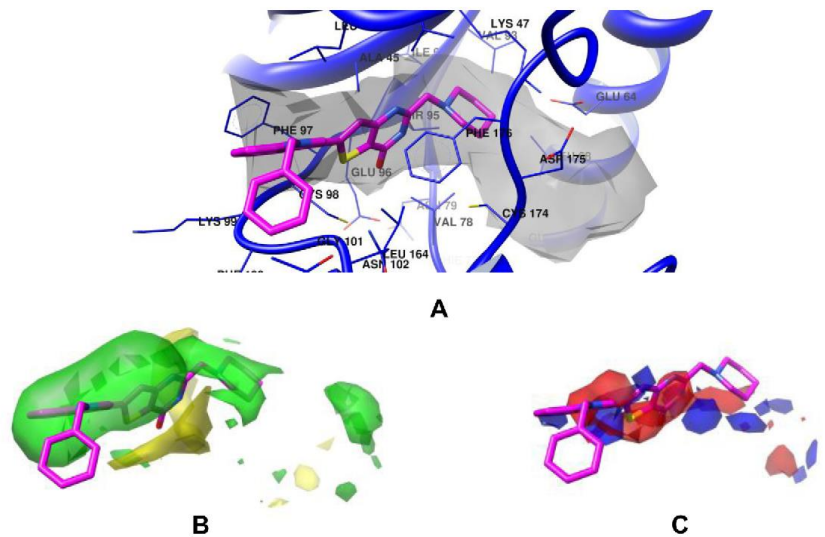
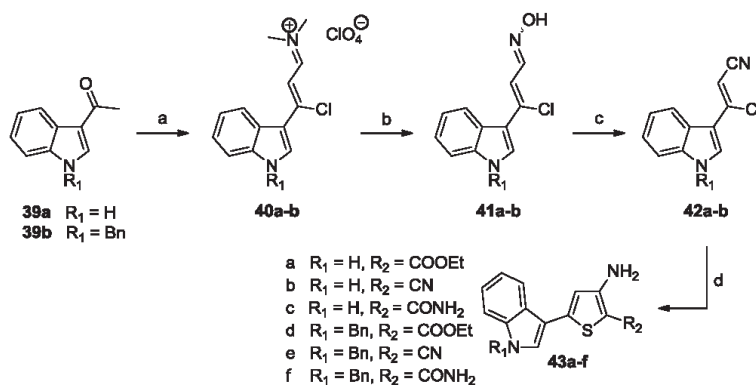


Fig. 5. Docked conformation of **2f** (magenta) (A) into VEGFR-2 (PDB ID 2qu5, blue ribbon). The ATP binding site is also shown in light gray. Merged in the steric (B) and electrostatic (C) 3-D QSAR maps. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



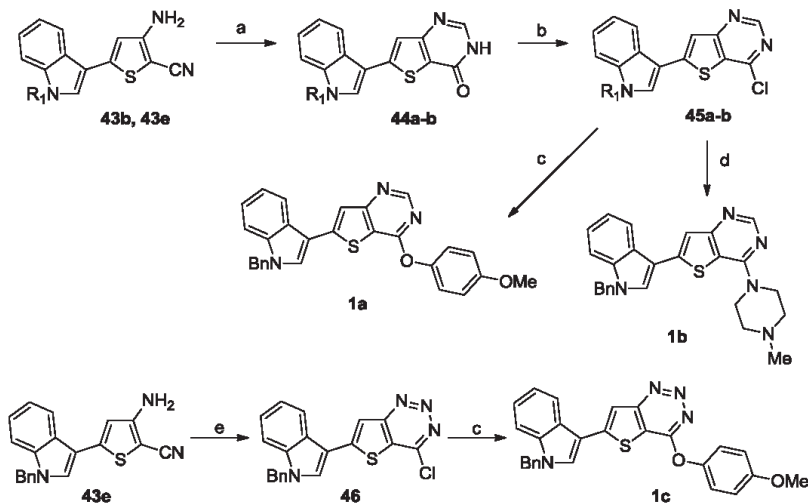
Scheme 1. Synthesis of intermediates **43a–f**. Reagents and conditions: (a) (i) Vilsmeier reagent: $POCl_3/DMF, 0^\circ C; 0^\circ C$ to RT. (ii) $NaClO_4, H_2O$. (b) $NH_2OH \cdot HCl, CH_3COONa \cdot 3H_2O, EtOH$. (c) Di-2-pyridyl thionocarbonate, DMAP, CH_2Cl_2 . (d) (i) $Na_2S_9H_2O$. (ii) $X-CH_2-R_2$. (iii) EtONa DMF, $50^\circ C$.

able to investigate the activity of those molecules, it was required to work with less than 0.1% DMSO in culture media supplemented with 2% SVF. As a matter of fact most of the newly prepared derivatives were poorly soluble in biological media containing 0.1% DMSO with precipitate formation upon dilution at physiological pH. Therefore we focused on derivative **2f** by preparing a series of salts (hydrochloride, methylsulfonate, phosphate and tartaric acid), among them only **2f** tartaric acid salt (**2f-tartaric**) showed the needed solubility in biological media to test its activity on our cellular model.

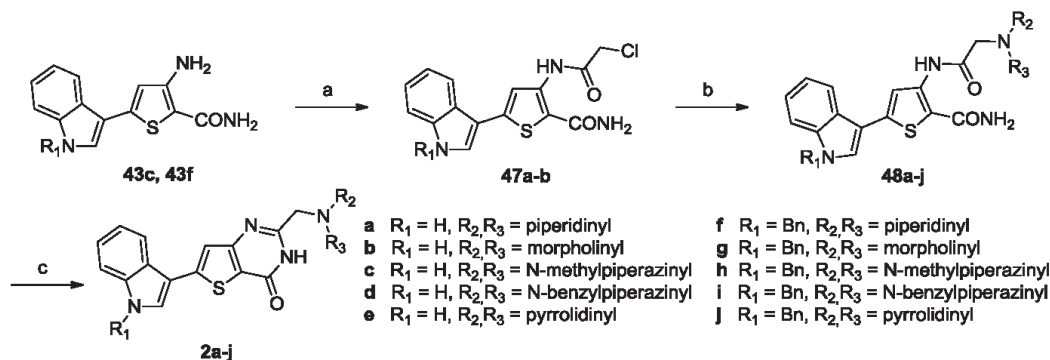
Study of the biological effect of **2f-tartaric** consists at the first step to the analysis of its effect on metabolic activity and viability of endothelial cells and at the second step to the analysis of the angiogenic activity of these cells.

Metabolic activity of HUVEC decreased with **2f-tartaric** concentration (–9% at $1.5 \mu M$ until –51% at $100 \mu M$) (Fig. 7) in a dose dependant manner (EC_{50} was estimated at about $12 \mu M$). In contrary, no variation in HUVEC viability was induced by **2f-tartaric** whatever the concentration used (Fig. 7). These results suggest that this compound had a cytostatic rather than cytotoxic activity on endothelial cells.

To further investigate the angiogenic activity of **2f-tartaric**, we used our *in vitro* model of endothelial cell tube formation in which HUVEC were plated onto matrigel[®]-coated well and cultured with or without **2f-tartaric** in the presence of VEGF during 24 h and using Sunitinib as anti-VEGFR-2 reference. As we can see in Fig. 8, VEGF alone (positive control) induced endothelial cell tube formation of HUVEC as compared with control medium without VEGF



Scheme 2. Synthesis of derivatives **1a–c**. Reagents and conditions: (a) $HCOOH, H_2SO_4, 40^\circ C$. (b) $POCl_3/DMF, 0^\circ C$ to r.t. (c) $MeOC_6H_4ONa, DMF, r.t.$ (d) *N*-Methylpiperazine, DMF, r.t. (e) $NaNO_2, HCl 37\%, 0^\circ C$ to r.t.



Scheme 3. Synthesis of derivatives 2a–j. Reagents and conditions: (a) Chloroacetyl chloride, Et₃N, THF, 0 °C to r.t. (b) R₂R₃NH, K₂CO₃, CH₃CN, reflux. (c) NaOH 2N, DMF, reflux.

(negative control). This was correlated with the decrease of area fraction found after VEGF exposure. Angiogenesis induced by VEGF was completely inhibited using 1 μM of **2f-tartaric** and 3 μM of Sunitinib (upper part of Fig. 8). **2f-tartaric** can inhibit endothelial cell tube formation induced by VEGF in a dose dependent manner showing an estimated EC₅₀ value of about 31 nM. Using the same model for Sunitinib the EC₅₀ resulted to be 645 nM. The number of contiguous cells was increased whereas the number of endothelial cell tubes was decreased and area fraction was equivalent to that of control media without VEGF (Fig. 8).

3. Conclusion

In this paper we present a multidisciplinary approach to the design, synthesis and biological characterization of thienopyrimidines as new class of VEGFR-2 inhibitors. The new compounds **1–3** were designed starting from a small fragment library (**4–38**) on which biological evaluation and SB 3-D QSAR studies were conducted in parallel. The modeling approach proved to be an effective tool to predict the activity of the preliminary series of compounds. Binding mode analysis of the most active derivatives (**14**, **30**, **33** and **35**) led to the design of a new series of thienopyrimidines (**1–3**) that were prepared by a multi-step synthetic pathway ad-hoc designed. Among the newly synthesized, compound **2f**, showed the highest activity, as predicted by the 3-D QSAR approach. Further biological assays on endothelial cell tube formation proved **2f** as a new anti-angiogenic lead compound that showed to be more efficient in inhibiting endothelial cell tube formation induced by VEGF compared with Sunitinib, in our *in vitro* model. Moreover **2f** did not cause any cytotoxic side effect to endothelial cells.

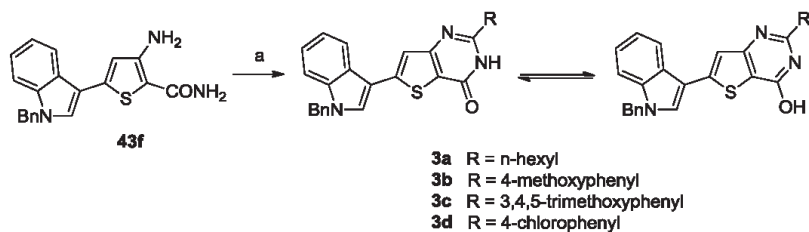
We have tested the specificity of **2f** for the inhibition of activated RTK in HUVEC using Human Phospho-RTK Proteome profiler array kit (RetD system, United Kingdom) and found that, as Sunitinib, **2f** decreased the phosphorylation of VEGFR-1 and VEGFR-2 (data not shown). Furthermore we found that **2f** decrease also the phosphorylation of others RTK activated in HUVEC such as EGFR, Axl, Dtk, ROR-2, Tie-1 and EphA6, receptors implicated in the development of cancer (Supplementary material Figure SM-5).

The binding mode of **2f** suggests that a further substitution on piperidine positions 3 and 4 would lead to better ATP binding pocket filling and, eventually, to more active derivatives (Fig. 9). To this aim a focused virtual screening on **2f** derivatives is currently ongoing and further synthetic efforts will be based on binding mode evaluation and 3-D QSAR predictions.

4. Experimental section

4.1. Chemistry

All starting materials and synthesis reagents were obtained commercially. Column chromatography was performed with silica gel 60 (particle size 70–200 μm). Thin-layer chromatography (TLC) was performed on Merck pre-coated TLC aluminum sheets with silica gel 60 F254. Melting points were determined on a Stuart SMP3 apparatus and are uncorrected. ¹H NMR spectra were measured at 250 MHz, and ¹³C NMR spectra were measured at 62.5 MHz on a Bruker Advance AC 250 spectrometer at 25 °C in CDCl₃, DMSO-*d*₆ or acetone-*d*₆, and chemical shifts are given in ppm (δ). The spectral splitting patterns are designated as follows: s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet; br s, broad singlet. Representative ¹H and ¹³C NMR spectra for **1a**, **2f**, **2g** and **2i**



Scheme 4. Synthesis of derivatives 3a–d. Reagents and conditions: (a) RCHO, MeOH, 6% HCl (w/w).

Table 5Structures and VEGFR-2 inhibitory activities of 5-indolylthienopyrimidine/triazines **1** and 5-indolylthieno[3,2-*d*]pyrimidines **2–3**.

#	X	R	R ₁	% Inhibition @ 200 μM ^a
1a	CH		Benzyl	38.8 ± 3.5
1b	CH	N-Methylpiperaziny	Benzyl	n.t. ^b
1c	N		Benzyl	1.7 ± 0.1
2a	–	1-Piperidinylmethyl	H	5.7 ± 0.4
2b	–	1-Morpholinylmethyl	H	n.a. ^c
2c	–	1-(4-Methylpiperazinyl)methyl	H	3.9 ± 0.2
2d	–	1-(4-Benzylpiperazinyl)methyl	H	2.1 ± 0.1
2e	–	1-Pyrrolidinylmethyl	H	n.a. ^c
2f	–	1-Piperidinylmethyl	Benzyl	99.2 ± 3.9
2g	–	1-Morpholinylmethyl	Benzyl	52.6 ± 3.7
2h	–	1-(4-Methylpiperazinyl)methyl	Benzyl	n.a. ^c
2i	–	1-(4-Benzylpiperazinyl)methyl	Benzyl	19.0 ± 0.7
2j	–	1-Pyrrolidinylmethyl	Benzyl	n.a. ^c
3a	–	Hexyl	Benzyl	35.2 ± 2.8
3b	–	4-Methoxyphenyl	Benzyl	40.9 ± 2.8
3c	–	3,4,5-Trimethoxyphenyl	Benzyl	52.0 ± 3.6
3d	–	4-Chlorophenyl	Benzyl	73.4 ± 6.5

^a Values are means ± SEM of two determinations carried out in triplicate.^b Not tested.^c Not active. No inhibition was observed up to 200 μM of the test compound.

are reported in Supplementary material Figures SM-7–SM-13. IR spectra were recorded for neat samples on KBr plates on a Perkin Elmer Spectrum Bx FTIR spectrophotometer or on a Perkin Elmer FTIR Baragon 1000PC equipped with a Graseby-Specac golden gate. HRMS were collected on a Bruker MICROTOF-Q ES/ESI/TOF spectrometer. Elemental analyses (C, H, N, S) were used to confirm the purity of all tested compounds (>95%) and were performed on a CHN ThermoScientific Flash 2000 apparatus.

4.1.1. N-Benzyl-3-acetylindole (**39b**)

To a solution of 3-acetylindole (382 mg, 2.40 mmol) in ethanol (20 mL) was added potassium hydroxide (168 mg, 3 mmol) and the reaction mixture was stirred at room temperature until dissolution. The solvent was removed under reduced pressure and dry acetone

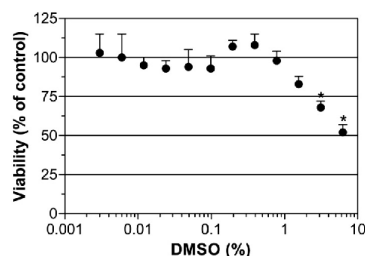


Fig. 6. Effect of DMSO on the viability of endothelial cells. Viability of HUVEC (black circle) was measured after 24 h of exposure of HUVEC to DMSO (0–6% v/v) in culture media containing 2% SVE. Results are presented as relative values to untreated cells ($n = 3$, triplicate). Differences are considered as significant when $p < 0.05$ using Bonferroni/Dunn test. *Versus control medium without DMSO.

was added followed by benzyl bromide (0.29 mL, 2.40 mmol). After 5 min, solid was filtered and filtrate was dried under reduced pressure to afford a white solid (586 mg, 2.35 mmol, 98%); mp 114 °C (Lit. [43]: 114–115 °C); ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 2.45 (s, 3H, CH₃), 5.49 (s, 2H, CH₂), 7.17–7.20 (m, 2H, 2 × CH), 7.17–7.33 (m, 5H, 5 × CH), 7.49–7.53 (m, 1H, CH), 8.16–8.20 (m, 1H, CH), 8.53 (s, 1H, CH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 27.3, 49.7, 111.0, 116.1, 121.6, 122.0, 122.9, 125.9, 127.2, 127.6, 128.6, 136.5, 137.0, 137.4, 192.3.

4.1.2. [3-Chloro-3-(1H-indol-3-yl)prop-2-enylidene]dimethyliminium perchlorate (**40a**)

To a solution of phosphorus oxychloride (27.96 mL, 300 mmol) was added dropwise dry DMF (23.20 mL, 300 mmol) at 0 °C. Then, the solution was stirred until formation of Vilsmeier–Haack reagent. After 30 min, 3-acetylindole (15.92 g, 100 mmol) dissolved in DMF (120 mL) was added dropwise. An orange solid appeared instantly. The solution was stirred at room temperature until dissolution of precipitated and after 2 h a pale orange solid formed in suspension of DMF. After 1 h, the reaction mixture was poured with good stirring into ice-water (800 mL) containing sodium perchlorate (24.49 g, 200 mmol). The precipitate was filtered and washed with Et₂O. The solid was purified in CHCl₃, filtered while hot to afford an orange solid (>95%, hygroscopic); mp 251 °C (Lit. [44]: 246–248 °C); ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 3.55 (s, 3H, CH₃), 3.65 (s, 3H, CH₃), 7.13 (d, 1H, CH, $J = 10.5$ Hz), 7.31–7.40 (m, 2H, 2 × CH), 7.56–7.61 (m, 1H, CH), 8.09–8.14 (m, 1H, CH), 8.60 (d, 1H, CH, $J = 3.4$ Hz), 8.79 (d, 1H, CH, $J = 10.5$ Hz), 12.87 (br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 41.2, 48.4, 107.1, 113.6, 114.2, 121.0, 123.8, 124.2, 125.0, 137.2, 138.1, 153.5, 163.3. HRMS (ESI): m/z calcd for C₁₃H₁₄ClN₂: 233.0840; found: 233.0835.

4.1.3. [3-Chloro-3-(1-benzyl-1H-indol-3-yl)prop-2-enylidene]dimethyliminium perchlorate (**40b**)

The same procedure as for **40a** was used and afforded an orange solid (>95%, hygroscopic); mp 168 °C; ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 3.51 (s, 3H, CH₃), 3.67 (s, 3H, CH₃), 5.61 (s, 2H, CH₂), 7.15 (d, 1H, CH, $J = 10.5$ Hz), 7.26–7.40 (m, 8H, 8 × CH), 7.67–7.70 (m, 1H, CH), 8.13–8.17 (m, 1H, CH), 8.84 (s, 1H, CH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 41.2, 48.5, 50.2, 107.7, 112.6, 113.6, 121.3, 123.6, 124.4, 127.3, 128.0, 128.7, 136.1, 137.8, 139.5, 152.5, 163.4. HRMS (ESI): m/z calcd for C₂₀H₂₀ClN₂: 323.1310; found: 323.1310.

4.1.4. 3-Chloro-3-(1H-indol-3-yl)-2-propenal oxime (**41a**)

To a solution of [3-chloro-3-(1H-indol-3-yl)prop-2-enylidene]dimethylammonium perchlorate **40a** (33.32 g, 100 mmol) in ethanol (150 mL) were added hydroxylamine hydrochloride (10.42 g, 150 mmol) followed by sodium acetate trihydrate (13.61 g, 100 mmol). The reaction mixture was heated at 60 °C for 6 h and kept at room temperature overnight. The solution was poured into ice-water, filtered and washed with cold water. The solid was triturated in CH₂Cl₂ to give a beige solid (17.87 g, 81 mmol, 81%); mp 99 °C; IR (neat): 3118 (OH), 3354 (NH) cm⁻¹; ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 6.84 + 7.33 (d + d, 1H, CH, $J = 9.4$ Hz), 7.12–7.23 (m, 2H, 2 × CH), 7.47 (d, 1H, CH, $J = 9.4$ Hz), 7.61 + 8.16 (d + d, 1H, CH, $J = 9.4$ Hz), 7.75–7.87 (m, 2H, 2 × CH), 11.32 + 11.51 (s + s, 1H, OH), 11.74 + 11.84 (br s + br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 20.6, 98.4, 117.2, 124.6, 129.5, 129.6, 131.2, 136.7, 141.8, 145.6, 161.6. HRMS (APCI): m/z calcd for C₁₁H₁₀ClN₂O: 221.0476; found: 221.0487.

4.1.5. 3-(1-Benzyl-1H-indol-3-yl)-3-chloro-2-propenal oxime (**41b**)

This compound was synthesized using the same procedure as for **41a** starting with [3-chloro-3-(1-benzyl-1H-indol-3-yl)prop-2-enylidene]dimethylammonium perchlorate **40b** to afford a pale

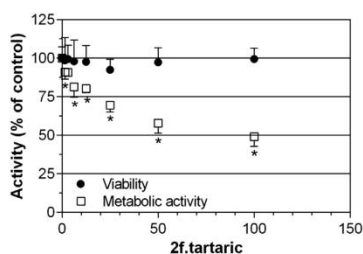


Fig. 7. Effect of **2f.tartaric** on metabolic activity and viability of endothelial cells. Metabolic activity (open square) and viability (black circle) were measured after 24 h of exposure of HUVEC to **2f.tartaric** (0–100 μM). Results are presented as relative values to untreated cells ($n = 3$, triplicate). Differences are considered as significant when $p < 0.05$ using Bonferroni/Dunn test. *Versus control medium without **2f.tartaric**.

yellow solid (20.82 g, 67 mmol, 67%); mp 69 °C; IR (neat): 3168 (OH) cm^{-1} ; $^1\text{H NMR}$ (250 MHz, $\text{DMSO-}d_6$) δ (ppm): 5.47 + 5.49 (s + s, 2H, CH_2), 6.87 + 7.36 (d + d, 1H, CH, $J = 9.4$ Hz), 7.17–7.31 (m, 7H, 7 × CH), 7.52–7.58 (m, 1H, CH), 7.62–8.17 (d + d, 1H, CH, $J = 9.4$ Hz), 7.78–7.88 (m, 1H, CH), 8.06 + 8.13 (s + s, 1H, CH), 11.36 + 11.55 (s + s, 1H, OH). $^{13}\text{C NMR}$ (62.5 MHz, $\text{DMSO-}d_6$) δ (ppm): 49.28, 49.34, 108.4, 111.4, 111.6, 112.86, 112.88, 115.4, 119.5, 119.8, 121.2, 121.5, 122.7, 122.8, 124.08, 124.14, 127.1, 127.2, 127.5, 127.6, 128.60, 128.62, 131.0, 131.4, 132.6, 132.9, 136.9, 137.0, 137.2, 137.3, 143.0, 146.9. HRMS (ESI): m/z calcd for $\text{C}_{18}\text{H}_{16}\text{ClN}_2\text{O}$: 311.0946; found: 311.0945.

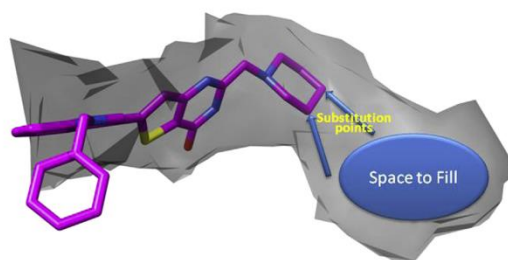


Fig. 9. Model for the molecules to be designed. In shaded gray is represented the ATP binding site. The two arrows indicate possible point for substitution on the **2f** piperidine ring. The new molecules will be designed to fill the blue oval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.1.6. 3-Chloro-3-(1H-indol-3-yl)-2-propenenitrile (**42a**)

To a solution of 3-chloro-3-(1H-indol-3-yl)-2-propenal oxime (**41a**) (3.00 g, 13.60 mmol) in CH_2Cl_2 (60 mL) were added di-2-pyridyl thionocarbonate (3.16 g, 13.60 mmol) followed by 4-*N,N*-dimethylaminopyridine (249 mg, 2.04 mmol). The solution was stirred overnight at room temperature. The reaction mixture was filtered on Celite and the filtrate was washed once with HCl 3% (30 mL) and once with NaHCO_3 10% (30 mL), dried on MgSO_4 and concentrated under reduce pressure. The residue was purified by silica gel column chromatography (cyclohexane:AcOEt = 80:20) to provide an orange solid (2.45 g, 12 mmol, 89%); mp 102 °C; IR (KBr): 2209 (CN) cm^{-1} ; $^1\text{H NMR}$ (250 MHz, $\text{DMSO-}d_6$) δ (ppm): 6.45 (s, 1H,

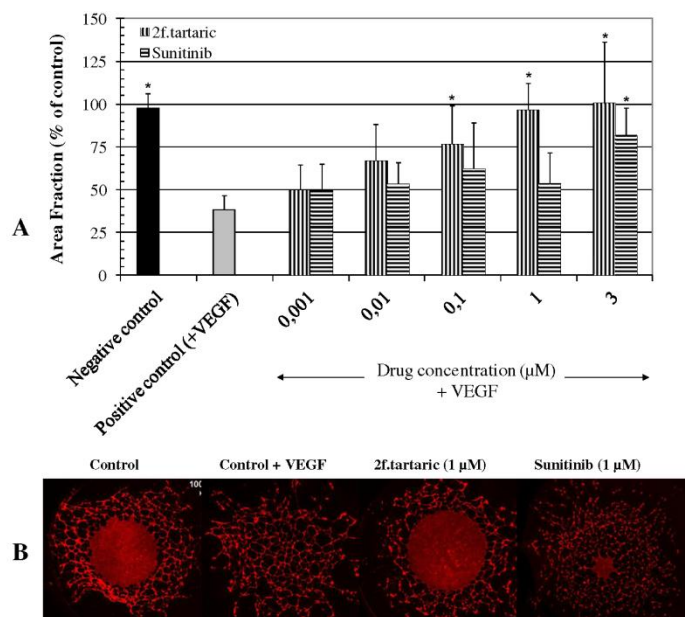


Fig. 8. Effect of **2f.tartaric** on endothelial cell tube formation. HUVEC were plated onto matrigel[®]-coated well and cultured in control media with 50 ng VEGF, **2f.tartaric** and Sunitinib as positive control. (A) Area fraction of endothelial cells was measured using NIS Element software and presented as percentage of variation with values obtained for control medium without VEGF as reference (negative control) ($n = 5$). Differences are considered as significant when $p < 0.05$ using PLSD Fisher statistical test. *Versus positive control medium without **2f.tartaric** (control + VEGF). (B) Photomicrographs were taken after fixation and phalloidin-sulfo-rhodamine staining and are representative of 5 experiments. Area fraction was calculated as area of cells in relation to whole field. Scale bar represent 1000 μm.

CH), 7.18–7.29 (m, 2H, 2× CH), 7.48–7.53 (m, 1H, CH), 7.92–7.97 (m, 1H, CH), 8.05 (s, 1H, CH), 12.18 (br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 89.8, 111.6, 112.9, 117.4, 119.8, 121.8, 123.1, 123.2, 132.0, 137.4, 145.6. HRMS (APCI): *m/z* calcd for C₁₁H₈ClN₂: 203.0371; found: 203.0373.

4.1.7. 3-(1-Benzyl-1H-indol-3-yl)-3-chloro-2-propenenitrile (**42b**)

This compound was synthesized using the same procedure as for **42a** starting with 3-(1-benzyl-1H-indol-3-yl)-3-chloro-2-propenal oxime **41b**. The residue was purified by silica gel column chromatography (cyclohexane:AcOEt = 80:20) to afford a beige solid (3.46 g, 11.83 mmol, 87%); mp 124 °C; IR (KBr): 2208 (CN) cm⁻¹; ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 5.52 (s, 2H, CH₂), 6.48 (s, 1H, CH), 7.20–7.34 (m, 7H, 7× CH), 7.56–7.60 (m, 1H, CH), 7.95–7.99 (m, 1H, CH), 8.31 (s, 1H, CH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 49.5, 90.4, 111.2, 111.8, 117.3, 120.1, 122.2, 123.3, 123.8, 127.2, 127.7, 128.7, 135.0, 136.8, 137.2, 145.0. HRMS (APCI): *m/z* calcd for C₁₈H₁₄ClN₂: 293.0840; found: 293.0839.

4.1.8. Ethyl 3-amino-5-(1H-indol-3-yl)-2-thiophenecarboxylate (**43a**)

A suspension of sodium sulfide nonahydrate (2.40 g, 10 mmol) in DMF (15 mL) was heated at 40 °C for 30 min. 3-Chloro-propenenitrile (2.03 g, 10 mmol) dissolved in DMF (15 mL) was added dropwise and the solution was heated at 50 °C for 2 h. Then, ethyl bromoacetate (1.11 mL, 10 mmol) in DMF (5 mL) was added dropwise and the reaction mixture was stirred at 50 °C for 2 h. A solution of NaOEt (681 mg, 10 mmol) in absolute ethanol was added and the solution was stirred at 50 °C for 2 h (the reaction was monitored by TLC). The reaction mixture was then cooled to room temperature, diluted with water (200 mL) and extracted three times with AcOEt (25 mL). The organic layer was dried over anhydrous magnesium sulfate and the solvent was evaporated under reduce pressure. The residue was purified by silica gel column chromatography (cyclohexane:AcOEt = 50:50) to provide a yellow oil (344 mg, 1.20 mmol, 12%); ¹H NMR (250 MHz, CDCl₃) δ (ppm): 1.29 (t, 3H, CH₃, *J* = 7.18 Hz), 4.24 (q, 2H, CH₂, *J* = 7.10 Hz), 5.09 (br s, 2H, NH₂), 6.64 (br s, 1H, CH), 7.11–7.21 (m, 2H, 2× CH), 7.28–7.33 (m, 1H, CH), 7.37–7.38 (m, 1H, CH), 7.87–7.91 (m, 1H, CH), 8.49 (br s, 1H, NH). ¹³C NMR (62.5 MHz, CDCl₃) δ (ppm): 14.7, 69.0, 98.3, 111.3, 111.7, 114.4, 120.0, 121.1, 123.0, 123.3, 124.9, 136.5, 143.8, 154.4, 164.9. HRMS (ESI): *m/z* calcd for C₁₅H₁₄N₂O₂SNa: 309.0668; found: 309.0661.

4.1.9. 3-Amino-5-(1H-indol-3-yl)-2-thiophenecarbonitrile (**43b**)

This compound was synthesized using the same procedure as for **43a** starting with 3-chloro-3-(1H-indol-3-yl)-2-propenenitrile **42a** and chloroacetonitrile. After cooling, the reaction mixture was poured into ice-water (200 mL) and the precipitate was filtered and washed with cold water. The residue was purified by silica gel column chromatography (cyclohexane:AcOEt = 70:30) to provide a beige solid (383 mg, 1.6 mmol, 16%); mp 138 °C; IR (neat): 2184 (CN), 3372, 3399 (NH₂) cm⁻¹; ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 6.42 (br s, 2H, NH₂), 6.83 (s, 1H, CH), 7.13–7.22 (m, 2H, 2× CH), 7.44–7.48 (m, 1H, CH), 7.76–7.80 (m, 1H, CH), 7.85–7.86 (m, 1H, CH), 11.66 (br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 71.4, 108.6, 112.4, 113.2, 116.4, 118.8, 120.5, 122.2, 124.1, 125.3, 136.6, 143.7, 157.8. HRMS (ESI): *m/z* calcd for C₁₅H₁₃N₃SNa: 262.0409; found: 262.0416.

4.1.10. 3-Amino-5-(1H-indol-3-yl)-2-thiophenecarboxamide (**43c**)

This compound was synthesized using the same procedure as for **43a** starting with 3-chloro-3-(1H-indol-3-yl)-2-propenenitrile **42a** and chloroacetamide diluted in DMF (5 mL). After cooling, the reaction mixture was poured into ice-water (200 mL) and the precipitate was filtered and washed with cold water. The solid was

trituted in Et₂O and filtered to afford a brown solid (1.78 g, 6.9 mmol, 69%); mp 196 °C; IR (KBr): 1632, 2203, 3177 (CONH₂), 3305, 3441 (NH₂) cm⁻¹; ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 6.46 (br s, 2H, NH₂), 6.57 (br s, 2H, NH₂), 6.82 (s, 1H, CH), 7.10–7.21 (m, 2H, 2× CH), 7.43–7.47 (m, 1H, CH), 7.75 (m, 1H, CH), 7.84–7.87 (m, 1H, CH), 11.53 (s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 97.5, 109.4, 112.2, 114.5, 119.0, 120.1, 121.9, 124.2, 124.4, 136.6, 139.2, 153.8, 166.4. HRMS (ESI): *m/z* calcd for C₁₅H₁₂N₃OS: 258.0696; found: 258.0676.

4.1.11. Ethyl 3-amino-5-(1-benzyl-1H-indol-3-yl)-2-thiophenecarboxylate (**43d**)

This compound was synthesized using the same procedure as for **43a** starting with 3-(1-benzyl-1H-indol-3-yl)-3-chloro-2-propenenitrile **42b**. The residue was purified by silica gel column chromatography (cyclohexane:AcOEt = 70:30) to afford a pale yellow solid (2.26 g, 6 mmol, 60%); mp 139 °C; IR (KBr): 1646 (CO), 3352, 3458 (NH₂) cm⁻¹; ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 1.26 (t, 3H, CH₃, *J* = 7.05 Hz), 4.19 (q, 2H, CH₂, *J* = 7.05 Hz), 5.46 (s, 2H, CH₂), 6.54 (br s, 2H, NH₂), 6.88 (s, 1H, CH), 7.17–7.24 (m, 2H, 2× CH), 7.28–7.34 (m, 5H, 5× CH), 7.53–7.57 (m, 1H, CH), 7.83–7.86 (m, 1H, CH), 8.09 (s, 1H, CH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 14.5, 49.3, 59.1, 93.8, 108.9, 111.1, 113.8, 119.4, 120.8, 122.3, 124.9, 127.2, 127.5, 128.4, 128.6, 136.3, 137.5, 142.6, 155.6, 163.6. HRMS (ESI): *m/z* calcd for C₂₂H₂₀N₂O₂SNa: 399.1138; found: 399.1137.

4.1.12. 3-Amino-5-(1-benzyl-1H-indol-3-yl)-2-thiophenecarbonitrile (**43e**)

This compound was synthesized using the same procedure as for **43a** starting with 3-(1-benzyl-1H-indol-3-yl)-3-chloro-2-propenenitrile **42b** and chloroacetonitrile. After cooling, the reaction mixture was poured into ice-water (200 mL) and the precipitate was filtered and washed with cold water. The solid was trituted in petroleum ether and filtered to afford a pale yellow solid (3.20 g, 9.7 mmol, 97%); mp 160 °C; IR (KBr): 2186 (CN), 3380, 3479 (NH₂) cm⁻¹; ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 5.46 (s, 2H, CH₂), 6.47 (br s, 2H, NH₂), 6.84 (s, 1H, CH), 7.16–7.22 (m, 2H, 2× CH), 7.24–7.34 (m, 5H, 5× CH), 7.53–7.57 (m, 1H, CH), 7.78–7.82 (m, 1H, CH), 8.09 (s, 1H, CH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 49.3, 71.5, 108.3, 111.2, 113.5, 116.3, 119.1, 120.9, 122.5, 124.7, 127.2, 127.5, 128.6, 136.3, 137.4, 143.0, 157.8, 157.9. HRMS (ESI): *m/z* calcd for C₂₀H₁₈N₃SNa: 352.0879; found: 352.0879.

4.1.13. 3-Amino-5-(1-benzyl-1H-indol-3-yl)-2-thiophenecarboxamide (**43f**)

This compound was synthesized using the same procedure as for **43a** starting with 3-(1-benzyl-1H-indol-3-yl)-3-chloro-2-propenenitrile **42b** and chloroacetamide diluted in DMF (5 mL). After cooling, the reaction mixture was poured into ice-water (200 mL) and the precipitate was filtered and washed with cold water. The residue was purified by silica gel column chromatography (cyclohexane:AcOEt = 50:50) to provide a pale brown solid (2.47 g, 7.1 mmol, 70%); mp 178 °C; IR (KBr): 1655, 2361, 3137 (CONH₂), 3314, 3455 (NH₂) cm⁻¹; ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 5.46 (s, 2H, CH₂), 6.47 (br s, 2H, NH₂), 6.76 (br s, 2H, NH₂), 6.82 (s, 1H, CH), 7.16–7.20 (m, 2H, 2× CH), 7.22–7.33 (m, 5H, 5× CH), 7.52–7.56 (m, 1H, CH), 7.85–7.89 (m, 1H, CH), 7.95 (s, 1H, CH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 49.2, 97.6, 109.2, 111.0, 114.8, 119.4, 120.5, 122.2, 125.0, 127.2, 127.5, 127.6, 128.6, 136.3, 137.6, 138.5, 153.9, 166.3. HRMS (ESI): *m/z* calcd for C₂₂H₁₇N₃OSNa: 370.0985; found: 370.0983.

4.1.14. 6-(1H-indol-3-yl)thieno[3,2-*d*]pyrimidin-4(3H)-one (**44a**)

In round-bottom flask was introduced formic acid (10 mL) and the solution was put at 0 °C. 3-Amino-5-(1H-indol-3-yl)-2-

thiophenecarbonitrile **43b** (239 mg, 1 mmol) was added slowly followed by 30 drops of concentrated sulfuric acid. After 30 min, the ice bath was removed and the reaction was heated at 40 °C for 24 h. The reaction mixture was then cooled to room temperature and poured into ice-water (50 mL). The precipitate was filtered, washed twice with water and once with Et₂O to afford a brown solid (225 mg, 0.84 mmol, 84%); mp 162 °C; IR (neat): 1637 (C=O), 3215 (NH) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 7.19–7.23 (m, 2H, 2 × CH), 7.47–7.51 (m, 1H, CH), 7.59 (s, 1H, CH), 7.95–7.99 (m, 1H, CH), 8.08–8.09 (m, 1H, CH), 8.13 (s, 1H, CH), 11.79 (br s, 1H, NH), 12.37 (br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 114.3, 117.6, 122.9, 124.1, 124.3, 126.0, 127.5, 129.4, 131.3, 142.0, 151.7, 152.0, 162.0, 163.9. HRMS (ESI): *m/z* calcd for C₁₄H₉N₃O₂Na: 290.0359; found: 290.0369.

4.1.15. 6-(1-Benzyl-1H-indol-3-yl)thieno[3,2-*d*]pyrimidin-4(3H)-one (**44b**)

This compound was synthesized using the same procedure as for **44a** starting with 3-amino-5-(1-benzyl-1H-indol-3-yl)-2-thiophene carbonitrile **43e** to provide a beige solid (315 mg, 0.88 mmol, 88%); mp 286 °C; IR (KBr): 1659 (C=O), 2789 (NH) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 5.50 (s, 2H, CH₂), 7.18–7.36 (m, 7H, 7 × CH), 7.54–7.59 (m, 1H, CH), 7.61 (s, 1H, CH), 7.97–8.04 (m, 1H, CH), 8.14 (s, 1H, CH), 8.30 (s, 1H, CH), 12.41 (br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 49.4, 108.8, 111.2, 118.1, 119.0, 119.5, 121.1, 122.5, 124.8, 127.1, 127.6, 128.6, 129.3, 136.5, 137.4, 145.7, 146.8, 156.8, 158.6. HRMS (APCI): *m/z* calcd for C₂₁H₁₆N₃O₂S: 358.1009; found: 358.1005.

4.1.16. 4-Chloro-6-(1H-indol-3-yl)thieno[3,2-*d*]pyrimidine (**45a**)

To a solution of dry DMF (5 mL) was added dropwise phosphorus oxychloride (0.47 mL, 5 mmol) at 0 °C. After few minutes of stirring, 6-(1H-indol-3-yl)thieno[3,2-*d*]pyrimidin-4(3H)-one **44a** (267 mg, 1 mmol) dissolved in DMF (5 mL) was added dropwise at 0 °C. The reaction mixture was heated at 50 °C overnight. After cooling, the solution was poured into ice-water (100 mL) and the precipitate was filtered and washed several times with cold water. The crude product was purified by silica gel column chromatography (cyclohexane:AcOEt = 60:40) to provide a yellow solid (129 mg, 0.45 mmol, 45%); mp 274 °C. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 7.47–7.54 (m, 2H, 2 × CH), 8.11–8.18 (m, 2H, 2 × CH), 8.29–8.35 (m, 1H, CH), 8.61 (s, 1H, CH), 9.01 (s, 1H, CH), 9.52 (s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 120.1, 120.8, 121.2, 121.6, 122.3, 125.0, 125.6, 127.2, 136.7, 142.4, 147.1, 156.5, 156.8, 157.8. HRMS (ESI): *m/z* calcd for C₁₄H₉ClN₃S: 286.0200; found: 286.0194.

4.1.17. 6-(1-Benzyl-1H-indol-3-yl)-4-chlorothieno[3,2-*d*]pyrimidine (**45b**)

This compound was synthesized using the same procedure as for **45a** starting with 6-(1-benzyl-1H-indol-3-yl)thieno[3,2-*d*]pyrimidin-4(3H)-one **44b**. The crude product was triturated in petroleum ether and filtered to afford yellow solid (368 mg, 0.98 mmol, 98%); mp 163 °C. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 5.62 (s, 2H, CH₂), 7.27–7.39 (m, 7H, 7 × CH), 7.67–7.73 (m, 1H, CH), 8.06–8.12 (m, 1H, CH), 8.96 (s, 1H, CH), 9.10 (s, 1H, CH), 10.41 (s, 1H, CH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 49.6, 106.8, 111.6, 119.3, 121.8, 123.1, 125.57, 125.64, 127.3, 127.7, 128.7, 134.1, 136.4, 137.0, 148.0, 153.6, 156.5, 157.5, 162.2. HRMS (ESI): *m/z* calcd for C₂₁H₁₅ClN₃S: 376.0670; found: 376.0679.

4.1.18. 6-(1-Benzyl-1H-indol-3-yl)-4-(4-methoxyphenoxy)thieno[3,2-*d*]pyrimidine (**1a**)

Preparation of 4-methoxyphenolate. In round bottom flask was placed sodium hydroxide (200 mg, 5 mmol) in water (1 mL). Then, *p*-methoxyphenol (621 mg, 5 mmol) was added to the solution. After

dissolution, the reaction mixture was dried under reduce pressure to give quantitative white solid. Preparation of the title compound. To a solution of 6-(1-benzyl-1H-indol-3-yl)-4-chlorothieno[3,2-*d*]pyrimidine **45b** (376 mg, 1 mmol) in dry DMF (5 mL) was added 4-methoxyphenolate (292 mg, 2 mmol) freshly prepared. The reaction mixture was stirred at room temperature overnight. The solution was poured into ice-water (100 mL) and the precipitate was filtered and washed with water. The crude product was purified by silica gel column chromatography (cyclohexane:AcOEt = 80:20) to provide yellow solid (375 mg, 0.81 mmol, 81%); mp 174 °C. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 3.79 (s, 3H, CH₃), 5.52 (s, 2H, CH₂), 7.01 (d, 2H, 2 × CH, *J* = 9.1 Hz), 7.23–7.35 (m, 9H, 9 × CH), 7.59–7.62 (m, 1H, CH), 7.83 (s, 1H, CH), 8.06–8.10 (m, 1H, CH), 8.41 (s, 1H, CH), 8.62 (s, 1H, CH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 49.5, 55.4, 108.5, 111.4, 113.8, 114.6, 116.4, 119.6, 121.4, 122.7, 122.9, 124.9, 127.2, 127.6, 128.6, 130.2, 136.6, 137.3, 145.2, 147.5, 154.4, 157.0, 163.0, 164.0. HRMS (APCI): *m/z* calcd for C₂₈H₂₂N₃O₂S: 464.1427; found: 464.1432.

4.1.19. 6-(1-Benzyl-1H-indol-3-yl)-4-(4-methyl-1-piperazinyl)thieno[3,2-*d*]pyrimidine (**1b**)

A solution of 6-(1-benzyl-1H-indol-3-yl)-4-chlorothieno[3,2-*d*]pyrimidine **45b** (38 mg, 0.1 mmol) in few drops of *N*-methylpiperazine was stirred at room temperature overnight. Then, water (10 mL) was added to the solution. The aqueous layer was extracted twice with AcOEt (10 mL). The organic layer was washed once with hydrochloric acid 1 N (10 mL) and once with brine (10 mL), dried over MgSO₄ and concentrated under reduce pressure to provide yellow solid (15 mg, 0.034 mmol, 34%); mp 110 °C. ¹H NMR (250 MHz, acetone-*d*₆) δ (ppm): 2.29 (s, 3H, CH₃), 2.52–2.56 (m, 4H, 2 × CH₂), 4.02–4.11 (m, 4H, 2 × CH₂), 5.69 (s, 2H, CH₂), 7.24–7.38 (m, 7H, 7 × CH), 7.58–7.63 (m, 1H, CH), 8.15–8.21 (m, 1H, CH), 8.56 (s, 1H, CH), 8.94 (s, 1H, CH), 10.62 (s, 1H, CH). ¹³C NMR (62.5 MHz, acetone-*d*₆) δ (ppm): 46.2, 46.8, 51.1, 55.6, 108.0, 112.0, 112.2, 121.4, 122.6, 123.9, 125.2, 127.4, 128.2, 128.7, 129.6, 135.9, 137.8, 138.0, 152.4, 155.5, 158.1, 186.4. HRMS (ESI): *m/z* calcd for C₂₆H₂₆N₅S: 440.1903; found: 440.1911.

4.1.20. 6-(1-Benzyl-1H-indol-3-yl)-4-chlorothieno[3,2-*d*]1,2,3-triazine (**46**)

To a solution of 3-amino-5-(1-benzyl-1H-indol-3-yl)-2-thiophenecarbonitrile **43e** (659 mg, 2 mmol) in concentrated hydrochloric acid (6 mL) was added dropwise sodium nitrite (193 mg, 2.80 mmol) dissolved in water (5 mL). After addition, the reaction mixture was stirred at room temperature overnight. The solution was poured into ice-water (150 mL) and the precipitate was filtered and washed with water. The residue was purified by silica gel column chromatography (cyclohexane:AcOEt = 90:10) to afford yellow solid (663 mg, 1.76 mmol, 88%); mp 173 °C. ¹H NMR (250 MHz, CDCl₃) δ (ppm): 5.43 (s, 2H, CH₂), 7.20–7.24 (m, 2H, 2 × CH), 7.33–7.45 (m, 6H, 6 × CH), 7.71 (s, 1H, CH), 7.93 (s, 1H, CH), 8.07–8.11 (m, 1H, CH). ¹³C NMR (62.5 MHz, CDCl₃) δ (ppm): 50.8, 109.2, 111.0, 115.1, 119.9, 122.5, 123.9, 125.4, 127.1, 127.4, 128.4, 129.0, 129.2, 129.5, 135.6, 137.4, 138.3, 150.9. HRMS (APCI): *m/z* calcd for C₂₀H₁₄ClN₄S: 377.0622; found: 377.0634.

4.1.21. 6-(1-Benzyl-1H-indol-3-yl)-4-(4-methoxybenzyl)thieno[3,2-*d*]1,2,3-triazine (**1c**)

This compound was synthesized using the same procedure as for **1a** starting with 6-(1-benzyl-1H-indol-3-yl)-4-chlorothieno[3,2-*d*]1,2,3-triazine **46** and 4-methoxyphenolate. The residue was purified by silica gel column chromatography (cyclohexane:AcOEt = 80:20) to afford yellow solid (372 mg, 0.80 mmol, 80%); mp 165 °C. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 3.81 (s, 3H, CH₃), 5.54 (s, 2H, CH₂), 7.07 (d, 2H, 2 × CH, *J* = 9.0 Hz), 7.26–7.38 (m, 9H, 9 × CH), 7.61–7.64 (m, 1H,

CH), 8.08–8.12 (m, 1H, CH), 8.17 (s, 1H, CH), 8.52 (s, 1H, CH). ^{13}C NMR (62.5 MHz, DMSO- d_6) δ (ppm): 49.6, 55.5, 107.8, 111.5, 114.4, 114.6, 114.8, 119.6, 121.7, 122.9, 123.0, 124.7, 127.2, 127.7, 128.7, 131.2, 136.7, 137.1, 144.9, 149.3, 157.4, 159.8, 160.1. HRMS (APCI): m/z calcd for $\text{C}_{27}\text{H}_{21}\text{N}_4\text{O}_2\text{S}$: 465.1380; found: 465.1372.

4.1.22. 3-[(Chloroacetyl)amino]-5-(1H-indol-3-yl)-2-thiophenecarboxamide (**47a**)

To a solution of 3-amino-5-(1H-indol-3-yl)-2-thiophenecarboxamide **43c** (772 mg, 3 mmol) and Et_3N (500 μL , 3.6 mmol) in dry THF was added dropwise chloroacetyl chloride (286 μL , 3.6 mmol) diluted in THF (2 mL) at 0 °C. The ice-bath was removed and the reaction mixture was stirred at room temperature 2 h. Then, the solution was poured into ice-water (30 mL) and the precipitate was filtered and washed once with cold water and once with petroleum ether to afford a yellow solid (411 mg, 1.23 mmol, 41%); mp 246 °C; IR (neat): 1652 (C=O), 2369, 2931, 3278 (NH + NH₂) cm^{-1} . ^1H NMR (250 MHz, DMSO- d_6) δ (ppm): 4.47 (s, 2H, CH₂), 7.16–7.24 (m, 2H, 2 \times CH), 7.41–7.46 (m, 1H, CH), 7.60 (br s, 2H, NH₂), 7.86–7.93 (m, 1H, CH), 7.93 (s, 1H, CH), 8.19 (s, 1H, CH), 11.69 (br s, 1H, NH), 12.04 (br s, 1H, NH). ^{13}C NMR (62.5 MHz, DMSO- d_6) δ (ppm): 43.1, 108.8, 109.4, 112.4, 115.0, 118.7, 120.4, 122.2, 124.2, 125.1, 136.6, 140.8, 142.1, 164.0, 164.4. HRMS (ESI): m/z calcd for $\text{C}_{15}\text{H}_{12}\text{ClN}_3\text{O}_2\text{SNa}$: 356.0231; found: 356.0236.

4.1.23. 5-(1-Benzyl-1H-indol-3-yl)-3-[(chloroacetyl)amino]-2-thiophenecarboxamide (**47b**)

This compound was synthesized using the same procedure as for **47a** starting with 3-amino-5-(1-benzyl-1H-indol-3-yl)-2-thiophenecarboxamide **43f** to afford a yellow solid (1.07 g, 2.52 mmol, 84%); mp 183 °C; IR (neat): 1652 (C=O), 2360, 3094 (NH + NH₂) cm^{-1} . ^1H NMR (250 MHz, DMSO- d_6) δ (ppm): 4.47 (s, 2H, CH₂), 5.49 (s, 2H, CH₂), 7.20–7.31 (m, 7H, 7 \times CH), 7.56–7.59 (m, 1H, CH), 7.60 (br s, 2H, NH₂), 7.87–7.91 (m, 1H, CH), 8.19 (s, 1H, CH), 8.21 (s, 1H, CH), 12.04 (br s, 1H, NH). ^{13}C NMR (62.5 MHz, DMSO- d_6) δ (ppm): 43.1, 49.3, 108.5, 109.6, 111.3, 115.1, 119.0, 120.8, 122.5, 124.8, 127.2, 127.5, 128.4, 128.6, 136.4, 137.5, 140.1, 142.1, 164.0, 165.4. HRMS (ESI): m/z calcd for $\text{C}_{22}\text{H}_{18}\text{ClN}_3\text{O}_2\text{SNa}$: 446.0700; found: 446.0713.

4.1.24. 5-(1H-indol-3-yl)-3-[(1-piperidinylacetyl)amino]-2-thiophenecarboxamide (**48a**)

A solution of 3-[(chloroacetyl)amino]-5-(1H-indol-3-yl)-2-thiophenecarboxamide **47a** (80 mg, 0.24 mmol), piperidine (24 μL , 0.24 mmol) and dry potassium carbonate (33 mg, 0.24 mmol) in dry CH_3CN (3 mL) was heated at reflux 1 h. After cooling, the reaction mixture was poured into ice-water (30 mL) and the precipitate was filtered and washed once with cold water and once with petroleum ether to give a pale red solid (91 mg, 0.238 mmol, 99%); mp 247 °C; IR (neat): 1667 (C=O), 2359, 2930, 3166 (NH + NH₂) cm^{-1} . ^1H NMR (250 MHz, DMSO- d_6) δ (ppm): 1.35–1.45 (m, 2H, CH₂), 1.58–1.68 (m, 4H, 2 \times CH₂), 2.42–2.48 (m, 4H, 2 \times CH₂), 3.08 (s, 2H, CH₂), 7.17–7.21 (m, 2H, 2 \times CH), 7.46–7.49 (m, 3H, NH₂ + CH), 7.87–7.89 (m, 2H, 2 \times CH), 8.29 (s, 1H, CH), 11.66 (br s, 1H, NH), 12.17 (br s, 1H, NH). ^{13}C NMR (62.5 MHz, DMSO- d_6) δ (ppm): 23.4, 25.3, 54.3, 62.3, 108.7, 109.0, 112.4, 115.3, 118.7, 120.3, 122.1, 124.3, 124.7, 136.6, 140.1, 142.5, 165.2, 168.6. HRMS (ESI): m/z calcd for $\text{C}_{20}\text{H}_{23}\text{N}_4\text{O}_2\text{S}$: 383.1536; found: 383.1529.

4.1.25. 5-(1H-indol-3-yl)-3-[(4-morpholinylacetyl)amino]-2-thiophenecarboxamide (**48b**)

This compound was synthesized using the same procedure as for **48a** starting with 3-[(chloroacetyl)amino]-5-(1H-indol-3-yl)-2-thiophenecarboxamide **47a** and morpholine to afford a beige solid (91 mg, 0.238 mmol, 99%); mp 261 °C; IR (neat): 1643 (C=O), 2808,

3150, 3277 (NH + NH₂) cm^{-1} . ^1H NMR (250 MHz, DMSO- d_6) δ (ppm): 2.49–2.51 (m, 4H, 2 \times CH₂), 3.15 (s, 2H, CH₂), 3.69–3.73 (m, 4H, 2 \times CH₂), 7.17–7.21 (m, 2H, 2 \times CH), 7.46–7.49 (m, 3H, NH₂ + CH), 7.85–7.88 (m, 2H, 2 \times CH), 8.28 (s, 1H, CH), 11.66 (br s, 1H, NH), 12.21 (br s, 1H, NH). ^{13}C NMR (62.5 MHz, DMSO- d_6) δ (ppm): 53.3, 61.6, 66.1, 108.8, 109.0, 112.4, 115.3, 118.7, 120.3, 122.1, 124.2, 124.8, 136.6, 140.2, 142.4, 165.3, 168.0. HRMS (ESI): m/z calcd for $\text{C}_{19}\text{H}_{21}\text{N}_4\text{O}_3\text{S}$: 385.1329; found: 385.1337.

4.1.26. 5-(1H-indol-3-yl)-3-[(4-methyl-1-piperazinyl)acetyl]amino]-2-thiophenecarboxamide (**48c**)

This compound was synthesized using the same procedure as for **48a** starting with 3-[(chloroacetyl)amino]-5-(1H-indol-3-yl)-2-thiophenecarboxamide **47a** and *N*-methylpiperazine to afford a pale red solid (93 mg, 0.235 mmol, 98%); mp 258 °C; IR (neat): 1658 (C=O), 2808, 3159, 3258 (NH + NH₂) cm^{-1} . ^1H NMR (250 MHz, DMSO- d_6) δ (ppm): 2.18 (s, 3H, CH₃), 2.44–2.47 (m, 8H, 4 \times CH₂), 3.12 (s, 2H, CH₂), 7.14–7.20 (m, 2H, 2 \times CH), 7.43–7.50 (m, 3H, NH₂ + CH), 7.83–7.88 (m, 2H, 2 \times CH), 8.27 (s, 1H, CH), 11.64 (br s, 1H, NH), 12.05 (br s, 1H, NH). ^{13}C NMR (62.5 MHz, DMSO- d_6) δ (ppm): 45.6, 52.8, 54.3, 61.4, 108.8, 108.9, 112.4, 115.4, 118.7, 120.3, 122.1, 124.2, 124.7, 136.6, 140.1, 142.4, 165.2, 168.1. HRMS (ESI): m/z calcd for $\text{C}_{20}\text{H}_{24}\text{N}_5\text{O}_2\text{S}$: 398.1645; found: 398.1645.

4.1.27. 3-[(4-Benzyl-1-piperazinyl)acetyl]amino]-5-(1H-indol-3-yl)-2-thiophenecarboxamide (**48d**)

This compound was synthesized using the same procedure as for **48a** starting with 3-[(chloroacetyl)amino]-5-(1H-indol-3-yl)-2-thiophenecarboxamide **47a** and *N*-benzylpiperazine to provide a pale brown solid (113 mg, 0.238 mmol, 99%); mp 268 °C; IR (neat): 1644 (C=O), 2360, 2812, 3275 (NH + NH₂) cm^{-1} . ^1H NMR (250 MHz, DMSO- d_6) δ (ppm): 2.49–2.51 (m, 8H, 4 \times CH₂), 3.15 (s, 2H, CH₂), 3.47 (s, 2H, CH₂), 7.17–7.32 (m, 7H, 7 \times CH), 7.46–7.49 (m, 3H, NH₂ + CH), 7.86–7.89 (m, 2H, 2 \times CH), 8.29 (s, 1H, CH), 11.65 (br s, 1H, NH), 12.08 (br s, 1H, NH). ^{13}C NMR (62.5 MHz, DMSO- d_6) δ (ppm): 52.5, 53.0, 61.5, 62.0, 108.8, 109.0, 112.4, 115.4, 118.7, 120.3, 122.1, 124.7, 126.9, 128.1, 128.7, 136.6, 138.5, 140.1, 142.4, 165.2, 168.1. HRMS (ESI): m/z calcd for $\text{C}_{26}\text{H}_{28}\text{N}_5\text{O}_2\text{S}$: 474.1958; found: 474.1981.

4.1.28. 5-(1H-indol-3-yl)-3-[(1-pyrrolidinylacetyl)amino]-2-thiophenecarboxamide (**48e**)

This compound was synthesized using the same procedure as for **48a** starting with 3-[(chloroacetyl)amino]-5-(1H-indol-3-yl)-2-thiophenecarboxamide **47a** and pyrrolidine to provide a yellow solid (87 mg, 0.235 mmol, 98%); mp 237 °C; IR (neat): 1672 (C=O), 2853, 2957, 3219 (NH + NH₂) cm^{-1} . ^1H NMR (250 MHz, DMSO- d_6) δ (ppm): 1.74–1.83 (m, 4H, 2 \times CH₂), 2.61–2.69 (m, 4H, 2 \times CH₂), 3.32 (s, 2H, CH₂), 7.16–7.23 (m, 2H, 2 \times CH), 7.45–7.50 (m, 3H, NH₂ + CH), 7.85–7.90 (m, 2H, 2 \times CH), 8.26 (s, 1H, CH), 11.66 (br s, 1H, NH), 12.03 (br s, 1H, NH). ^{13}C NMR (62.5 MHz, DMSO- d_6) δ (ppm): 23.5, 53.7, 58.7, 108.2, 108.8, 109.0, 112.4, 115.4, 118.7, 120.3, 122.2, 124.3, 124.8, 136.6, 140.2, 142.5, 165.3. HRMS (ESI): m/z calcd for $\text{C}_{19}\text{H}_{21}\text{N}_4\text{O}_2\text{S}$: 369.1380; found: 369.1380.

4.1.29. 5-(1-Benzyl-1H-indol-3-yl)-3-[(1-piperidinylacetyl)amino]-2-thiophenecarboxamide (**48f**)

This compound was synthesized using the same procedure as for **48a** starting with 5-(1-benzyl-1H-indol-3-yl)-3-[(chloroacetyl)amino]-2-thiophenecarboxamide **47b** and piperidine to afford a beige solid (104 mg, 0.221 mmol, 92%); mp 112 °C; IR (neat): 1644 (C=O), 2933, 3176 (NH + NH₂) cm^{-1} . ^1H NMR (250 MHz, CDCl_3) δ (ppm): 1.35–1.47 (m, 2H, CH₂), 1.61–1.72 (m, 4H, 2 \times CH₂), 2.43–2.52 (m, 4H, 2 \times CH₂), 3.08 (s, 2H, CH₂), 5.28 (s, 2H, CH₂), 5.39 (sl, 2H, NH₂), 7.07–7.29 (m, 8H, 8 \times CH), 7.47 (s, 1H, CH), 7.94–8.02 (m, 1H, CH), 8.36 (s, 1H, CH), 11.93 (br s, 1H, NH). ^{13}C NMR (62.5 MHz, CDCl_3) δ (ppm): 23.9,

26.0, 50.4, 55.1, 63.0, 107.9, 110.2, 110.3, 117.3, 120.2, 121.1, 122.9, 125.7, 127.0, 127.3, 128.0, 129.0, 136.4, 137.0, 140.6, 143.9, 165.5, 169.8. HRMS (ESI): *m/z* calcd for C₂₇H₂₉N₄O₂S: 473.2006; found: 473.2006.

4.1.30. 5-(1-Benzyl-1H-indol-3-yl)-3-[(4-morpholinylacetyl)amino]-2-thiophenecarboxamide (48g)

This compound was synthesized using the same procedure as for **48a** starting with 5-(1-benzyl-1H-indol-3-yl)-3-[(chloroacetyl)amino]-2-thiophenecarboxamide **47b** and morpholine to afford a beige solid (105 mg, 0.221 mmol, 92%); mp 116 °C; IR (neat): 1645 (C=O), 2816, 3188 (NH – NH₂) cm⁻¹. ¹H NMR (250 MHz, CDCl₃) δ (ppm): 2.62–2.65 (m, 4H, 2 × CH₂), 3.21 (s, 2H, CH₂), 3.85–3.89 (m, 4H, 2 × CH₂), 5.33 (s, 2H, CH₂), 5.57 (sl, 2H, NH₂), 7.15–7.35 (m, 8H, 8 × CH), 7.52 (s, 1H, CH), 8.01–8.06 (m, 1H, CH), 8.41 (s, 1H, CH), 12.04 (br s, 1H, NH). ¹³C NMR (62.5 MHz, CDCl₃) δ (ppm): 50.4, 53.8, 62.5, 67.0, 108.0, 110.1, 110.4, 117.1, 120.2, 121.1, 123.0, 125.6, 127.0, 127.3, 128.0, 129.0, 136.4, 137.0, 140.9, 143.7, 165.6, 168.7. HRMS (ESI): *m/z* calcd for C₂₆H₂₆N₄O₃SNa: 497.1618; found: 497.1610.

4.1.31. 5-(1-Benzyl-1H-indol-3-yl)-3-[(4-methyl-1-piperazinyl)acetyl]amino-2-thiophenecarboxamide (48h)

This compound was synthesized using the same procedure as for **48a** starting with 5-(1-benzyl-1H-indol-3-yl)-3-[(chloroacetyl)amino]-2-thiophenecarboxamide **47b** and *N*-methylpiperazine to afford a beige solid (99 mg, 0.204 mmol, 85%); mp 128 °C; IR (neat): 1644 (C=O), 2799, 3192 (NH + NH₂) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 2.36 (s, 3H, CH₃), 2.59–2.72 (m, 8H, 4 × CH₂), 3.22 (s, 2H, CH₂), 5.34 (s, 2H, CH₂), 5.47 (sl, 2H, NH₂), 7.15–7.36 (m, 8H, 8 × CH), 7.53 (s, 1H, CH), 8.02–8.06 (m, 1H, CH), 8.41 (s, 1H, CH), 11.92 (br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 45.9, 50.4, 53.4, 54.9, 62.1, 108.0, 110.1, 110.3, 117.3, 120.2, 121.1, 122.9, 125.6, 127.0, 127.3, 128.0, 129.0, 136.4, 137.0, 140.7, 143.8, 165.4, 169.0. HRMS (ESI): *m/z* calcd for C₂₇H₃₀N₅O₂S: 488.2115; found: 488.2119.

4.1.32. 5-(1-Benzyl-1H-indol-3-yl)-3-[(4-benzyl-1-piperazinyl)acetyl]amino-2-thiophenecarboxamide (48i)

This compound was synthesized using the same procedure as for **48a** starting with 5-(1-benzyl-1H-indol-3-yl)-3-[(chloroacetyl)amino]-2-thiophenecarboxamide **47b** and *N*-benzylpiperazine to afford a beige solid (134 mg, 0.238 mmol, 99%); mp 204 °C; IR (neat): 1658 (C=O), 2816, 3182 (NH₂) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 2.60–2.71 (m, 8H, 4 × CH₂), 3.21 (s, 2H, CH₂), 3.59 (s, 2H, CH₂), 5.34 (s, 2H, CH₂), 5.51 (sl, 2H, NH₂), 7.15–7.37 (m, 13H, 13 × CH), 7.53 (s, 1H, CH), 8.02–8.06 (m, 1H, CH), 8.41 (s, 1H, CH), 11.94 (br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 50.4, 53.1, 53.5, 62.2, 62.8, 108.0, 110.1, 110.3, 117.3, 120.2, 121.1, 122.9, 125.6, 126.98, 126.99, 127.1, 127.3, 128.0, 128.2, 129.0, 129.2, 136.4, 137.0, 140.7, 143.8, 165.5, 169.1. HRMS (ESI): *m/z* calcd for C₃₃H₃₄N₅O₂S: 564.2428; found: 564.2453.

4.1.33. 5-(1-Benzyl-1H-indol-3-yl)-3-[(1-pyrrolidinylacetyl)amino]-2-thiophenecarboxamide (48j)

This compound was synthesized using the same procedure as for **48a** starting with 5-(1-benzyl-1H-indol-3-yl)-3-[(chloroacetyl)amino]-2-thiophenecarboxamide **47b** and pyrrolidine to afford a beige solid (108 mg, 0.235 mmol, 98%); mp 180 °C; IR (neat): 1662 (C=O), 2787, 3149 (NH + NH₂) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 1.81–1.89 (m, 4H, 2 × CH₂), 2.63–2.70 (s, 4H, 2 × CH₂), 3.35 (s, 2H, CH₂), 5.28 (s, 2H, CH₂), 5.39 (sl, 2H, NH₂), 7.08–7.30 (m, 8H, 8 × CH), 7.46 (s, 1H, CH), 7.95–8.00 (m, 1H, CH), 8.35 (s, 1H, CH), 11.86 (br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 24.2, 50.4, 54.5, 59.6, 107.8, 110.1, 110.3, 117.3, 120.2, 121.1, 122.9, 125.6, 127.0, 127.3, 128.0, 129.0, 136.4, 137.0, 140.7, 144.0, 165.6, 169.7. HRMS (ESI): *m/z* calcd for C₂₆H₂₇N₄O₂S: 459.1849; found: 459.1867.

4.1.34. 6-(1H-Indol-3-yl)-2-(1-piperidinylmethyl)thieno[3,2-*d*]pyrimidin-4(3H)-one (2a)

To a solution of 5-(1H-Indol-3-yl)-3-[(1-piperidinylacetyl)amino]-2-thiophenecarboxamide **48a** (36.0 mg, 0.1 mmol) in DMF (2.5 mL) was added NaOH 2 N (2.5 mL) and heated at reflux for 30 min. After cooling, the reaction mixture was poured into ice-water under good stirring. The precipitate was filtered, washed once with water and once with petroleum ether to provide a beige solid (35.7 mg, 0.098 mmol, 98%); mp 350 °C (dec.); IR (neat): 1660 (C=O), 2937 (NH) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 1.31–1.33 (m, 2H, CH₂), 1.44–1.46 (m, 4H, 2 × CH₂), 2.38–2.46 (m, 4H, 2 × CH₂), 3.37 (s, 2H, CH₂), 7.12–7.17 (m, 2H, 2 × CH), 7.44–7.47 (m, 1H, CH), 7.52 (s, 1H, CH), 7.90–7.93 (m, 1H, CH), 8.01 (s, 1H, CH), 8.44 (br s, 1H, NH), 12.03 (br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 23.6, 25.4, 53.8, 61.0, 109.0, 112.4, 117.2, 117.6, 119.0, 120.7, 122.2, 124.1, 125.9, 136.8, 146.5, 156.3, 157.6, 158.4. HRMS (ESI): *m/z* calcd for C₂₀H₂₁N₄O₂S: 365.1431; found: 365.1431.

4.1.35. 6-(1H-Indol-3-yl)-2-(4-morpholinylmethyl)thieno[3,2-*d*]pyrimidin-4(3H)-one (2b)

This compound was synthesized using the same procedure as for **2a** starting with 5-(1H-Indol-3-yl)-3-[(4-morpholinylacetyl)amino]-2-thiophenecarboxamide **48b** to afford a beige solid (35.9 mg, 0.098 mmol, 98%); mp 262 °C; IR (neat): 1652 (C=O), 2835 (NH) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 2.47–2.51 (m, 2H, CH₂), 3.31–3.36 (m, 2H, CH₂), 3.46 (s, 2H, CH₂), 3.57–3.61 (m, 4H, 2 × CH₂), 7.18–7.22 (m, 2H, 2 × CH), 7.48–7.51 (m, 1H, CH), 7.57 (s, 1H, CH), 7.95–7.98 (m, 1H, CH), 8.06 (s, 1H, CH), 8.49 (br s, 1H, NH), 11.90 (br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 53.0, 60.5, 66.1, 109.0, 112.4, 117.4, 117.6, 119.0, 120.7, 122.2, 124.1, 126.0, 136.7, 146.5, 155.7, 157.5, 158.3. HRMS (ESI): *m/z* calcd for C₁₉H₁₈N₄O₂SNa: 389.1043; found: 389.1045.

4.1.36. 6-(1H-Indol-3-yl)-2-[(4-methyl-1-piperazinyl)methyl]thieno[3,2-*d*]pyrimidin-4(3H)-one (2c)

This compound was synthesized using the same procedure as for **2a** starting with 5-(1H-Indol-3-yl)-3-[(4-methyl-1-piperazinyl)acetyl]amino-2-thiophenecarboxamide **48c** to afford a beige solid (36.8 mg, 0.097 mmol, 97%); mp 350 °C (dec.); IR (neat): 1651 (C=O), 3125 (NH) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 2.45 (m, 3H, CH₃), 2.66–2.71 (m, 4H, 2 × CH₂), 2.75–2.81 (m, 4H, 2 × CH₂), 3.52 (s, 2H, CH₂), 7.15–7.24 (m, 2H, 2 × CH), 7.48–7.52 (m, 1H, CH), 7.57 (s, 1H, CH), 7.94–7.97 (m, 1H, NH), 8.06 (s, 1H, NH), 8.18 (s, 1H, CH), 12.04 (br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 43.1, 50.1, 52.9, 59.4, 109.0, 112.5, 117.4, 117.6, 118.8, 120.7, 122.3, 124.1, 126.0, 136.7, 146.6, 155.6, 157.4, 158.2. HRMS (ESI): *m/z* calcd for C₂₀H₂₂N₅O₂S: 380.1540; found: 380.1549.

4.1.37. 2-[(4-Benzyl-1-piperazinyl)methyl]-6-(1H-indol-3-yl)thieno[3,2-*d*]pyrimidin-4(3H)-one (2d)

This compound was synthesized using the same procedure as for **2a** starting with 3-[(4-benzyl-1-piperazinyl)acetyl]amino-5-(1H-indol-3-yl)-2-thiophenecarboxamide **48d** to afford a beige solid (44.2 mg, 0.097 mmol, 97%); mp 350 °C (dec.); IR (neat): 1659 (C=O), 2815 (NH) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 2.34–2.42 (m, 4H, 2 × CH₂), 2.46–2.53 (m, 4H, 2 × CH₂), 3.44 (s, 2H, CH₂), 3.46 (s, 2H, CH₂), 7.14–7.33 (m, 7H, 7 × CH), 7.47–7.52 (m, 1H, CH), 7.56 (s, 1H, CH), 7.92–7.98 (m, 1H, CH), 8.04 (s, 1H, CH), 8.47 (br s, 1H, NH), 12.12 (br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 52.4, 52.6, 60.2, 62.0, 109.0, 112.4, 117.3, 117.6, 119.0, 120.7, 122.2, 124.1, 125.9, 126.8, 128.1, 128.8, 136.8, 138.1, 146.4, 156.1, 157.7, 158.4. HRMS (ESI): *m/z* calcd for C₂₆H₂₆N₅O₂S: 456.1853; found: 456.1856.

4.1.38. 6-(1*H*-Indol-3-yl)-2-(1-pyrrolidinylmethyl)thieno[3,2-*d*]pyrimidin-4(3*H*)-one (**2e**)

This compound was synthesized using the same procedure as for **2a** starting with 5-(1*H*-indol-3-yl)-3-[(1-pyrrolidinylacetyl)amino]-2-thiophenecarboxamide **48e** to afford a beige solid (34.3 mg, 0.098 mmol, 98%); mp 350 °C (dec.); IR (neat): 1660 (C=O), 2817 (NH) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 1.64 (1.70 (m, 4H, 2 × CH₂), 2.47–2.55 (m, 4H, 2 × CH₂), 3.53 (s, 2H, CH₂), 7.09–7.21 (m, 2H, 2 × CH), 7.43–7.47 (m, 1H, CH), 7.52 (s, 1H, CH), 7.89–7.95 (m, 1H, CH), 8.01 (s, 1H, CH), 8.44 (br s, 1H, NH), 12.01 (br s, 1H, NH). ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 23.5, 53.7, 57.9, 109.0, 112.4, 117.2, 117.6, 119.0, 120.7, 122.2, 124.2, 125.9, 136.7, 146.5, 156.7, 157.6, 158.5. HRMS (ESI): *m/z* calcd for C₁₉H₁₈N₄OSNa: 373.1094; found: 373.1101.

4.1.39. 6-(1-Benzyl-1*H*-indol-3-yl)-2-(1-piperidinylmethyl)thieno[3,2-*d*]pyrimidin-4(3*H*)-one (**2f**)

This compound was synthesized using the same procedure as for **2a** starting with 5-(1-benzyl-1*H*-indol-3-yl)-3-[(1-piperidinylacetyl)amino]-2-thiophenecarboxamide **48f** to afford a beige solid (45.0 mg, 0.099 mmol, 99%); mp 114 °C; IR (neat): 1667 (C=O), 2930 (NH) cm⁻¹. ¹H NMR (250 MHz, CDCl₃) δ (ppm): 1.38–1.48 (m, 2H, CH₂), 1.52–1.64 (m, 4H, 2 × CH₂), 2.44–2.52 (m, 4H, 2 × CH₂), 3.48 (s, 2H, CH₂), 5.32 (s, 2H, CH₂), 7.08–7.15 (m, 2H, 2 × CH), 7.18–7.30 (m, 6H, 6 × CH), 7.37 (s, 1H, CH), 7.48 (s, 1H, CH), 7.93–8.02 (m, 1H, CH), 10.02 (br s, 1H, NH). ¹³C NMR (62.5 MHz, CDCl₃) δ (ppm): 23.6, 25.8, 50.5, 54.8, 60.8, 110.4, 110.5, 118.4, 118.8, 120.0, 121.4, 123.1, 125.7, 126.9, 127.6, 128.0, 129.0, 136.3, 137.1, 147.3, 157.4, 159.0, 169.0. HRMS (ESI): *m/z* calcd for C₂₇H₂₇N₄OS: 455.1900; found: 455.1912.

4.1.40. 6-(1-Benzyl-1*H*-indol-3-yl)-2-(4-morpholinylmethyl)thieno[3,2-*d*]pyrimidin-4(3*H*)-one (**2g**)

This compound was synthesized using the same procedure as for **2a** starting with 5-(1-benzyl-1*H*-indol-3-yl)-3-[(4-morpholinylacetyl)amino]-2-thiophenecarboxamide **48g** to afford a beige solid (44.7 mg, 0.098 mmol, 98%); mp 236 °C; IR (neat): 1661 (C=O), 2848 (NH) cm⁻¹. ¹H NMR (250 MHz, CDCl₃) δ (ppm): 2.61–2.65 (m, 4H, 2 × CH₂), 3.61 (s, 2H, CH₂), 3.77–3.81 (m, 4H, 2 × CH₂), 5.35 (s, 2H, CH₂), 7.16–7.23 (m, 2H, 2 × CH), 7.28–7.37 (m, 6H, 6 × CH), 7.44 (s, 1H, CH), 7.56 (s, 1H, CH), 8.01–8.06 (m, 1H, CH), 10.07 (br s, 1H, NH). ¹³C NMR (62.5 MHz, CDCl₃) δ (ppm): 50.5, 53.6, 60.7, 66.8, 110.3, 110.5, 118.4, 118.7, 120.0, 121.5, 123.1, 125.6, 126.9, 127.6, 128.1, 129.0, 136.3, 137.1, 147.6, 155.0, 157.4, 158.9. HRMS (ESI): *m/z* calcd for C₂₈H₂₅N₄O₂S: 457.1693; found: 457.1676.

4.1.41. 6-(1-Benzyl-1*H*-indol-3-yl)-2-[(4-methyl-1-piperazinyl)methyl]thieno[3,2-*d*]pyrimidin-4(3*H*)-one (**2h**)

This compound was synthesized using the same procedure as for **2a** starting with 5-(1-benzyl-1*H*-indol-3-yl)-3-[(4-methyl-1-piperazinyl)acetyl]amino]-2-thiophenecarboxamide **48h** to afford a beige solid (46.5 mg, 0.099 mmol, 99%); mp 132 °C; IR (neat): 1660 (C=O), 2796 (NH) cm⁻¹. ¹H NMR (250 MHz, CDCl₃) δ (ppm): 2.33 (s, 3H, CH₃), 2.48–2.58 (m, 4H, 2 × CH₂), 2.63–2.71 (m, 4H, 2 × CH₂), 3.61 (s, 2H, CH₂), 5.38 (s, 2H, CH₂), 7.15–7.21 (m, 2H, 2 × CH), 7.24–7.38 (m, 6H, 6 × CH), 7.44 (s, 1H, CH), 7.55 (s, 1H, CH), 8.00–8.09 (m, 1H, CH), 10.01 (s, 1H, NH). ¹³C NMR (62.5 MHz, CDCl₃) δ (ppm): 45.9, 50.5, 53.5, 54.9, 60.2, 110.4, 110.5, 118.4, 118.7, 120.0, 121.4, 123.1, 125.7, 126.9, 127.6, 128.1, 129.0, 136.3, 137.1, 147.4, 155.4, 157.4, 159.0. HRMS (ESI): *m/z* calcd for C₂₇H₂₈N₅O₂S: 470.2009; found: 470.2008.

4.1.42. 6-(1-Benzyl-1*H*-indol-3-yl)-2-[(4-benzyl-1-piperazinyl)methyl]thieno[3,2-*d*]pyrimidin-4(3*H*)-one (**2i**)

This compound was synthesized using the same procedure as for **2a** starting with 5-(1-benzyl-1*H*-indol-3-yl)-3-[(4-benzyl-1-piperazinyl)acetyl]amino]-2-thiophenecarboxamide **48i** to afford

a beige solid (53.5 mg, 0.098 mmol, 98%); mp 350 °C (dec.); IR (neat): 1667 (C=O), 2814 (NH) cm⁻¹. ¹H NMR (250 MHz, CDCl₃) δ (ppm): 2.51–2.59 (m, 4H, 2 × CH₂), 2.62–2.69 (m, 4H, 2 × CH₂), 3.56 (s, 2H, CH₂), 3.61 (s, 2H, CH₂), 5.39 (s, 2H, CH₂), 7.15–7.22 (m, 2H, 2 × CH), 7.24–7.38 (m, 11H, 11 × CH), 7.43 (s, 1H, CH), 7.55 (s, 1H, CH), 8.03–8.07 (m, 1H, CH), 10.04 (s, 1H, NH). ¹³C NMR (62.5 MHz, CDCl₃) δ (ppm): 50.5, 52.9, 53.4, 60.2, 62.9, 110.4, 110.5, 118.4, 118.7, 120.0, 121.4, 123.1, 125.7, 126.9, 127.2, 127.6, 128.1, 128.3, 129.0, 129.2, 136.3, 137.1, 137.9, 147.4, 155.5, 157.4, 159.0. HRMS (ESI): *m/z* calcd for C₃₃H₃₂N₅O₂S: 546.2322; found: 546.2331.

4.1.43. 6-(1-Benzyl-1*H*-indol-3-yl)-2-(1-pyrrolidinylmethyl)thieno[3,2-*d*]pyrimidin-4(3*H*)-one (**2j**)

This compound was synthesized using the same procedure as for **2a** starting with 5-(1-benzyl-1*H*-indol-3-yl)-3-[(1-pyrrolidinylacetyl)amino]-2-thiophenecarboxamide **48j** to afford a beige solid (43.6 mg, 0.099 mmol, 99%); mp 212 °C; IR (neat): 1661 (C=O), 2809 (NH) cm⁻¹. ¹H NMR (250 MHz, CDCl₃) δ (ppm): 1.85–1.90 (m, 4H, 2 × CH₂), 2.66–2.73 (m, 4H, 2 × CH₂), 3.75 (s, 2H, CH₂), 5.38 (s, 2H, CH₂), 7.15–7.21 (m, 2H, 2 × CH), 7.26–7.36 (m, 6H, 6 × CH), 7.44 (s, 1H, CH), 7.55 (s, 1H, CH), 8.00–8.07 (m, 1H, CH), 8.43 (br s, 1H, NH). ¹³C NMR (62.5 MHz, CDCl₃) δ (ppm): 24.0, 50.5, 54.3, 57.4, 110.4, 110.5, 118.4, 118.6, 120.0, 121.4, 123.1, 125.7, 126.9, 127.6, 128.0, 129.0, 136.3, 137.1, 147.3, 156.3, 157.5, 159.1. HRMS (ESI): *m/z* calcd for C₂₆H₂₅N₄O₂S: 441.1744; found: 441.1734.

4.1.44. 6-(1-Benzyl-1*H*-indol-3-yl)-2-hexylthieno[3,2-*d*]pyrimidin-4-ol (**3a**)

To a solution of 3-amino-5-(1-benzyl-1*H*-indol-3-yl)-2-thiophenecarboxamide **43f** (347 mg, 1 mmol) in MeOH containing 6% of concentrated HCl, was added heptaldehyde (209 μL, 1.5 mmol). The reaction mixture was heated at reflux for 24 h. After cooling, the precipitate formed was filtered and washed once with cold MeOH to afford a yellow solid (93 mg, 0.21 mmol, 21%); mp 236 °C; IR (KBr): 2424 (OH) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 0.83–0.87 (m, 3H, CH₃), 1.25–1.32 (m, 6H, 3 × CH₂), 1.71–1.74 (m, 2H, CH₂), 2.67–2.73 (t, 2H, CH₂, *J* = 7.43 Hz), 5.50 (s, 2H, CH₂), 7.21–7.29 (m, 7H, 7 × CH), 7.57–7.60 (m, 2H, 2 × CH), 7.95–7.98 (m, 1H, CH), 8.35 (s, 1H, CH). Proton of hydroxyl group exchange with deuterium of NMR solvent. ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 13.9, 21.9, 27.0, 28.1, 30.8, 33.2, 48.5, 49.4, 108.6, 111.3, 116.0, 116.6, 119.4, 121.3, 122.6, 124.7, 127.2, 127.6, 128.6, 129.7, 136.5, 137.33, 146.7, 157.0, 160.5. HRMS (ESI): *m/z* calcd for C₂₇H₂₈N₃O₂S: 442.1948; found: 442.1963.

4.1.45. 6-(1-Benzyl-1*H*-indol-3-yl)-2-(4-methoxyphenyl)thieno[3,2-*d*]pyrimidin-4-ol (**3b**)

This compound was synthesized using the same procedure as for **3a** starting with 3-amino-5-(1-benzyl-1*H*-indol-3-yl)-2-thiophenecarboxamide **43f** and *p*-anisaldehyde to give a yellow solid (144 mg, 0.31 mmol, 31%); mp 278 °C; IR (KBr): 2518 (OH) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 3.85 (s, 3H, CH₃), 5.51 (s, 2H, CH₂), 7.09 (d, 2H, 2 × CH, *J* = 8.93 Hz), 7.22–7.33 (m, 7H, 7 × CH), 7.56–7.60 (m, 1H, CH), 7.65 (s, 1H, CH), 8.00–8.04 (m, 1H, CH), 8.15 (d, 2 × CH, *J* = 8.93 Hz), 8.32 (s, 1H, CH). Proton of hydroxyl group exchange with deuterium of NMR solvent. ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 49.4, 55.5, 108.8, 111.3, 114.0, 116.6, 117.7, 119.5, 121.2, 122.6, 124.1, 124.8, 127.2, 127.6, 128.6, 129.4, 129.6, 136.5, 137.4, 146.2, 154.3, 157.86, 157.93, 161.9. HRMS (ESI): *m/z* calcd for C₂₈H₂₂N₃O₂S: 464.1427; found: 464.1425.

4.1.46. 6-(1-Benzyl-1*H*-indol-3-yl)-2-(3,4,5-trimethoxyphenyl)thieno[3,2-*d*]pyrimidin-4-ol (**3c**)

This compound was synthesized using the same procedure as for **3a** starting with 3-amino-5-(1-benzyl-1*H*-indol-3-yl)-2-

thiophenecarboxamide **43f** and 3,4,5-trimethoxybenzaldehyde to provide a yellow solid (188 mg, 0.36 mmol, 36%); mp 289 °C; IR (KBr): 2939 (OH) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 3.74 (s, 3H, CH₃), 3.89 (s, 6H, 2 × CH₃), 5.51 (s, 2H, CH₂), 7.22–7.33 (m, 7H, 7 × CH), 7.54 (s, 2H, 2 × CH), 7.57–7.61 (m, 1H, CH), 7.69 (s, 1H, CH), 8.01–8.06 (m, 1H, CH), 8.32 (s, 1H, CH). Proton of hydroxyl group exchange with deuterium of NMR solvent. ¹³C NMR (62.5 MHz, DMSO-*d*₆) δ (ppm): 49.4, 56.1, 60.1, 105.1, 108.9, 111.3, 117.0, 118.3, 119.6, 121.2, 122.7, 124.8, 127.2, 127.5, 127.6, 128.6, 136.5, 137.4, 140.1, 146.1, 152.8, 153.9, 158.0, 158.6. HRMS (ESI): *m/z* calcd for C₃₀H₂₆N₃O₄S: 524.1639; found: 524.1636.

4.1.47. 6-(1-Benzyl-1*H*-indol-3-yl)-2-(4-chlorophenyl)thieno[3,2-*d*]pyrimidin-4(3*H*)-one (**3d**)

This compound was synthesized using the same procedure as for **3a** starting with 3-amino-5-(1-benzyl-1*H*-indol-3-yl)-2-thiophenecarboxamide **43f** and *p*-chlorobenzaldehyde to afford a brown solid (56 mg, 0.12 mmol, 12%); mp 314 °C; IR (KBr): 1658 (C=O) cm⁻¹. ¹H NMR (250 MHz, DMSO-*d*₆) δ (ppm): 5.51 (s, 2H, CH₂), 7.22–7.33 (m, 7H, 7 × CH), 7.55–7.67 (m, 5H, 5 × CH), 8.00–8.06 (m, 1H, CH), 8.16 (d, 1H, CH, *J* = 8.65 Hz), 8.32 (s, 1H, CH), 12.72 (br s, 1H, NH). HRMS (APCI): *m/z* calcd for C₂₇H₁₉ClN₃O₂S: 468.0932; found: 468.0939.

4.2. Molecular modeling, docking and 3-D QSAR predictions

All computational details will be detailed elsewhere [33]. Briefly all the molecules listed in Tables 1–3, 5 were modeled using the chemaxon msketch module (<http://www.chemaxon.com>) and directly used for the docking studies. AutoDock version 4.2 [45] was used and cross-docking experiments were carried out similarly as described by Musmuca [46]. The 3-D QSAR model was carried out using an in house procedure [47]. Nineteen VEGFR-2 co-crystallized complexes were retrieved from the PDB (www.rcsb.org), cleaned from any solvents and ions residues. The tyrosine phosphate non standard residue was kept in the structures. The complexes were subjected to a previous reported procedure for geometry optimization [46]. The minimized complex was aligned using Chimera 1.5.2 [48] and the ligands were then extracted obtaining the SB alignment. The AutoGrid module of the AutoDock suite was used to calculate the molecular interaction fields (MIFs) on the aligned molecules. The ligand MIFs and the corresponding activities were submitted to an R [49] script which performed PLS and cross-validations to assess the 3-D QSAR model.

SB alignment assessment was performed though extensive re-docking and cross-docking experiments on the cleaned and minimized co-crystals. The modeled compounds **1–38** and the intermediates **44b**, **45b** and **48a–j** were then cross-docked. The built 3-D QSAR model was applied to lowest energy docked poses and the corresponding pIC₅₀ were predicted. The binding modes of the **1–38**, **44b**, **45b** and **48a–j** were analyzed by the mean of the Chimera software and the interaction profiles were obtained with the Ligplot software [50].

4.3. Biology. Materials and methods

Human recombinant protein tyrosine kinase VEGFR-2 and Omnia™ Tyr Peptide Kit 7 were from Invitrogen. Vandetanib, exploited as the reference standard, was from Sequoia research product Ltd.

4.3.1. Tyrosine kinase assays

Assays were performed in 96-well microtiter plates using the Omnia Tyr Peptide 7 Kit, according to the manufacturer's protocol and following a previously reported procedure, standardized for

EGFR [51]. Briefly, the kinase activity was determined fluorimetrically by monitoring the increase in fluorescence resulting from phosphorylation of the peptide substrate, carrying the fluorophore 8-hydroxy-5-(*N,N*-dimethylsulfonamido)-2-methylquinoline, catalyzed by VEGFR-2 in the presence of ATP.

Assays were carried out at 30 °C in a reaction mixture containing 5 μL of Tyrosine Kinase Reaction Buffer, 5 μL of Tyrosine Kinase Substrate, 5 μL of 1 mM ATP, 5 μL of 1 mM DTT, 25 μL of ultrapure water and 5 μL of 3 mU/μL VEGFR-2, in a total volume of 50 μL. All the above reagents, except VEGFR-2, were incubated at 30 °C for 5 min. VEGFR-2 was then added to start the reaction, which was monitored with the fluorescence meter Victor3™ Perkin Elmer at 360 nm (excitation filter) and 485 nm (emission filter). Kinase activity was calculated from a linear least-squares fit of the data for fluorescence intensity versus time.

4.3.2. Enzymatic inhibition

The inhibitory activity of titled compounds against VEGFR-2 was assayed by adding 5 μL of the inhibitor solution to the reaction mixture described above. All the products were dissolved in 100% DMSO and diluted to the appropriate concentrations with Tyrosine Kinase Reaction Buffer, provided by the kit. Final concentration of DMSO in assay solutions never exceed 1%, and proved to have no effects on protein activity. The inhibitory effect of the new derivatives was routinely estimated at a concentration of 200 μM. Results are expressed as means ± SEM of percentage inhibition values, obtained through two determinations carried out in triplicate (Tables 1–3, 5). Those compounds found to be active were then tested at additional concentrations between 200 μM and 20 nM. For a proper comparison Vandetanib was employed as the reference standard. The determination of the IC₅₀ values was performed by linear regression analysis of the log–dose response curve, which was generated using at least five concentrations of the inhibitor causing an inhibition between 20% and 80%, with three replicates at each concentration. Results are expressed as means ± SEM (Table 4 compounds **2f** and **3d**). GraphPad 5.0 software was used for the statistical analysis.

4.3.3. Endothelial cell culture

Umbilical cords were cut after delivery in compliance with relevant laws in respect with consent of women. Rapidly, human umbilical vein endothelial cells (HUVEC) were collected from umbilical cords as previously described by Jaffe et al., 1973 [51]. HUVEC were cultured and used until passage 5 in "HUVEC/complete medium" consisting in 50% M199 (GibcoBRL, France) and 50% RPMI 1640 (v/v) (Sigma, France) supplemented with 20% heat inactivated human AB serum (EFS, Nancy, France), 2 mM L-glutamine (GibcoBRL), 100 U/mL penicillin (GibcoBRL, France), 100 μg/mL streptomycin (GibcoBRL, France), 2.5 μg/mL amphotericin B (GibcoBRL, France), and 20 mM HEPES (Sigma, France).

4.3.4. Effect of **2f tartaric** on metabolic activity and viability of endothelial cells

As previously described [52], HUVEC were plated at 20,000 cells/cm² in 24 well plates in complete medium. After 24 h, the medium was changed and cells were cultured with **2f tartaric** (0–100 μM) in RPMI 1640 medium supplemented with 2% heat inactivated fetal calf serum (SVF) during 24 h.

Metabolic activity (MTT assay): cells were incubated with 0.4 mg/mL of 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyl tetrazolium bromide (MTT) (Sigma, France) at 37 °C during 3 h. Formazan crystals were dissolved by dimethylsulfoxide (Fisher Scientific, France) and absorbance was measured at 570/630 nm (EL800 Universal microplate reader, Bio-Tek instrument, USA). Results are presented as percent of variation with values obtained for control

medium without **2f-tartaric** as reference (Means \pm SD, $n = 3$ in triplicate).

Viability (Hoechst assay): cells were lysed by freezing at -80°C . Cell lysates were incubated with $2\ \mu\text{g}$ of Hoechst 33258 (Sigma, France) during 1 h under agitation and the fluorescence intensity was measured (λ excitation: $360\ \text{nm}$ / λ emission: $460\ \text{nm}$, Berthold Twinkle LB970 system). Results are presented as percent of variation with values obtained for control medium without **2f-tartaric** as reference (Means \pm SD, $n = 3$ in triplicate).

Effect of DMSO on the viability of endothelial cells was evaluated after dilution in RPMI 1640 medium supplemented with 2% SVF.

4.3.5. Endothelial cell tube formation

As previously described [52], HUVEC were plated (90,000 cells/ cm^2) onto 24-well plate pre-coated with Matrigel[®] (BD Biosciences, France). After 1 h, media were removed and replaced by **2f-tartaric** ($0-3\ \mu\text{M}$) in RPMI 1640 medium supplemented with 2% heat inactivated fetal calf serum and 50 ng VEGF during 24 h before HUVEC being fixed with 4% paraformaldehyde (Sigma, France). Photomicrographs (Nikon AZ100, Digital Sight DS-Qi1Mc camera, Nikon, France) were taken after phalloidin-sulforhodamine staining (Fluoprobes[®], Interchim, France). Area fraction of endothelial cells was calculated with NIS Element Software (Nikon, France) and represented as the area fraction of cells in relation to whole field. It was expressed as a percentage of variation with values obtained for control medium without VEGF and without **2f-tartaric** as reference (negative control) (Means \pm SEM, $n = 5$). For comparison purpose, Sunitinib (LC Laboratories, USA) was used as reference drug using the same protocol.

Funding sources

This study was supported by grants from the French 'Ligue Nationale contre le Cancer' and the 'Region Lorraine', a PhD grant to E.P. from the French 'Ministère de l'Enseignement Supérieur et de la Recherche' and from Italian Ministry of University and Research (MIUR Grant 2008F8T894_002 and 2008ZTN724_003, PRIN Grant 20105YY2HL). One of us (F.B.) acknowledge SapienzaUniversità di Roma (grant "Progetti per Avvio alla Ricerca" prot. C26N12JZCT).

Acknowledgments

The authors thank the staff of obstetric department of Nancy maternity hospital and pregnant women for providing umbilical cords.

Appendix A. Supplementary material

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.ejmech.2013.03.022>.

References

- [1] D. Ribatti, A. Vacca, B. Nico, L. Roncali, F. Dammacco, Postnatal vasculogenesis, *Mech. Dev.* 100 (2001) 157–163.
- [2] W. Risau, Mechanisms of angiogenesis, *Nature* 386 (1997) 671–674.
- [3] D.A. Walsh, L. Haywood, Angiogenesis: a therapeutic target in arthritis, *Curr. Opin. Invest. Drugs* 2 (2001) 1054–1063.
- [4] G.A. Fava, Affective disorders and endocrine disease. New insights from psychosomatic studies, *Psychosomatics* 35 (1994) 341–353.
- [5] L.P. Aiello, R.L. Avery, P.G. Arrigg, B.A. Keyt, H.D. Jampel, S.T. Shah, L.R. Pasquale, H. Thieme, M.A. Iwamoto, J.E. Park, et al., Vascular endothelial growth factor in ocular fluid of patients with diabetic retinopathy and other retinal disorders, *N. Engl. J. Med.* 331 (1994) 1480–1487.
- [6] M. Detmar, The role of VEGF and thrombospondins in skin angiogenesis, *J. Dermatol. Sci.* 24 (Suppl. 1) (2000) S78–S84.
- [7] J. Folkman, Anti-angiogenesis: new concept for therapy of solid tumors, *Ann. Surg.* 175 (1972) 409–416.
- [8] L.A. Liotta, P.S. Steeg, W.G. Stetler-Stevenson, Cancer metastasis and angiogenesis: an imbalance of positive and negative regulation, *Cell* 64 (1991) 327–336.
- [9] N. Ferrara, VEGF and the quest for tumour angiogenesis factors, *Nat. Rev. Cancer* 2 (2002) 795–803.
- [10] N. Ferrara, R.S. Kerbel, Angiogenesis as a therapeutic target, *Nature* 438 (2005) 967–974.
- [11] R.S. Kerbel, Tumor angiogenesis, *N. Engl. J. Med.* 358 (2008) 2039–2049.
- [12] N. Ferrara, H.P. Gerber, J. LeCouter, The biology of VEGF and its receptors, *Nat. Med.* 9 (2003) 669–676.
- [13] N. Almog, L. Ma, R. Raychowdhury, C. Schwager, R. Erber, S. Short, L. Hlatky, P. Vajkoczy, P.E. Huber, J. Folkman, A. Abdollahi, Transcriptional switch of dormant tumors to fast-growing angiogenic phenotype, *Cancer Res.* 69 (2009) 836–844.
- [14] M.A. Gimbrone Jr., S.B. Leapman, R.S. Cotran, J. Folkman, Tumor dormancy in vivo by prevention of neovascularization, *J. Exp. Med.* 136 (1972) 261–276.
- [15] S. Baka, A.R. Clamp, G.C. Jayson, A review of the latest clinical compounds to inhibit VEGF in pathological angiogenesis, *Expert Opin. Ther. Targets* 10 (2006) 867–876.
- [16] B.M. Klebl, G. Muller, Second-generation kinase inhibitors, *Expert Opin. Ther. Targets* 9 (2005) 975–993.
- [17] L. Sepp-Lorenzino, K.A. Thomas, Antiangiogenic agents targeting vascular endothelial growth factor and its receptors in clinical development, *Expert Opin. Invest. Drugs* 11 (2002) 1447–1465.
- [18] C.T. Supuran, A. Scozzafava, Protein tyrosine kinase inhibitors as anticancer agents, *Expert Opin. Ther. Patents* 14 (2004) 35–53.
- [19] P. Carmeliet, L. Moons, A. Luttun, V. Vincenzi, M. Compennolle, M. De Mol, Y. Wu, F. Bon, L. Devy, H. Beck, D. Scholz, T. Acker, T. DiPalma, M. Dewerschin, A. Noel, I. Stalmans, A. Barra, S. Blacher, T. Vandendriessche, A. Ponten, U. Eriksson, K.H. Plate, J.M. Foidart, W. Schaper, D.S. Charnock-Jones, D.J. Hicklin, J.M. Herbert, D. Collen, M.G. Persico, Synergism between vascular endothelial growth factor and placental growth factor contributes to angiogenesis and plasma extravasation in pathological conditions, *Nat. Med.* 7 (2001) 575–583.
- [20] N. Ferrara, K.J. Hillan, H.P. Gerber, W. Novotny, Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer, *Nat. Rev. Drug Discov.* 3 (2004) 391–400.
- [21] S.M. Wilhelm, C. Carter, L.Y. Tang, D. Wilkie, A. McNabola, H. Rong, C. Chen, X.M. Zhang, P. Vincent, M. McHugh, Y.C. Cao, J. Shujath, S. Gawlak, D. Eveleigh, B. Rowley, L. Liu, L. Adnane, M. Lynch, D. Auclair, I. Taylor, R. Gedrich, A. Voznesensky, B. Riedl, L.E. Post, G. Bollag, P.A. Trail, BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis, *Cancer Res.* 64 (2004) 7099–7109.
- [22] R.J. Motzer, M.D. Michaelson, B.G. Redman, G.R. Hudes, G. Wilding, R.A. Figlin, M.S. Ginsberg, S.T. Kim, C.M. Baum, S.E. DePrimo, J.Z. Li, C.L. Bello, C.P. Thuer, D.J. George, B.I. Rini, Activity of SU11248, a multitargeted inhibitor of vascular endothelial growth factor receptor and platelet-derived growth factor receptor, in patients with metastatic renal cell carcinoma, *J. Clin. Oncol.* 24 (2006) 16–24.
- [23] S. Faivre, C. Delbaldo, K. Vera, C. Robert, S. Lozahic, N. Lassau, C. Bello, S. DePrimo, N. Brega, G. Massimini, J.P. Armand, P. Scigalla, E. Raymond, Safety, pharmacokinetic, and antitumor activity of SU11248, a novel oral multitarget tyrosine kinase inhibitor, in patients with cancer, *J. Clin. Oncol.* 24 (2006) 25–35.
- [24] D. Strumberg, H. Richly, R.A. Hilger, N. Schleucher, S. Korfee, M. Tewes, M. Faghii, E. Brendel, D. Voliotis, C.G. Haase, B. Schwartz, A. Awada, R. Voigtman, M.E. Scheulen, S. Seeber, Phase I clinical and pharmacokinetic study of the Novel Raf kinase and vascular endothelial growth factor receptor inhibitor BAY 43-9006 in patients with advanced refractory solid tumors, *J. Clin. Oncol.* 23 (2005) 965–972.
- [25] R. Rahman, S. Smith, C. Rahman, R. Grundy, Antiangiogenic therapy and mechanisms of tumor resistance in malignant glioma, *J. Oncol.* 2010 (2010) 251231.
- [26] J. Robert, L'angiogenèse, Application Médicale du GENE Collection, John Libbey Eurotext, 2009.
- [27] A. Garofalo, L. Goossens, P. Six, A. Lemoine, S. Ravez, A. Farce, P. Depreux, Impact of aryloxy-linked quinazolines: a novel series of selective VEGFR-2 receptor tyrosine kinase inhibitors, *Bioorg. Med. Chem. Lett.* 21 (2011) 2106–2112.
- [28] E. Hu, A. Tasker, R.D. White, R.K. Kunz, J. Human, N. Chen, R. Buerli, R. Hungeate, P. Novak, A. Itano, X.X. Zhang, V. Yu, Y. Nguyen, Y. Tudor, M. Plant, S. Flynn, Y. Xu, K.L. Meagher, D.A. Whittington, G.Y. Ng, Discovery of aryl amino-quinazoline pyridones as potent selective, and orally efficacious inhibitors of receptor tyrosine kinase c-Kit, *J. Med. Chem.* 51 (2008) 3065–3068.
- [29] D.S. La, J. Belzile, J.V. Bready, A. Coxon, T. DeMelfi, N. Doerr, J. Estrada, J.C. Flynn, S.R. Flynn, R.F. Graceffa, S.P. Harriman, J.F. Larrow, A.M. Long, M.W. Martin, M.J. Morrison, V.F. Patel, P.M. Roveto, L. Wang, M.M. Weiss, D.A. Whittington, Y. Teffera, Z. Zhao, A.J. Polverino, J.C. Harnange, Novel 2,3-dihydro-1,4-benzoxazines as potent and orally bioavailable inhibitors of tumor-driven angiogenesis, *J. Med. Chem.* 51 (2008) 1695–1705.
- [30] Y. Miyazaki, S. Matsunaga, J. Tang, Y. Maeda, M. Nakano, R.J. Philippe, M. Shibahara, W. Liu, H. Sato, L. Wang, R.T. Nolte, Novel 4-amino-furo[2,3-d]

- pyrimidines as Tie-2 and VEGFR2 dual inhibitors, *Bioorg. Med. Chem. Lett.* 15 (2005) 2203–2207.
- [31] B.I. Hodocs, S.D. Geuns-Meyer, P.E. Hughes, B.K. Albrecht, S. Bellon, J. Bready, S. Caenepeel, V.J. Gee, S.C. Chaffee, A. Coxon, M. Emery, J. Fretland, P. Gallant, Y. Gu, D. Hoffman, R.E. Johnson, R. Kendall, J.L. Kim, A.M. Long, M. Morrison, P.R. Olivieri, V.F. Patel, A. Polverino, P. Rose, P. Tempest, L. Wang, D.A. Whittington, H. Zhao, Evolution of a highly selective and potent 2-(pyridin-2-yl)-1,3,5-triazine Tie-2 kinase inhibitor, *J. Med. Chem.* 50 (2007) 611–626.
- [32] P.A. Harris, A. Boloor, M. Cheung, R. Kumar, R.M. Crosby, R.G. Davis-Ward, A.H. Epperly, K.W. Hinkle, R.N. Hunter 3rd, J.H. Johnson, V.B. Knick, C.P. Laudeman, D.K. Luttrell, R.A. Mook, R.I. Nolte, S.K. Rudolph, J.R. Szewczyk, A.T. Truesdale, J.M. Veal, L. Wang, J.A. Stafford, Discovery of 5-[[4-[(2,3-dimethyl 2H indazol-6-yl)methylamino]-2-pyrimidinyl]amino]-2-methylbenzenesulfonamide (Pazopanib), a novel and potent vascular endothelial growth factor receptor inhibitor, *J. Med. Chem.* 51 (2008) 4632–4640.
- [33] R. Ragno, VEGFR 2 inhibitors. Ligand Based, structure based and 3 D QSAR studies as tools to design new small molecules, in: XXIII Congresso Nazionale della Società Chimica Italiana, Sorrento, Italy, 2009.
- [34] S. Hesse, E. Perspicace, G. Kirsch, Microwave-assisted synthesis of 2-aminothiophene-3-carboxylic acid derivatives, 3H-thieno[2,3-d]pyrimidin-4-one and 4-chlorothieno[2,3-d]pyrimidine, *Tetrahedron Lett.* 48 (2007) 5261–5264.
- [35] E. Perspicace, S. Hesse, G. Kirsch, M. Yemlou, C. Leconte, Unexpected C–O bond formation in Suzuki coupling of 4-chlorothieno[2,3-d]pyrimidines, *J. Heterocycl. Chem.* 46 (2009) 459–464.
- [36] D. Thoma, E. Perspicace, S. Hesse, G. Kirsch, P. Seck, Synthesis of substituted [1,3]thiazolo[4,5-b]pyridines and [1,3]thiazolo[4,5-d][1,2,3]triazines, *Tetrahedron* 64 (2008) 9309–9314.
- [37] E. Perspicace, D. Thoma, G. Hamm, S. Hesse, G. Kirsch, P. Seck, Synthesis of substituted selenolo[3,2-d][1,2,3]triazines and [1,3]selenazolo[4,5-d][1,2,3]triazines, *Synthesis* (2009) 3472–3476.
- [38] M.D. Shults, B. Imperiali, Versatile fluorescence probes of protein kinase activity, *J. Am. Chem. Soc.* 125 (2003) 14248–14249.
- [39] E. Migiani, G. Kirsch, Synthesis of new thieno[6]azepinediones from α -methylene ketones, *Synthesis* (2002) 1096–1100.
- [40] H. Hartmann, J. Liebscher, A simple method or the synthesis of 5-aryl-3-amino-2-alkoxy-carbonylthiophenes, *Synthesis-stuttgart* (1984) 275–276.
- [41] S. Kim, K.Y. Yi, Synthetic applications of di-2-pyridyl thionocarbonate as a dehydration, a dehydrosulfurization, and a thiocarbonyl transfer reagent, *Bull. Korean Chem. Soc.* 8 (1987) 466–470.
- [42] M.I. Crespo, L. Pages, A. Vega, V. Segarra, M. Lopez, T. Domenech, M. Miralpeix, J. Beleta, H. Ryder, J.M. Palacios, Design, synthesis, and biological activities of new thieno 3,2-d pyrimidines as selective type 4 phosphodiesterase inhibitors, *J. Med. Chem.* 41 (1998) 4021–4035.
- [43] O. Ottoni, R. Cruz, R. Alves, Efficient and simple methods for the introduction of the sulfonyl, acyl and alkyl protecting groups on the nitrogen of indole and its derivatives, *Tetrahedron* 54 (1998) 13915–13928.
- [44] V.O. Ili, Phase transfer-catalyzed N-sulfonation of indole, *Synthesis* (1979) 136.
- [45] G.M. Morris, R. Huey, A.J. Olson, Using AutoDock for Ligand–receptor Docking, *Curr. Protoc. Bioinformatics* (2008) Chapter 8, Unit 8.14.
- [46] I. Musmuca, A. Caroli, A. Mai, N. Kaushik-Basu, P. Arora, R. Ragno, Combining 3 D quantitative structure activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors, *J. Chem. Inf. Model.* 50 (2010) 662–676.
- [47] F. Ballante, R. Ragno, 3-D QSAutoGrid/R: an alternative procedure to build 3-D QSAR models. methodologies and applications, *J. Chem. Inf. Model.* 52 (2012) 1674–1685.
- [48] Z. Yang, K. Lasker, D. Schneidman-Dubovny, B. Webb, C.C. Huang, E.F. Pettersen, T.D. Goddard, E.C. Meng, A. Sali, T.E. Ferrin, UCSF Chimera, MODELLER, and IMP: an integrated modeling system, *J. Struct. Biol.* 179 (2011) 259–278.
- [49] R.D.C. Team, The R Foundation for statistical computing, (2002).
- [50] A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions, *Protein Eng.* 8 (1995) 127–134.
- [51] E.A. Jaffe, R.I. Nachman, C.G. Becker, C.R. Minick, Culture of human endothelial cells derived from umbilical veins. Identification by morphologic and immunologic criteria, *J. Clin. Invest.* 52 (1973) 2745–2756.
- [52] V. Jouan-Hureau, C. Boura, J.-L. Merlin, B. Faivre, Modulation of endothelial cell network formation in vitro by molecular signaling of head and neck squamous cell carcinoma (HNSCC) exposed to cetuximab, *Microvasc. Res.* 83 (2012) 131–137.

Pharmacophore Assessment Through 3-D QSAR: Evaluation of the Predictive Ability on New Derivatives by the Application on a Series of Antitubercular Agents

Laura Friggeri,^{§,†} Flavio Ballante,^{*,§,‡} Rino Ragno,^{*,‡} Ira Musmuca,[‡] Daniela De Vita,[†] Fabrizio Manetti,[○] Mariangela Biava,[†] Luigi Scipione,[†] Roberto Di Santo,^{‡,†} Roberta Costi,^{‡,†} Marta Feroci,^{||} and Silvano Tortorella[†]

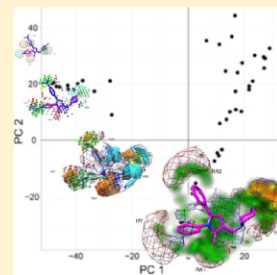
[‡]Rome Center for Molecular Design and [†]Istituto Pasteur-Fondazione Cenci Bolognetti, [†]Dipartimento di Chimica e Tecnologie del Farmaco, Sapienza Università di Roma, P. le A. Moro 5, 00185 Rome, Italy

[○]Dipartimento di Biotecnologie, Chimica e Farmacia, Università degli Studi di Siena, Via Aldo Moro 2, I-53100 Siena, Italy

^{||}Dipartimento di Scienze di Base e Applicate per l'Ingegneria, Sapienza Università di Roma, Via Castro Laurenziano 7, I-00161 Rome, Italy

Supporting Information

ABSTRACT: Pharmacophoric mapping is a useful procedure to frame, especially when crystallographic receptor structures are unavailable as in ligand-based studies, the hypothetical site of interaction. In this study, 71 pyrrole derivatives active against *M. tuberculosis* were used to derive through a recent new 3-D QSAR protocol, 3-D QSAutogrid/R, several predictive 3-D QSAR models on compounds aligned by a previously reported pharmacophoric application. A final multiprobe (MP) 3-D QSAR model was then obtained configuring itself as a tool to derive pharmacophoric quantitative models. To stress the applicability of the described models, an external test set of unrelated and newly synthesized series of R-4-amino-3-isoxazolidinone derivatives found to be active at micromolar level against *M. tuberculosis* was used, and the predicted bioactivities were in good agreement with the experimental values. The 3-D QSAutogrid/R procedure proved to be able to correlate by a single multi-informative scenario the different activity molecular profiles thus confirming its usefulness in the rational drug design approach.



1. INTRODUCTION

Tuberculosis (TB), an infectious disease mainly caused by *Mycobacterium tuberculosis* (MTB), remains a major public health problem and causes ill-health among millions of people each year. TB ranks as the second leading cause of death from an infectious disease worldwide, after the human immunodeficiency virus (HIV-1). The 2012WHO Global tuberculosis report estimates there are almost 9 million new cases and 1.4 million TB deaths.¹ Moreover, two billion people are estimated to be latently infected with MTB, and 10% of them reactivate to active TB with major risk relative to immigrants from endemic areas, people with HIV-1 infection, and individuals with underlying diseases (silicosis, diabetes mellitus, malignant conditions).²

Currently, the standard treatment comprises: first line drugs, such as isoniazid (INH), pyrazinamide (PZA), ethambutol (EMB), and rifampin (RIF); and second line drugs, such as ethionamide (ETH), *p*-aminosalicylic acid (PAS), capreomycin, aminoglycosides, *D*-cycloserine (DCS), and fluoroquinolones.³

The required long-term drug treatment, due to the high persistence of MTB, combined with poor compliance of the patients, highly contributes to developing drug resistant strains, particularly multidrug-resistant (MDR, resistant at least to INH and RIF) and extensively drug-resistant (XDR, resistant at least

to INH, RIF, and three of second line class of anti-TB drugs). Recently a more dangerous form of bacilli, named totally drug-resistant (TDR) showing in vitro resistance to all first- and second-line drugs tested have been isolated.^{4,5}

To reduce this increasing problem, antitubercular drugs are used with specific therapeutic protocols under direct observation therapy short course (DOTS) conditions.⁶

The need for new shorter therapeutic regimens and new classes of drugs active on MDR, XDR, and TDR MTB drives pharmaceutical research to accelerate in the development process of new anti-TB drugs.⁷ Continuing our research on anti-TB agents,^{8,9} here we report the assessment of a previously reported pharmacophore model¹⁰ through 3-D QSAutogrid/R, a recent introduced quantitative ligand-based design protocol.¹¹ The developed 3-D QSAR models were tested for their predictive ability on a series of new independently synthesized R-4-amino-3-isoxazolidinone derivatives 1a–e, 2a–f, and 3h,i (Table 1). These compounds have been designed to evaluate the effects on antitubercular activity due by the introduction of acyl substituents on N(2) atom of oxoisoxazolidine ring and on amino group.

Received: March 1, 2013

Published: April 25, 2013

Table 1. R-4-Amino-3-isoxazolidinone Derivatives: Monocarbamates (1a–e), Dicarbamates (2a–f), and Amides (3h,i)

Monocarbamates 1a–e			Dicarbamates 2a–f		
#	R ₁	R ₂	#	R ₁	R ₂
1a		H	2a		
1b		H	2b		
1c		H	2c		
1d		H	2d		
1e		H	2e		
			2f		

Amides 3h, 3i	
3h	
3i	

2. RESULTS AND DISCUSSION

2.1. Ligand-Based Design. A first pharmacophore model for anti-TB activity was previously developed by us¹² using a series of 32 imidazole derivatives with interesting antitubercular activity, adopting the HipHop¹³ method. The final model was then optimized,¹⁴ finally characterized by four pharmacophoric features as follows—a hydrogen bond acceptor feature (HBA), two aromatic ring features (RA1, RA2), and a hydrophobic feature (HY)—and applied recently to different antimycobacterial agents.^{10,15} Even if this model is able to describe the needed structural properties for antitubercular activity and identify the possible antimycobacterial candidates within large molecular databases, it does not permit correlation of quantitative biological activity of the compounds with their structural features. This limitation is due to the fact that the model was obtained by application of the qualitative approach referred to as the common feature hypothesis generation method. In addition, as for the specific case of the newly synthesized monocarbamates (1a–e), dicarbamates (2a–f), and amides (3h,i) of R-4-amino-3-isoxazolidinone (Figure 1, discussion in the External Test Set Prediction Analysis paragraph), a proper evaluation may be difficult when a partial overlap of the investigated compounds with the defined pharmacophore areas is established. In this perspective the use of a three-dimensional quantitative approach is useful, and several 3-D QSAR partial least squares (PLS) models, characterized by a training set (Table 2, Table S1 for numeric reference) of 71 published antitubercular agents,^{10,15–19} were

built through the 3-D QSAutogrid/R¹¹ protocol: 8 monoprobe (see Supporting Information Table S5 for probes' definitions) 3-D QSAR PLS models were generated and optimized via the CAPP¹¹ procedure (Tables 3 and 4) and a final multiprobe (MPGRS)¹¹ model (Tables 5 and S5) was then derived to correlate the pharmacophoric features required for antitubercular activity with molecular structures. Activity data, originally determined as MIC ($\mu\text{g}/\text{mL}$) values, were transformed to pMIC values on molar basis.

Three of the best monoprobe 3-D QSAR models, A, HD, and NA (Table 4 and Supporting Information Figure S1), accounting for different interaction patterns, were selected for further analysis, and the relative 3-D plots were inspected (Figure 2, Figures S2–S4). A comparison between these plots and the original pharmacophoric model¹⁴ was performed to check for spatial superposition of plot regions and pharmacophoric features (compare Figure 2A–C with D). Interpretation of the PLS-coefficients plots could be helped considering eq 1 where C_n is the C_{PLS} coefficient in the n th grid point, X_n is the actual field in the n th grid point, Y is the biological activity, and n is the number of grid points.

General equation for QSARs

$$Y = C_1X_1 + C_2X_2 + \dots + C_nX_n + e \quad (1)$$

As addressed by eq 1, C_{PLS} coefficients provide both interpretation of training set data (explaining the relative influence of each grid point by means of size and sign) and prediction of test set molecules' biological activity, Y ; an

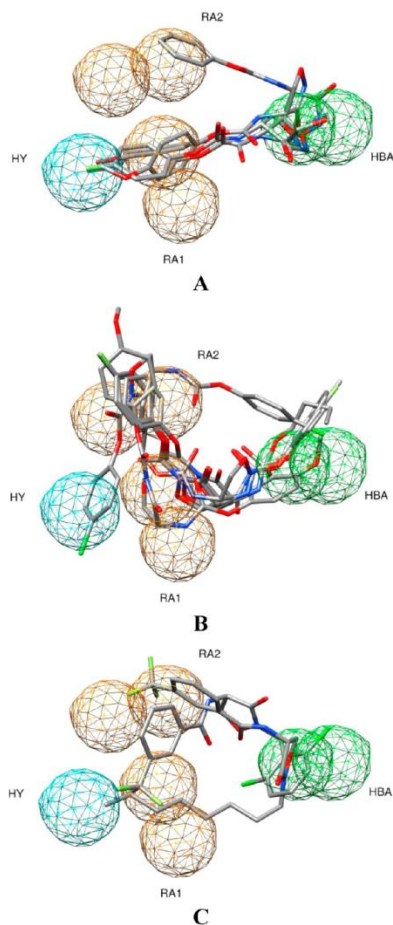


Figure 1. (A) Monocarbamates (1a–e). (B) Dicarbamates (2a–f). (C) Amides (3h,i) of D-4-amino-3-isoxazolidinone placed in the reported pharmacophoric model.¹⁴ HY (hydrophobic feature), RA (aromatic feature), HBA (hydrogen bond acceptor feature). The four pharmacophoric features are color-coded according to the original reference.

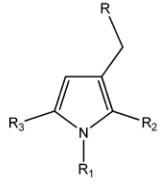
interaction characterized by a positive (repulsive) field X_n into a region with positive PLS-coefficient C_n will produce a positive effect ($C_n \times X_n$ product is positive), denoting a positive influence on Y (higher pMIC); the opposite is valid if the field or the PLS-coefficient have negative X_n or C_n , respectively. A positive effect (still considering pMIC activities) could be produced as well by a negative (attractive) field X_n into a region with negative PLS-coefficient C_n ($-C_n \times -X_n$ product is positive) and the opposite effect if the field or the PLS-coefficient have positive X_n or C_n , respectively.

As a result, four areas of the PLS-coefficients plots were distinguished over the N1, C2, C3, and C5 substituents of the pyrrole ring which overlap the pharmacophoric features HY, RA1, HBA, and RA2, respectively, thus suggesting a good agreement between QSAR and pharmacophoric models.

All monoprobe 3-D QSAR models clearly suggest that the presence of bulky groups as substituent at N1, C2, C5 is preferred, especially in N1 and C5 (HY and RA2 features). Moreover, the PLS-coefficient plots within thiomorpholines and methylpiperazinyl derivatives, (i.e., compounds **60** and **21** in Figure 2) and even more both PLS-coefficients and activity contribution plots (Supporting Information Figures S2–S4) show that at the C3 position steric features are required within certain limits. In particular, concerning the HBA feature, the HD model clearly shows that attractive interactions involving the sulfur of the thiomorpholine group of **60** (the most active compound, Figure S3A) increase the biological activity, while if these are missing or replaced by repulsive interactions, as for **21** (the least active compound), the biological effect decreases (Figure S3B). Therefore, a bulky group like thiomorpholinomethyl, also able to participate in electrostatic interactions, such as hydrogen bonds, is preferred at the C3 position. As for the structural features required for activity, the simultaneous analysis of PLS-loading and score plots were very useful to carry out the most relevant variables from the models (loading plots) and interpret the patterns seen in the score plots. Interesting is the case of the A probe model: starting from the first principal component (PC1), the presence of two clusters, differing each other for their conformational properties, is clearly showed in the score plot (Figure S5A). As shown in Figure S6 by superimposing the most influencing compounds (absolute higher score values) for each cluster to the PLS-loadings, the most important molecular feature in the PC1 space is related to both different spatial orientations and conformations. Light gray molecules, that are in the positive field cluster (positive PC1), had a higher activity and fill the lower part of the region between the RA2 and HBA features but not between RA1 and HBA. On the contrary, molecules with lower pMICs fill the area between the RA1 and HBA. PC2 and PC3 respectively gave information about substituents at C3 (Figures S5B, S7, and S8), suggesting the presence of bulky groups in the upper areas between the RA2 and HBA and over the HBA features have a detrimental effect on the biological activity.

Application of Multi-Probe Guided Region-Variable Selection. By application of the Multi Probe Guided Region Selection (MPGRS package), as implemented in 3-D QSAutogrid/R,¹¹ a multiprobe (MP) 3-D QSAR model was derived, representing, to the best of our knowledge, the first quantitative pharmacophoric model able to correlate the structural features of pyrrole derivatives with their biological data. The optimal MP 3-D QSAR model was characterized by a $PC_{FL,SL} = 1:3$,¹¹ and as previously reported,¹¹ its associated statistical coefficients (Table S, Figure S9) were similar to those of the monoprobe models, but the interpretation was greatly enhanced. Applying a q^2 threshold value of 0.4, the most relevant MIF subregions were selected (Figure 3) to build the multiprobe MIF and the resulting MP model condensing into one all the suggestions retrieved by the analysis conducted on the monoprobe models. In particular, the MP PLS-loadings in association with the MP score plots identified the same conformational differences, addressed by the monoprobe models, as the most discriminating aspect in molecular clustering, for example: starting from $PC_{1,1}$ to $PC_{1,2}$ a similar clustering in the score plots and in meaning for the descriptors

Table 2. Structure and Antimycobacterial Activity against *M. tuberculosis* 103471 of the Pyrrole Derivatives Used As a Training Set for the Generation of the 3-D QSAR Models



compd ^a	R ^b	R ₁	R ₂	R ₃	pMIC ^c
1	B	2-F-Ph	CH ₃	2-F-Ph	4.68
2	A	2-Cl-Ph	CH ₃	2-F-Ph	5
3	B	2-Cl-Ph	CH ₃	2-F-Ph	4.09
4	A	2-F-Ph	CH ₃	2-Cl-Ph	5
5	B	2-F-Ph	CH ₃	2-Cl-Ph	4.4
6	A	2-Cl-Ph	CH ₃	2-Cl-Ph	5.02
7	B	2-Cl-Ph	CH ₃	2-Cl-Ph	4.41
8	A	2-F-Ph	CH ₃	α -naphthyl	4.11
9	B	2-F-Ph	CH ₃	α -naphthyl	4.11
10	A	2-Cl-Ph	CH ₃	α -naphthyl	4.13
11	B	2-Cl-Ph	CH ₃	α -naphthyl	4.13
12	A	α -naphthyl	CH ₃	2-Cl-Ph	4.13
13	B	4-F-Ph	CH ₃	Ph	4.36
14	B	Ph	CH ₃	4-F-Ph	4.36
15	A	4-Cl-Ph	CH ₃	4-F-Ph	5.30
16	B	4-F-Ph	CH ₃	4-F-Ph	4.47
17	A	4-F-Ph	CH ₃	4-F-Ph	5.58
18	B	4-F-Ph	CH ₃	4-Cl-Ph	5.30
19	A	4-F-Ph	CH ₃	4-Cl-Ph	5.60
20	A	2-F-Ph	CH ₃	Ph	4.66
21	B	2-F-Ph	CH ₃	Ph	4.06
22	A	Ph	CH ₃	2-F-Ph	4.96
23	B	Ph	CH ₃	2-F-Ph	4.36
24	A	2-Cl-Ph	CH ₃	Ph	4.38
25	B	2-Cl-Ph	CH ₃	Ph	4.07
26	B	Ph	CH ₃	2-Cl-Ph	4.07
27	A	α -naphthyl	CH ₃	Ph	4.1
28	B	α -naphthyl	CH ₃	Ph	4.09
29	A	Ph	CH ₃	α -naphthyl	4.10
30	B	Ph	CH ₃	α -naphthyl	4.09
31	B	Ph	CH ₃	Ph	4.33
32	A	4-F-Ph	CH ₃	2-Cl-Ph	5.00
33	B	4-F-Ph	CH ₃	2-Cl-Ph	4.70
34	B	4-F-Ph	CH ₃	2-F-Ph	4.08
35	A	4-F-Ph	CH ₃	4-CH ₃ -Ph	5.98
36	B	4-F-Ph	CH ₃	3-CH ₃ -Ph	4.37
37	A	4-F-Ph	CH ₃	2-CH ₃ -Ph	4.98
38	B	4-F-Ph	CH ₃	2-CH ₃ -Ph	4.07
39	A	4-F-Ph	CH ₃	2,4-Cl ₂ -Ph	5.34
40	B	4-F-Ph	CH ₃	2,4-F ₂ -Ph	5.00
41	A	2-Cl-Ph	CH ₃	4-F-Ph	5.30
42	B	2-Cl-Ph	CH ₃	4-F-Ph	5.00
43	B	2-F-Ph	CH ₃	4-F-Ph	4.68
44	A	4-CH ₃ -Ph	CH ₃	4-F-Ph	5.58
45	A	3-CH ₃ -Ph	CH ₃	4-F-Ph	4.98
46	B	3-CH ₃ -Ph	CH ₃	4-F-Ph	4.40
47	A	2-CH ₃ -Ph	CH ₃	4-F-Ph	4.68
48	B	2-CH ₃ -Ph	CH ₃	4-F-Ph	4.10
49	A	2,4-Cl ₂ -Ph	CH ₃	4-F-Ph	5.64
50	B	2,4-Cl ₂ -Ph	CH ₃	4-F-Ph	5.03
51	A	2,4-F ₂ -Ph	CH ₃	4-F-Ph	5.30
52	B	2,4-F ₂ -Ph	CH ₃	4-F-Ph	4.40
53	A	4-F-Ph	CH ₃	4-C ₂ H ₅ -Ph	5.60
54	A	4-F-Ph	CH ₃	4- <i>i</i> -propyl-Ph	6.21
55	A	4-C ₂ H ₅ -Ph	CH ₃	4-F-Ph	5.30
56	A	4-C ₃ H ₇ -Ph	CH ₃	4-F-Ph	5.61
57	A	4-Cl-Ph	CH ₃	4-CH ₃ -Ph	5.90
58	A	4-Cl-Ph	CH ₃	4-C ₂ H ₅ -Ph	6.22
59	A	4-Cl-Ph	CH ₃	4-C ₃ H ₇ -Ph	6.23
60	A	4-Cl-Ph	CH ₃	4- <i>i</i> -propyl-Ph	6.53
61	A	4-CH ₃ -Ph	CH ₃	4-Cl-Ph	5.90
62	A	4-C ₂ H ₅ -Ph	CH ₃	4-Cl-Ph	5.91
63	A	4-C ₃ H ₇ -Ph	CH ₃	4-Cl-Ph	6.23
64	A	4- <i>i</i> -propyl-Ph	CH ₃	4-Cl-Ph	6.23
65	B	4-Cl-Ph	C ₂ H ₅	4-Cl-Ph	5.33
66	A	4-F-Ph	C ₂ H ₅	4-CH ₃ -Ph	6.20
67	A	Ph	C ₂ H ₅	Ph	5.26
68	A	Ph	C ₂ H ₅	4-F-Ph	5.58
69	A	4-F-Ph	C ₂ H ₅	Ph	5.28
70	A	2-F-Ph	C ₂ H ₅	4-F-Ph	5.30
71	A	2-F-Ph	C ₂ H ₅	2-F-Ph	5.00

^aA = thiomorpholin-4-yl and B = 4-methylpiperazin-1-yl. ^bpMIC = $-\text{Log}[\text{MIC}(\mu\text{M}) \times 10^{-6}]$. ^cCompound enumeration was assigned on the basis of the original increasing numbering from the oldest to the most recent reference. Supporting Information Table S1 shows the connections between the new and original enumerations.

Table 3. CAPP Settings

min value	parameter ^a	max value	step
0	PCO	10	1.0
0	Zeroing	0.05	0.005
0	MSDCO	5	1

^aPCO: positive cut off. Zeroing: zeroing of very low data points. MSDCO: minimum standard deviation cut off.

to those in the A monoprobe model was noticed (compare Supporting Information Figures S10–S12 and Figures S5–S7), confirming the above assumptions (effect of difference in spatial arrangement and conformation). The MP PLS-coefficient plot showed that the most important regions were spatially and chemically overlapping with the pharmacophoric model¹⁴ (compare Figure 3B and D). Taking into account the probe

type with the associated PLS-coefficient sign, bulky groups seemed to be required at the N1, C2, and C5 positions (positive coefficients); furthermore negative PLS-coefficients were spread in the proximity of these areas and chlorine and fluorine substituents are associated to activity enhancement, these additional areas can be related to some electrostatic molecular environment (Figure 3B and C) in agreement with the pharmacophoric model.^{16,18} In addition, regarding the C3 position (in the lower part of HBA and slightly extended toward RA2) the model indicates that a limited steric repulsion is tolerated and electrostatic endowed groups could be profitable for the activity. Further information about the HBA feature was derived overlapping the clustered molecules (Figure S10B) with both PLS-loadings and PLS-coefficients at PC_{1,3} (Figure 4A and B): the implementation of different probes (such as NA

Table 4. 3-D QSAutogrid/R PLS Model Statistical Results^{a,b}

model	P	PC	r^2	q_{LOO}^2	q_{KSCV}^2	r_{YS}^2	q_{YS}^2	V
1	A	3	0.92	0.86	0.85	0.36	-0.33	3758
2	C	3	0.92	0.86	0.85	0.37	-0.33	4492
3	HD	3	0.91	0.85	0.84	0.39	-0.31	1217
4	NA	3	0.91	0.86	0.85	0.31	-0.33	531
5	N	3	0.91	0.85	0.85	0.32	-0.30	477
6	OA	3	0.91	0.85	0.85	0.36	-0.33	658
7	e	4	0.88	0.78	0.76	0.40	-0.48	468
8	d	4	0.91	0.85	0.84	0.35	-0.44	4412

^aCAPP process was applied. ^bP: Autogrid Probe. PC: optimal number of principal components/latent variables. r^2 : conventional square-correlation coefficient. q_{LOO}^2 : cross-validation correlation coefficient using the leave-one-out method. q_{KSCV}^2 : cross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations. r_{YS}^2 : average square-correlation coefficient obtained after Y-scrambling process using 100 iterations; q_{YS}^2 : average cross-validation correlation coefficient using the leave-one-out method obtained after Y-scrambling process using 100 iterations. V: number of active variables.

and HD) suggested that the presence of bulky groups in the upper areas between the RA2 and HBA and over the HBA feature might have a detrimental effect on the biological activity; i.e. considering the methylpiperazinyl moiety (characterizing most of the negative clustered molecules, Figure 4B2) the methyl group fits the HD areas characterized by a negative PLS-coefficient, while the thiomorpholinomethyl moiety (that discriminate the positive clustered molecules, Figure 4B1) satisfies both steric and electrostatic features leading to higher activities. In this scenario the MP model was able to increase the resolution of the HBA region revealing an extra partial steric role.

External Test Set Prediction Analysis. The 8 3-D QSAutogrid/R monoprobe models were externally validated using the 13 newly synthesized monocarbamates (1a–e), dicarbamates (2a–f), and amides (3h–i) (Table 1). A fact must be emphasized: since the training set was composed only by pyrrole derivatives to directly compare the quantitative models with the original pharmacophoric assumptions, the resulting quantitative structure–activity relationships were based, mostly, on the characteristics of the scaffolds composing the training set. This may result in a limitation of the models to predict the activities of other molecular classes: in this case, specifically, a major difference between the two sets, training set and test set, was represented by the fact that the former was characterized by the pyrrole ring, which permits a quadruple branching able to satisfy simultaneously the different pharmacophoric areas; on the contrary, the compounds of the test set were characterized by a double branching. Despite this fact, and considering also that the test set molecules showed similar activity values (total pMIC activity range = 1.31 log unit), acceptable errors of prediction (SDEP coefficients all

below the unit except for the d model) were obtained (Table 6); but an analysis focused only on the statistical SDEP values or experimental vs predicted plots (Figure S13) could be misleading. In fact, analyzing only the statistical results might seem that the PLS models had good predictive ability toward the isoxazolidinone derivatives, while considering only the experimental vs predicted plots the same conclusion could not be reached. In this case, it was helpful to analyze both of these pieces of information, together with the average absolute error of predictions (AAEP, Supporting Information Table S6), for each molecule from all the 3-D QSAR monoprobe models and finally their placement in the 3-D space of PLS-coefficients. Analyzing, for each test set molecule, the AAEP from all the eight monoprobe models stood out the good predictive capacity toward 10 of these, while for 1a, 2c, and 2e, the AAEPs were 1.05, 1.50, and 1.72, respectively. These compounds were overpredicted, and this can be sought precisely in the dependence of the model from the training set congeners and from consequent inevitable alignment limitations. Furthermore 1a was predicted more active than 1d likely due to the fact that its isoxazolidinonic carbonyl group was perfectly superimposed to the training set most active compound (60) thiomorpholinic sulfur atom, showing how important was for the models the presence of a group capable to accept hydrogen bonds in the HBA space. It should be stressed, however, that such molecules (1a, 2c, and 2e), probably, would have been discarded by adopting the original pharmacophoric model;¹⁴ in fact, compound 1a misses the RA1 and HY features while RA2 and HBA are satisfied with the presence of a phenyl in R1 and carbonyl group of the isoxazolidinone ring, respectively; 2c satisfies only the RA2 feature, and partially the HBA feature with the presence of the *p*-methoxyphenylic oxygen; 2e accomplishes the HY and partially the RA1 feature; whereas all the 3-D QSAR models are able to frame their level of activity: for this reason and for the above considerations, the models show a good predictive ability although different scaffold endowed molecules were used as test set. As examples of the 3-D QSAR model application the most and least active monocarbamate derivatives (1d and 1a, respectively) overlapped with the A probe model PLS-coefficients are depicted in Figure 5A, while the most and the least active dicarbamate molecules (2d and 2f, respectively) are reported in Figure 5B. Despite the above considerations differences in experimental activities appears to be determined by a better overlap of the aromatic ring with the *p*-fluorine on the HY and RA2 regions. Analogously, the MP model predictions were in good agreement with those of the monoprobe models (Table 7). The 3-D QSAR MP plots showing the more and less active monocarbamates and dicarbamates derivatives (Figure 6) indicated the lack of a simultaneous coverage of the different regions addressed by the PLS-coefficients. However, as shown in Figure 6 was confirmed the importance of hydrophobic substituents in the HY and RA2 areas, which should determine the highest activities of 1d and 2d.

Table 5. MPGRS: Multi Probe Model Statistical Results

PC _{FL/SL} ^a	r^2 ^b	MPGRS 3-D QSAR					
		q_{LOO}^2 ^c	q_{KSCV}^2 ^d	SDEP _{LOO}	SDEP _{KSCV}	r_{YS}^2	q_{YS}^2
1:3	0.88	0.80	0.80	0.32	0.32	0.31	-0.31

^aOptimal number of principal first level (FL) and second level (SL) components for the MPGRS model; ^bConventional square-correlation coefficient. ^cCross-validation correlation coefficient using the leave-one-out method. ^dCross-validation correlation coefficient using the k -fold cross-validation with 5 random groups and 100 iterations.

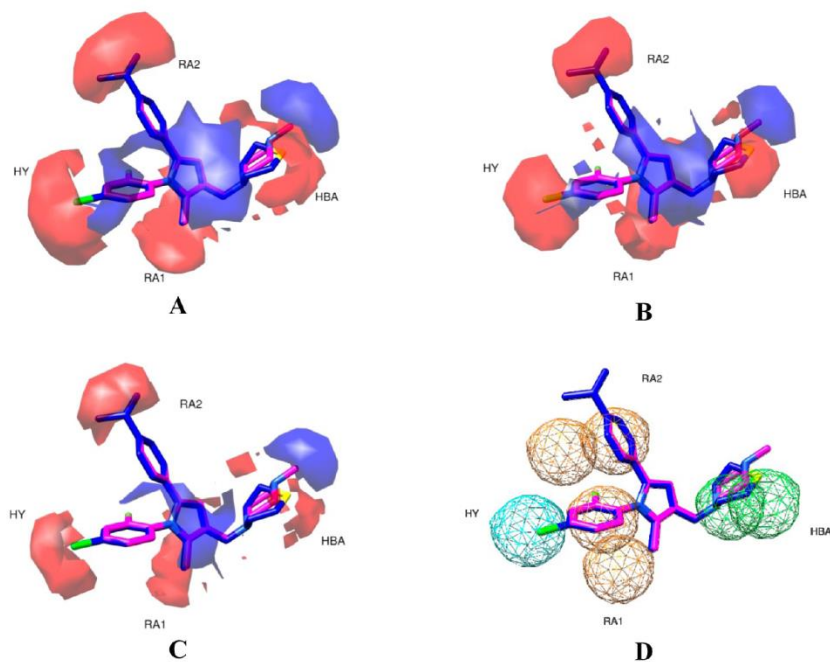


Figure 2. Most active (**60** in blue) and least active (**21** in magenta) compounds. (A) PLS-coefficients contour maps derived from A probe analysis (contour levels 80%; positive red, negative blue). (B) PLS-coefficients contour maps derived from HD probe analysis (contour levels 85%; positive red, negative blue). (C) PLS-coefficients contour maps derived from NA probe analysis (contour levels 75%; positive red, negative blue). (D) Pharmacophoric features derived from the original pharmacophoric model:¹⁴ HY (hydrophobic feature), RA (aromatic feature), HBA (hydrogen bond acceptor feature).

2.2. Chemistry. The synthesis of compounds (**1a–e**) and (**2a–f**) was carried out modifying a literature procedure described by Stammer and co-workers,²⁰ by treatment of D-4-amino-3-isoxazolidinone in weakly alkaline media (1 M NaHCO₃) with the corresponding chloroformate to obtain both mono- and dicarbamate derivatives, as illustrated in Scheme 1.

The derivatives **1a–f** were prepared by regioselective acylation of the 4-amino group using the appropriate chloroformates at low temperature for short reaction time (0 °C for 3 h); then pure solids **1a–f** were obtained by acidification with 4 M HCl. Dicarbamate derivatives **2a–e** were synthesized dissolving the D-4-amino-3-isoxazolidinone in a basic solution at 0 °C and the selected chloroformates were dropwise added. **2a–e** gradually precipitate in 12 h from the aqueous solution.

The compound **2f** was obtained by treatment of **1f** with *n*-butylchloroformate in alkaline solution.

N-2-(alkyl)-4-amino-3-oxoisoxazolidinone **4h,i** were synthesized by an electrochemical reaction as previously reported.²¹ Then crude **4h,i** were acylated with 3-trifluoromethyl benzoyl chloride in chloroform/TEA to give the amide derivatives **3h,i**.

In order to verify the racemization of the α carbon of D-4-amino-3-isoxazolidinone in the reaction conditions, we have analyzed by chiral HPLC the enantiopurity of (*R*)-**3h** and (*S*)-**3h**, obtained with the same synthetic procedure starting from (*R*)-4-amino-3-isoxazolidinone and (*S*)-4-amino-3-isoxazolidinone. Chiralpak Column IC 250 mm \times 4.6 mm I.D. was used with *n*-hexane-2-propanol 75/25 (v/v) as eluent at flow

rate of 1.0 mL/min at the temperature of 25 °C. In both the chromatogram of (*R*)-**3h** and (*S*)-**3h** an enantiomeric excess >99.0% was observed (Supporting Information Figure S15).

2.3. Biological Activity of Synthesized Compounds.

The compounds were assayed for their antimycobacteria activity toward *M. tuberculosis* H37Rv (ATCC 27294). The minimal concentration inhibiting visible growth of mycobacteria was determined for each compound.

Concerning the data reported in Table 8, only the acylation of 4-amino group and acylation or alkylation of N-2 of D-4-amino-3-isoxazolidinone influenced the antitubercular activity leading to a MIC value of 3.1 μ g/mL (as in the case of **1d** and **2d**). Most of the tested compounds showed the same activity of the 4-amino-3-isoxazolidinone (32 μ g/mL), while **1d** and **2d** were more active (3.1 μ g/mL) and only the compound **3h** was less active (64 μ g/mL).

3. CONCLUSION

In this paper we present the first application of a quantitative pharmacophoric model able to define and correlate the needed chemical characteristics with antitubercular activity of a previously reported class of antimycobacterial agents.^{8,13–17} Eight 3-D QSAR monoprobe models and a multi probe (MP) model were built showing appreciable statistical coefficients and allowing an accurate definition of the structure–activity relationships on the basis of pyrrole derivatives used as training set. The MP 3-D QSAR model allowed defining the training set

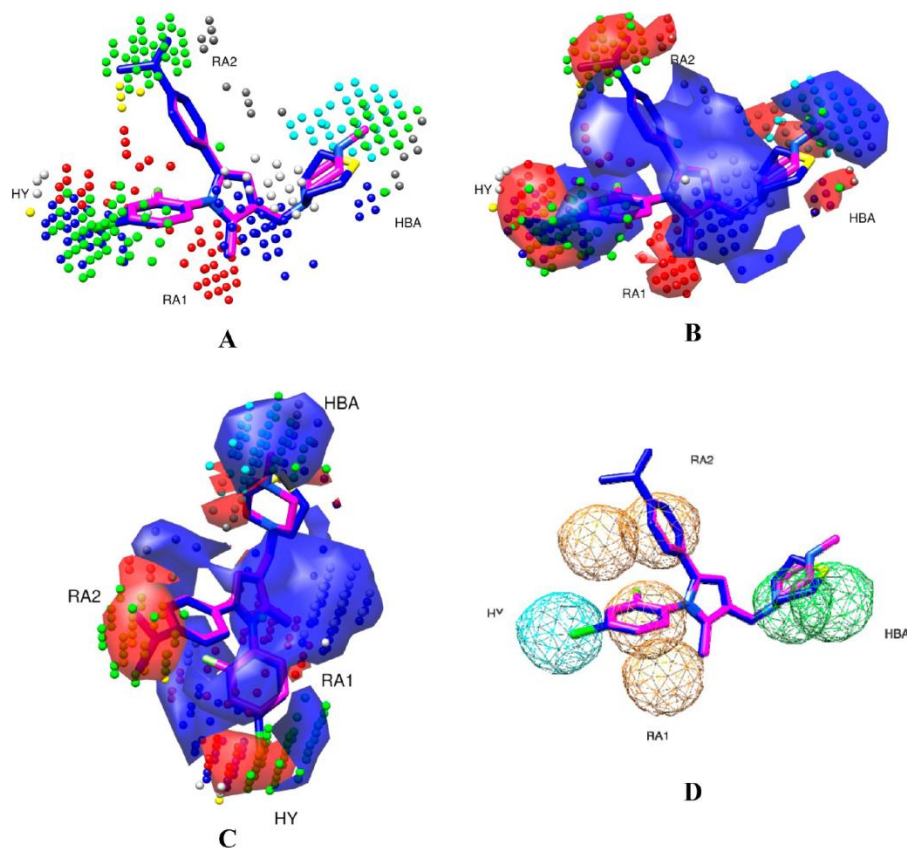


Figure 3. MPGRS. (A) Key points. The points are color-coded according to that reported in Supporting Information Table S5. (B) Key points with PLS-coefficients contour maps (contour levels: positive 85%, red; negative 95%, blue). (C) Top view, key points with PLS-coefficients solid contour maps (contour levels: positive 85%, red; negative 95%, blue). The most active (60 in blue) and the least active (21 in magenta) compounds are shown. (D) Pharmacophoric features derived from the original pharmacophoric model:¹⁴ HY (hydrophobic feature); RA (aromatic feature), HBA (hydrogen bond acceptor feature).

molecular features and their three-dimensional positioning, configuring itself as a quantitative pharmacophoric model. Furthermore it was possible to elucidate the effect of conformational differences on the biological activity. As a further assessment the multi probe information was compared with the original pharmacophoric model, previously developed by us,¹⁴ showing an high degree of correspondence.

Independently, a series of 13 isoxazolidinone derivatives 1–3 (Table 1) was synthesized and tested as new antitubercular compounds. The new compounds showed MIC values in the micromolar range. In particular among the monocarbamates and dicarbamates, derivatives 1d and 2d showed the higher biological activities. Although there are limitations due to structural differences between the molecules of the training set and those of derivatives 1–3, the latter were used as an external test set to evaluate the models' predictive capabilities. All the 3-D QSAR models showed prediction errors (Tables 6

and 7), against these structurally unrelated molecules, with an acceptable degree of approximation.

The application of the models allowed to clarify the role of halogens and phenyl rings in 1a–e and 2a–f. Considering all these outcomes, the MP 3-D QSAR model could represent a useful tool for the design of new antitubercular drugs.

4. EXPERIMENTAL SECTION

4.1. Molecular Modeling and 3-D QSAR. All calculations used a 6 blade (8 Intel-Xeon E5520 2.27 GHz CPU and 24 GB DDR3 RAM each) cluster (48 CPU total) running Debian GNU/Linux 6.0 64 bit operating system. A series of 71 previously described pyrrole derivatives^{10,15,19} were used to build 8 single probe and a multi probe 3-D QSAR models using the 3-D QSAutogrid/R procedure.¹¹ The obtained models were tested predicting the activities of the monocarbamates (1a–e), dicarbamates (2a–f), and amides (3h,i) of D-4-amino-3-isoxazolidinone derivatives.

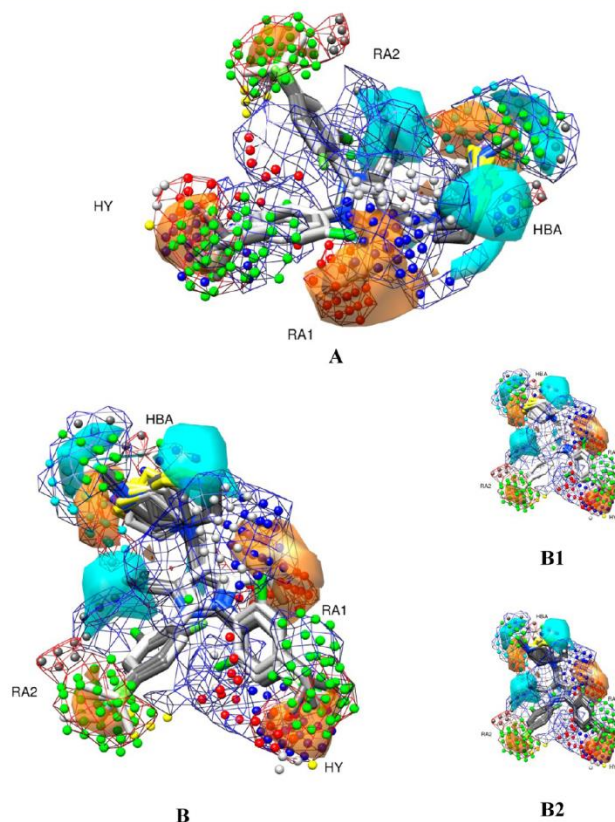


Figure 4. MPGRS. PLS-loadings contour maps at $PC_{1,3}$ (contour levels 75%; positive orange, negative cyan) with PLS-coefficients (mesh levels positive 85%, red; negative 95%, blue) and key points (see Table S5 for color coding). The ten most important molecules for each cluster are plotted and color-coded (compounds in the positive loading field in light gray; compounds in the negative loading field in dark gray): (A) side view; (B) frontal view; (B1) frontal view of only positive clustered molecules; (B2) frontal view of only negative clustered molecules. HY (hydrophobic feature), RA (aromatic feature), HBA (hydrogen bond acceptor feature).

Table 6. Test Set Predictions^a

P	PC	SDEP _{EXT}
A	3	0.88
C	3	0.88
HD	3	0.81
NA	3	0.82
N	3	0.83
OA	3	0.84
e	4	0.90
d	4	1.51

^aSDEP values considering the optimal PCs. P: AutoGrid Probe. PC: optimal number of principal components/latent variables. SDEP_{EXT}: standard deviation error of prediction (or root mean squared error of prediction, RMSEP) for the external test set.

Training Set Selection. Starting from a training set composed by 90 pyrrole derivatives,^{10,15,19} a selection based on inner relationship analysis were conducted to improve the robustness and prediction capabilities of the 3-D QSAR

Table 7. MPGRS: Multi Probe Model Test Set Predictions^a

P	PC _{FL,SL}	SDEP _{EXT}
AutoGrid MP	1:3	0.89

^aSDEP values considering the optimal first level and second level PCs. P: AutoGrid Multi-Probe. PC_{FL,SL}: optimal first level and second level PCs. SDEP_{EXT}: standard deviation error of prediction (or root mean squared error of prediction, RMSEP) for the external test set.

models: this leads to a final training set composed by 71 molecules (Table 2).

Alignment Rules. Training Set: Training set compounds were first submitted to a conformational search following a computational protocol previously described.¹⁰ Next, each compound with its conformational models was aligned to the pharmacophoric model with the flexible fitting method implemented in Discovery Studio (version 3.0, Accelrys, Inc., San Diego, CA), that allows slight modification of each conformation to better fit the pharmacophore.

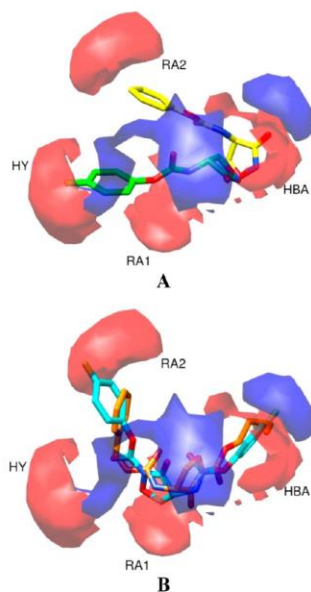


Figure 5. PLS-coefficients contour maps. (A) AutoGrid/R PLS-coefficients contour maps derived from the A probe analysis (contour levels 80%; positive red, negative blue; **1a** yellow; **1d** green); (B) AutoGrid/R PLS-coefficients contour maps derived from the A probe analysis (contour levels 80%; positive red, negative blue; **2d** cyan; **2f** orange). HY (hydrophobic feature), RA (aromatic feature), HBA (hydrogen bond acceptor feature).

Test Set: The new 13 derivatives were aligned using the Surflex-Sim²² software which has been chosen since it is a valuable tool in ligand-based drug discovery, free for academics, and its alignment process is based on morphological similarities. The query molecules' poses were optimized to the compounds used as training set to maximize 3-D similarity.

As shown in Supporting Information Figures S16 and S17 and Tables S9 and S10, similar results were obtained using the same alignment software (pharmacophoric alignment) adopted for the training set. The choice of Surflex alignment was not dictated by the improvement (although negligible respect the pharmacophoric one) in prediction, but by the fact that, in our view, this is a further confirmation of the robustness of the models, always taking into account the above limits. Indeed, using two different procedures of alignment, results are comparable and this should show that the predictive capability of the models is stable and in the specific case scarcely influenced. Further clarifications on the differences in prediction (as in the case of **1d** and **2d**) would be only speculative since the presence of similar activity values with a limited total pMIC activity range.

Strategies for different alignments were also tried leading to not consistent prediction supporting that the best alignment/prediction is that reported above.

Molecular Interaction Fields Calculation. As reported,¹¹ MIFs were generated using the AutoGrid Software (AutoDock Suite,²³ based on the AMBER united-atom Force Field) implemented in the 3-D QSAutoGrid/R procedure, considering

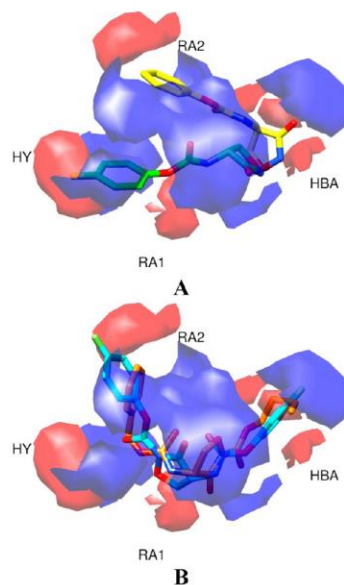


Figure 6. MPGRS. (A) PLS-coefficients contour maps at PC_{1.3} (contour levels: positive 85%, red; negative 95%, blue. **1a**: yellow. **1d**: green). (B) PLS-coefficients contour maps at PC_{1.3} (contour levels: positive 85%, red; negative 95%, blue. **2d**: cyan. **2f**: orange). HY (hydrophobic feature), RA (aromatic feature), HBA (hydrogen bond acceptor feature).

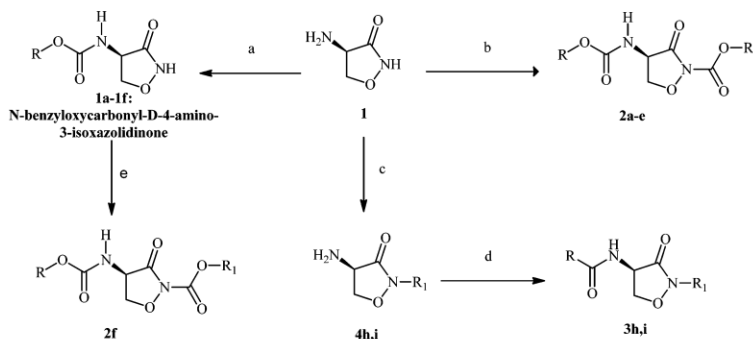
Table 8. MIC Data for D-4-Amino-3-isoxazolidinone Derivatives

compd	MIC ($\mu\text{g/mL}$) ^a	pMIC ^b
1a	32	3.84
1b	32	3.87
1c	32	3.90
1d	3.1	4.89
1e	32	3.97
2a	32	4.03
2b	32	4.06
2c	32	4.10
2d	3.1	5.09
2e	32	4.11
2f	32	4.02
3h	64	3.78
3i	32	4.13

^a*M. tuberculosis* H37Rv (ATCC 27294) was used. MIC values represent the minimal concentrations of compounds completely inhibiting visible growth of mycobacteria. ^bpMIC = $-\text{Log}[\text{MIC} (\mu\text{M}) \times 10^{-6}]$.

8 different probes. Interaction energies between the selected probes and each molecule were computed using a grid spacing of 1 Å (Supporting Information Table S8). The xyz coordinates (in angstroms) of the grid rectangular box used for the computation were Xmin/Xmax = -9.828/12.172, Ymin/Ymax = -9.021/8.979, and Zmin/Zmax = -10.481/9.519.

Statistical Analysis. Through the D2M package,¹¹ eight 3-D QSAR PLS models were built. During the model definition the

Scheme 1^a

^a(a) R-O-CO-Cl, 1 M NaHCO₃, 0 °C 3 h, 4 M HCl; (b) R-O-CO-Cl, NaHCO₃, 0 °C, 15 h; (c) 0.1 M TEAHFP in CH₃CN, 30 mA cm⁻², D-4-amino-3-isoxazolidinone 1 eq (15 min), R₁-Br; (d) 3-CF₃-C₆H₄COCl, TEA, CHCl₃; (e) 1f, *n*-Bu-O-CO-Cl, 1 M NaHCO₃, 0 °C, 24 h.

assessment of quality and robustness was conducted via two cross-validation (CV package)¹¹ procedures as follows: (1) leave-one-out (LOO) and (2) *k*-fold (KF, 5-random groups and 100 iterations) methodologies. Initially, the raw models (Supporting Information Table S2) were optimized through the Combinatorial Analysis of Pretreatment Parameters (CAPP package)¹¹ setting the pretreatment intervals as listed in Table 3, using the *k*-fold cross-validation with 5-random groups and 100 iterations and monitoring the *q*² and SDEP values. A total of 726 combinations, for each 3-D QSAR model, were processed using 5% sPRESS reduction¹¹ to select the optimal pretreatment combination and derive the pretreated PLS models; this led to an average *q*² KSFCV value increment equal to 14% (Supporting Information Table S4). Furthermore the scrambling approach, Y-scrambling, (package YS)¹¹ was applied to investigate the presence of chance correlations using 100 iterations. Considering the obtained good overall statistical coefficients together with the absence of chance correlations (Table 4, Supporting Information Table S3 and Figure S1), no further variable selection steps were performed. By the application of the MPGRS package,¹¹ a MP 3-D QSAR PLS model was then derived by selecting the most informative subregions for each of the eight considered probe; the same CV and scrambling procedures as those of the monoprobe models were performed and the optimal MP 3-D QSAR model was selected according to the *q*_{PL:SL}² values.¹¹ Similar statistical coefficients to those of the monoprobe models were obtained (Table 5, Supporting Information Figure S9) and no further variable selection were performed; finally the most relevant MIFs subregions were selected applying a *q*² threshold value of 0.4.

4.2. Chemistry. D-4-Amino-3-isoxazolidinone and all chloroformates were purchased from Sigma-Aldrich (Milano, Italy). All other reagents and solvents were of higher analytical grade. *N*-Benzyloxycarbonyl-D-4-amino-3-isoxazolidinone (**1f**) was prepared according to the procedure of Stammer et al.²⁰ Melting points were determined on Tottoli apparatus (Buchi) and are uncorrected. Vibrational spectra were recorded on a Spectrum One ATR Perkin-Elmer FT-IR spectrometer. ¹H and ¹³C NMR spectra were acquired on a Bruker AVANCE-400 spectrometer at 9.4 T, in DMSO-*d*₆ or CDCl₃ at 27 °C; chemical shift values are given in δ (ppm) relatively to TMS as internal reference, coupling constants are given in hertz.

Mass spectra were recorded on a API-TOF Mariner by Perspective Biosystem (Straford, Texas, USA), and samples were injected by a Harvard pump using a flow rate of 5–10 μ L/min, infused in the electrospray system, a TSQ quadrupole mass spectrometer by Thermofinnigan (San Jose, California, USA) operating in CH₄/CI conditions; samples were introduced in the CI source by a direct insertion probe. Elemental analyses were obtained by a PE 2400 (Perkin-Elmer) analyzer.

General Procedure for Synthesis of Monocarbamates 1a–e. Compounds **1a–e** were prepared by a modified procedure described by Stammer et al.²⁰ briefly, to 0.5 mmol of D-4-amino-3-isoxazolidinone, dissolved in 1.2 mL of aqueous 1 M NaHCO₃, cooled in an ice bath, 1.0 mmol of the corresponding chloroformate was added and the solution stirred for 3 h at 0 °C. Little amounts of precipitate that could be formed were filtered off, and the solution, kept to 0 °C, was acidified to pH 4 with 4 M aqueous HCl. After 30 min, the white fine precipitate was collected by centrifugation and washed with cold water. The collected carbamates **1a–e** were crystallized from water.

Synthesis of (*R*)-Phenyl-(3-oxoisoxazolidin-4-yl)carbamate (1a). **1a** was obtained as white crystalline solid in 75% yield, mp 148–149 °C. ¹H NMR (DMSO-*d*₆): 11.52 (1H, bs, D₂O exchange); 8.37 (1H, s, D₂O exchange); 7.38 (t, 2H, *J* = 8.56 Hz); 7.21 (t, 1H, *J* = 7.58 Hz); 7.12 (d, 2H, *J* = 8.56 Hz); 4.65–4.60 (m, 1H); 4.56 (t, 1H, *J* = 9.78 Hz); 4.03 (t, 1H, *J* = 8.32 Hz). ¹³C NMR (DMSO-*d*₆): 170.5; 154.9; 151.4; 129.9; 125.9; 122.3; 72.2; 53.5. FT-IR (cm⁻¹): 3339, 1709, 1655. MS/ESI: (M + H)⁺ 223.0730 (*m/z*).

Synthesis of (*R*)-4-Methylphenyl-(3-oxoisoxazolidin-4-yl)carbamate (1b). **1b** was obtained as a white crystalline solid in 75% yield, mp 195–196 °C. ¹H NMR (DMSO-*d*₆): 11.54 (bs, 1H, D₂O exchange); 8.30 (s, 1H, D₂O exchange); 7.16 (d, 2H, *J* = 8.56 Hz); 6.99 (d, 2H, *J* = 8.56 Hz); 4.66–4.59 (m, 1H); 4.55 (t, 1H, *J* = 10.35 Hz); 4.02 (t, 1H, *J* = 8.32 Hz); 2.28 (s, 3H). ¹³C NMR (DMSO-*d*₆): 169.9; 154.4; 148.5; 134.2; 129.6; 121.3; 71.6; 52.8; 20.3. FT-IR (cm⁻¹): 3305, 1709, 1654. MS/ESI: (M + H)⁺ 237.0874 (*m/z*).

Synthesis of (*R*)-4-Methoxyphenyl-(3-oxoisoxazolidin-4-yl)carbamate (1c). **1c** was obtained as a white crystalline solid in 70% yield, mp 155–156 °C. ¹H NMR (DMSO-*d*₆): 11.45 (bs, 1H, D₂O exchange); 8.27 (s, 1H, D₂O exchange); 7.02 (d, 2H, *J* = 9.06 Hz); 6.91 (d, 2H, *J* = 9.06 Hz); 4.64–4.59

(m, 1H); 4.54 (t, 1H, $J = 9.29$ Hz); 4.02 (t, 1H, $J = 9.29$ Hz); 3.73 (s, 3H). ^{13}C NMR (DMSO- d_6): 170.0; 156.5; 154.6; 144.2; 122.5; 114.2; 71.6; 55.4; 52.8. FT-IR (cm^{-1}): 3314, 1709, 1655. MS/ESI: $(\text{M} + \text{H})^+$ 253.0790 (m/z).

Synthesis of (R)-4-Fluorophenyl-(3-oxoisoxazolidin-4-yl)-carbamate (1d). **1d** was obtained as a white crystalline solid in 70% yield, mp 175–176 °C. ^1H NMR (DMSO- d_6): 11.54 (bs, 1H, D_2O exchange); 8.39 (s, 1H, D_2O exchange); 7.16 (d, 2H, $J = 8.56$ Hz); 6.98 (d, 2H, $J = 8.56$ Hz); 4.68–4.63 (m, 1H); 4.55 (t, 1H, $J = 9.56$ Hz); 4.03 (t, 1H, $J = 8.80$ Hz). NMR (DMSO- d_6): 155.9 (d, $J = 237.8$ Hz); 154.1; 147.3; 123.9 (d, $J = 7.32$ Hz); 116.5; 116.0 (d, $J = 22.7$ Hz); 72.0; 53.4. FT-IR (cm^{-1}): 3320, 1712, 1699. MS/CI: $(\text{M} + \text{H})^+$ 241 (m/z).

Synthesis of (R)-4-Bromophenyl-(3-oxoisoxazolidin-4-yl)-carbamate (1e). **1e** was obtained as a white crystalline solid in 75% yield, mp 180–181 °C. ^1H NMR (DMSO- d_6): 11.55 (bs, 1H, D_2O exchange); 8.42 (s, 1H, D_2O exchange); 7.22 (d, 2H, $J = 8.56$ Hz); 7.16 (d, 2H, $J = 8.56$ Hz); 4.69–4.62 (m, 1H); 4.56 (t, 1H, $J = 9.88$ Hz); 4.04 (t, 1H, $J = 8.80$ Hz). ^{13}C NMR (DMSO- d_6): 170.0; 156.5; 154.6; 144.2; 122.5; 114.2; 71.6; 52.8. FT-IR (cm^{-1}): 3330, 1715, 1705. MS/CI: $(\text{M} + \text{H})^+$ 302 (m/z).

General Procedure for Synthesis of Dicarbamates 2a–f. A 50 mg (0.5 mmol) portion of D-4-amino-3-isoxazolidinone were dissolved in 1.2 mL of aqueous 1 M NaHCO_3 and cooled at 0 °C, and 1.0 mmol of the corresponding chloroformate was added; the solution was stirred for 15 h. The obtained precipitates were collected by centrifugation, washed three times with 2 mL of water, and dried under reduced pressure to give compounds **2a–f** subsequently crystallized from benzene.

Synthesis of (R)-Phenyl-3-oxo-4-[(phenoxy)carbonyl]aminoisoxazolidine-2-carboxylate (2a). **2a** was obtained as a white crystalline solid in 85% yield, mp 159–160 °C. ^1H NMR (DMSO- d_6): 8.58 (bs, 1H, D_2O exchange); 7.47 (t, 2H, $J = 7.83$ Hz); 7.41 (t, 2H, $J = 7.58$ Hz); 7.33 (t, 1H, $J = 6.48$ Hz); 7.24 (d, 2H, $J = 7.83$ Hz); 7.23 (d, 1H, $J = 6.70$ Hz); 7.16 (d, 2H, $J = 7.58$ Hz); 5.03–4.97 (m, 1H); 4.80 (t, 1H, $J = 10.15$ Hz); 4.31 (t, 1H, $J = 8.68$ Hz). ^{13}C NMR (DMSO- d_6): 166.4; 154.1; 150.6; 149.5; 145.7; 129.8; 129.4; 126.6; 125.4; 121.6; 121.4; 69.8; 53.0. FT-IR (cm^{-1}): 3344, 1789, 1736, 1713; MS/ESI: $(\text{M} + \text{H})^+$ 343.0914 (m/z).

Synthesis of (R)-4-Methylphenyl-4-[(4-methylphenoxy)carbonyl]amino-3-oxoisoxazolidine-2-carboxylate (2b). **2b** was obtained as a white crystalline solid in 75% yield, mp 169–170 °C. ^1H NMR (DMSO- d_6): 8.48 (bs, 1H, D_2O exchange); 7.25 (d, 2H, $J = 7.83$ Hz); 7.19 (d, 2H, $J = 7.58$ Hz); 7.11 (d, 2H, $J = 7.83$ Hz); 7.01 (d, 2H, $J = 7.58$ Hz); 5.01–4.94 (m, 1H); 4.78 (t, 1H, $J = 10.55$ Hz); 4.29 (t, 1H, $J = 8.69$ Hz); 2.31 (s, 3H); 2.29 (s, 3H). ^{13}C NMR (DMSO- d_6): 166.4; 154.3; 148.4; 145.8; 147.3; 135.9; 134.5; 130.1; 129.7; 121.3; 121.1; 69.8; 53.0; 20.4; 20.3. FT-IR (cm^{-1}): 3339, 1788, 1737, 1713. MS/ESI: $(\text{M} + \text{H})^+$ 371.1085 (m/z).

Synthesis of (R)-4-Methoxyphenyl-4-[(4-methoxyphenoxy)carbonyl]amino-3-oxoisoxazolidine-2-carboxylate (2c). **2c** was obtained as a white crystalline solid in 70% yield, mp 144–145 °C. ^1H NMR (DMSO- d_6): 8.45 (1H, bs, D_2O exchange); 7.16 (d, 2H, $J = 7.83$ Hz); 7.06 (d, 2H, $J = 7.58$ Hz); 6.98 (d, 2H, $J = 7.83$ Hz); 6.93 (d, 2H, $J = 7.58$ Hz); 5.02–4.95 (m, 1H); 4.77 (t, 1H, $J = 10.27$ Hz); 4.28 (t, 1H, $J = 9.23$ Hz); 3.76 (s, 3H); 3.75 (3H, s). ^{13}C NMR (DMSO- d_6): 166.3; 157.3; 156.6; 156.5; 154.5; 144.0; 142.8; 122.5; 122.3; 114.6; 114.3; 69.8; 55.4; 55.3; 53.0. FT-IR (cm^{-1}): 3346, 1790, 1731, 1715. MS/ESI: $(\text{M} + \text{H})^+$ 403.1017 (m/z).

Synthesis of (R)-4-fluorophenyl-4-[(4-fluorophenoxy)carbonyl]amino-3-oxoisoxazolidine-2-carboxylate (2d). **2d**

was obtained as a white crystalline solid in 75% yield, mp 179–180 °C. ^1H NMR (DMSO- d_6): 8.65 (bs, 1H, D_2O exchange); 7.50 (d, 2H, $J = 8.80$ Hz); 7.44 (d, 2H, $J = 8.80$ Hz); 7.29 (d, 2H, $J = 8.80$ Hz); 7.18 (d, 2H, $J = 8.80$ Hz); 5.03–4.97 (m, 1H, Hz); 4.79 (t, 1H, $J = 10.52$ Hz); 4.31 (t, 1H, $J = 8.32$ Hz). ^{13}C NMR (DMSO- d_6): 166.8; 160.0 (d, $J = 243.1$ Hz); 159.5 (d, $J = 242.0$ Hz); 154.6; 147.2 (d, $J = 2.7$ Hz); 146.2; 146.0 (d, $J = 2.7$ Hz); 123.9; 123.8 (d, $J = 8.7$ Hz); 116.9 (d, $J = 23.6$ Hz); 116.4 (d, $J = 23.2$ Hz); 70.2; 53.5. FT-IR (cm^{-1}): 3350, 1740, 1735, 1720. MS/CI: $(\text{M} + \text{H})^+$ 379 (m/z).

Synthesis of (R)-4-Chlorophenyl-4-[(4-chlorophenoxy)carbonyl]amino-3-oxoisoxazolidine-2-carboxylate (2e). **2e** was obtained as a white crystalline solid in 75% yield, mp 160–161 °C. ^1H NMR (DMSO- d_6): 8.65 (bs, 1H, D_2O exchange); 7.52 (d, 2H, $J = 8.32$ Hz); 7.45 (d, 2H, $J = 8.32$ Hz); 7.18 (d, 2H, $J = 8.28$ Hz); 7.15 (d, 2H, $J = 8.28$ Hz); 5.01–4.97 (m, 1H); 4.80 (t, 1H, $J = 9.52$ Hz); 4.30 (t, 1H, $J = 8.32$ Hz). ^{13}C NMR (DMSO- d_6): 167.3; 159.7; 159.6; 155.1; 155.0; 147.7; 147.6; 124.4; 124.3; 117.5; 116.8; 70.7; 53.9. FT-IR (cm^{-1}): 3330, 1740, 1730, 1725. MS/CI: $(\text{M} + \text{H})^+$ 412 (m/z).

Synthesis of (R)-butyl-4-[(benzyloxy)carbonyl]amino-3-oxoisoxazolidine-2-carboxylate (2f). A 0.5 mmol portion of *N*-benzyloxycarbonyl-D-4-amino-3-isoxazolidinone **1f** were dissolved in 2 mL of 1 M aqueous NaHCO_3 and at 0 °C and 1.0 mmol of *n*-butylchloroformate was added. The resulting suspension was stirred at room temperature for 24 h. The obtained white crystalline solid was separated by centrifugation, washed with cold water, dried, and crystallized from benzene. Mp 170–174 °C; yield 75%. ^1H NMR (DMSO- d_6): 7.35 (s, 5H); 5.06 (s, 2H); 4.87–4.80 (m, 1H); 4.62 (t, 1H, $J = 8.08$ Hz); 4.12 (t, 2H, $J = 6.12$ Hz); 4.10 (t, 1H, $J = 9.4$ Hz); 1.61 (m, 2H); 1.35 (m, 2H); 0.88 (t, 3H, $J = 7.90$ Hz). ^{13}C NMR (DMSO- d_6): 167.3; 158.0; 148.3; 143.4; 129.1; 127.6; 127.4; 66.3; 65.9; 65.2; 63.9; 31.1; 19.4; 14.4. FT-IR (cm^{-1}): 3330, 1720, 1715, 1696. MS/CI: $(\text{M} + \text{H})^+$ 336 (m/z).

Synthesis of Amide Derivatives 3h,i. The cyanomethyl anion was generated by electrochemical reduction of anhydrous acetonitrile as previously reported.²¹ Anhydrous acetonitrile, containing 0.1 M tetraethylammoniumhexafluoro phosphate (TEAHFP) was electrolyzed in N_2 atmosphere, at 30 mA cm^{-2} current until 1.0 F/mol was consumed (calculated relative to D-4-amino-3-isoxazolidinone); thereafter, 1.0 mmol of D-4-amino-3-isoxazolidinone were added and the solution stirred for 15 min. A 1.0 mmol portion of alkyl-bromide was added, and the solution was stirred for 1.5 h at room temperature. Removal of the solvent under reduced pressure gave a crude solid that was extracted with three portions of Et_2O , and the combined organic extracts were dried over anhydrous Na_2SO_4 and evaporated under reduced pressure to give a crude residue that was used, without further purification, for the synthesis of amides **3h,i**.

The crude residue (**4h** or **4i**) was dissolved in 20 mL of CHCl_3 and 1.2 mmol of TEA and 1.2 mmol of 3-(trifluoromethyl)benzoyl chloride were added. The obtained solution was stirred at room temperature for 24 h. The solution was washed with saturated Na_2CO_3 , dried over anhydrous Na_2SO_4 and removed under reduced pressure; the obtained residue was purified on silica gel column chromatography to afford pure **3h** and **3i**.

N-[2-(*n*-Octyl)-3-oxoisoxazolidin-4-yl]-3-(trifluoromethyl)benzamide (**3h**) was purified on silica gel column ($\text{CH}_2\text{Cl}_2/\text{ethyl acetate}$, 9:1). Mp 95–96 °C; yield 70%. ^1H NMR (DMSO- d_6): 9.28 (bs, 1H, D_2O exchange); 8.22 (m, 1H, $J = 1.78$ Hz); 8.18 (dd, 1H, $J = 7.71$ Hz, $J = 1.58$ Hz); 7.95 (dd, 1H,

$J = 7.91$ Hz, $J = 2.09$ Hz); 7.74 (m, 1H); 5.09 (m, 1H, $J = 10.15$ Hz); 4.61 (t, 1H, $J = 8.56$ Hz); 4.11 (t, 1H, $J = 8.68$ Hz); 3.50 (d, 2H, $J = 7.25$ Hz); 1.56 (m, 2H); 1.26 (m, 10H); 0.85 (3H, t, $J = 7.00$ Hz). ^{13}C NMR (DMSO- d_6): 166.8; 165.4; 134.7; 132.1; 130.3; 130.0 (q , $J = 32.4$ Hz); 128.7; 125.8; 124.4; 70.2; 52.3; 45.2; 31.7; 29.0; 28.9; 26.8; 26.5; 22.5; 14.4. FT-IR (cm^{-1}): 3301, 1671, 1662. MS/CI: (M + H) $^+$ 387 (m/z).

N -[2-(2,6-Dichlorobenzyl)-3-oxoisoxazolidin-4-yl]-3-(trifluoromethyl)benzamide (**3i**) was purified on silica gel column (CH_2Cl_2 /ethyl acetate, 9:1). Mp 162–164 °C; yield 75%. ^1H NMR (DMSO- d_6): 7.99 (bs, 1H, D_2O exchange); 7.80 (m, 1H, $J = 1.58$ Hz); 7.74 (dd, 1H, $J = 7.55$ Hz, $J = 1.68$ Hz); 7.59 (dd, 1H, $J = 6.77$ Hz, $J = 1.98$ Hz); 7.37 (m, 1H); 7.27 (t, 2H, $J = 6.30$ Hz); 7.24 (d, 2H); 5.15 (s, 2H); 4.97 (m, 1H, $J = 9.87$ Hz); 4.87 (dd, 1H, $J = 8.33$ Hz); 3.98 (t, 1H, $J = 8.68$ Hz). ^{13}C NMR (DMSO- d_6): 166.6; 166.1; 136.9; 133.6; 131.1 (q , $J = 32.9$); 130.7; 129.2; 129.1; 128.7; 128.6; 125.2; 124.7; 72.8; 52.8; 44.7. FT-IR (cm^{-1}): 3253, 1713, 1635. MS/CI: (M + H) $^+$ 433 (m/z).

4.3. Microbiology. The biological activity of the synthesized compounds has been evaluated as reported elsewhere²⁴ toward *M. tuberculosis* H37Rv ATCC 27294.

■ ASSOCIATED CONTENT

Supporting Information

Tables S1–S10 and Figures S1–S17. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: flavio.ballante@uniroma1.it (F.B.); rino.ragno@uniroma1.it (R.R.).

Author Contributions

[§]L.F. and F.B. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Roberto Cirilli (Dipartimento del Farmaco, Istituto Superiore di Sanità, Rome, Italy) for the chiral HPLC analysis. We also thank Prof. A. De Logu from “Dipartimento di Scienze e Tecnologie Biomediche”, University of Cagliari, for the biological activity data. One of us (F.B.) acknowledge Sapienza Università di Roma (grant “Progetti per Avvio alla Ricerca” prot. C26N12JZCT).

■ REFERENCES

- (1) WHO *Global tuberculosis report* 2012. http://www.who.int/tb/publications/global_report/en/index.html (accessed February 21, 2013).
- (2) Ehlers, S. Lazy, dynamic or minimally recrudescens? On the elusive nature and location of the mycobacterium responsible for latent tuberculosis. *Infection* **2009**, *37*, 87–95.
- (3) Di Perri, G.; Bonora, S. Which agents should we use for the treatment of multidrug-resistant Mycobacterium tuberculosis? *J. Antimicrob. Chemother.* **2004**, *54*, 593–602.
- (4) Velayati, A. A.; Masjedi, M. R.; Farnia, P.; Tabarsi, P.; Ghanavi, J.; Ziaiarif, A. H.; Hoffner, S. E. Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran. *Chest* **2009**, *136*, 420–425.
- (5) Udawadia, Z. F.; Amale, R. A.; Ajbani, K. K.; Rodrigues, C. Totally drug-resistant tuberculosis in India. *Clin. Infect. Diseases* **2012**, *54*, 579–581.

(6) Guy, E. S.; Mallampalli, A. Managing TB in the 21st century: existing and novel drug therapies. *Ther. Adv. Respir. Dis.* **2008**, *2*, 401–408.

(7) Bhowruth, V.; Dover, L. G.; Besra, G. S. Tuberculosis chemotherapy: recent developments and future perspectives. *Prog. Med. Chem.* **2007**, *45*, 169–203.

(8) Piccaro, G.; Filippini, P.; Giannoni, F.; Scipione, L.; Tortorella, S.; De Vita, D.; Mellini, P.; Fattorini, L. Activity of drugs against dormant Mycobacterium tuberculosis. *J. Chemother.* **2011**, *23*, 175–178.

(9) Ragno, R.; Marshall, G. R.; Di Santo, R.; Costi, R.; Massa, S.; Rompei, R.; Artico, M. Antimycobacterial pyrroles: synthesis, anti-Mycobacterium tuberculosis activity and QSAR studies. *Bioorg. Med. Chem.* **2000**, *8*, 1423–1432.

(10) Biava, M.; Porretta, G. C.; Poce, G.; Supino, S.; Deidda, D.; Pompei, R.; Mollicotti, P.; Manetti, F.; Botta, M. Antimycobacterial agents. Novel diarylpyrrole derivatives of BM212 endowed with high activity toward Mycobacterium tuberculosis and low cytotoxicity. *J. Med. Chem.* **2006**, *49*, 4946–4952.

(11) Ballante, F.; Ragno, R. 3-D QSAutogrid/R: an alternative procedure to build 3-D QSAR models. Methodologies and applications. *J. Chem. Inf. Model.* **2012**, *52*, 1674–1685.

(12) Manetti, F.; Corelli, F.; Biava, M.; Fioravanti, R.; Porretta, G. C.; Botta, M. Building a pharmacophore model for a novel class of antitubercular compounds. *Farmaco* **2000**, *55*, 484–491.

(13) Barnum, D.; Greene, J.; Smellie, A.; Sprague, P. Identification of common functional configurations among molecules. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 563–571.

(14) Biava, M.; Fioravanti, R.; Porretta, G. C.; Deidda, D.; Lampis, G.; Pompei, R.; Tafi, A.; Manetti, F. New derivatives of tolidine: Synthesis, antitubercular activity and pharmacophore hypothesis. *Med. Chem. Res.* **2002**, *11* (1), 50–66.

(15) Biava, M.; Porretta, G. C.; Poce, G.; De Logu, A.; Saggi, M.; Meleddu, R.; Manetti, F.; De Rossi, E.; Botta, M. 1,5-Diphenylpyrrole derivatives as antimycobacterial agents. Probing the influence on antimycobacterial activity of lipophilic substituents at the phenyl rings. *J. Med. Chem.* **2008**, *51*, 3644–8.

(16) Biava, M.; Porretta, G. C.; Deidda, D.; Pompei, R.; Tafi, A.; Manetti, F. Antimycobacterial compounds. New pyrrole derivatives of BM212. *Bioorg. Med. Chem.* **2004**, *12*, 1453–8.

(17) Biava, M.; Porretta, G. C.; Poce, G.; Deidda, D.; Pompei, R.; Tafi, A.; Manetti, F. Antimycobacterial compounds. Optimization of the BM 212 structure, the lead compound for a new pyrrole derivative class. *Bioorg. Med. Chem.* **2005**, *13*, 1221–30.

(18) Biava, M.; Cesare Porretta, G.; Deidda, D.; Pompei, R.; Tafi, A.; Manetti, F. Importance of the thiomorpholine introduction in new pyrrole derivatives as antimycobacterial agents analogues of BM 212. *Bioorg. Med. Chem.* **2003**, *11*, 515–20.

(19) Biava, M.; Porretta, G. C.; Poce, G.; De Logu, A.; Meleddu, R.; De Rossi, E.; Manetti, F.; Botta, M. 1,5-Diaryl-2-ethyl pyrrole derivatives as antimycobacterial agents: design, synthesis, and microbiological evaluation. *Eur. J. Med. Chem.* **2009**, *44*, 4734–8.

(20) Stammer, C. H.; Kartha, C. C.; Chaturvedi, N. C.; McKinney, J. D. Cycloserine derivatives. *J. Med. Chem.* **1970**, *13*, 1013–5.

(21) Feroci, M.; De Vita, D.; Scipione, L.; Sotgiu, G.; Tortorella, S. Electrogenated acetonitrile anion induced selective N-alkylation of bifunctional compounds. *Tetrahedron Lett.* **2012**, *53*, 2564–2567.

(22) Jain, A. N. Ligand-based structural hypotheses for virtual screening. *J. Med. Chem.* **2004**, *47*, 947–61.

(23) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–91.

(24) Castagnolo, D.; De Logu, A.; Radi, M.; Bechi, B.; Manetti, F.; Magnani, M.; Supino, S.; Meleddu, R.; Chisu, L.; Botta, M. Synthesis, biological evaluation and SAR study of novel pyrazole analogues as inhibitors of Mycobacterium tuberculosis. *Bioorg. Med. Chem.* **2008**, *16*, 8587–91.

**Abstracts presented at the 13th International Congress
on Amino Acids, Peptides and Proteins**

Galveston, Texas

October 5–7, 2013

President: Guoyao Wu

immunity of male reproductive tract but also regulate sperm fertilization process by interacting heparin.

L-Amino acids as chiral auxiliaries to develop reagents for enantioseparation of DL-selenomethionine by reversed-phase high-performance liquid chromatography

Hariom Nagar and Ravi Bhushan

Department of Chemistry, Indian Institute of Technology Roorkee, Roorkee – 247667, India

Optically pure amino acids L-Ala, L-Val, L-Leu and S-methyl-L-Cysteine were used to synthesise chiral derivatizing reagents with fluoro dinitro benzene as the chromophore; these are FDNP-L-Ala, FDNP-L-Val, FDNP-L-Leu and FDNP-SMLC. The reagents were characterised using UV, IR, CHN, and ¹H NMR. Diastereomers of selenomethionine were synthesized with the nucleophilic substitution of remaining fluorine atom in these CDRs under microwave irradiation for 55 s at 75 % (of 800 W) and also by stirring for 50 min at 45 °C. The diastereomers were enantioseparated by reversed-phase high-performance liquid chromatography on a C₁₈ column with detection at 340 nm using gradient elution with mobile phases containing *aq* TFA (0.1 %)-MeCN and by reversed-phase thin layer chromatography with mobile phases containing *aq* TEAP (50 mM)-MeCN in different compositions. The conditions of derivatization and chromatographic separation were optimized. The method was validated for accuracy, precision, limit of detection and limit of quantification.

Computational strategies to design new highly potential BSAO polyamine substrates

Flavio Ballante^{1,*}, Giampiero Tempera¹, Enzo Agostinelli¹, Rino Ragno²

²Rome Center for Molecular Design, Department of Drug Chemistry and Technologies, “Sapienza” University of Rome, P.le A. Moro 5, 00185 Rome, Italy;

¹Istituto Pasteur Fondazione Cenci Bolognietti, Department of Biochemical Sciences “A. Rossi Fanelli”, “Sapienza” University of Rome and CNR, Biology and Molecular Pathology Institute, Piazzale Aldo Moro 5, 00185 Rome, Italy

Natural polyamines putrescine, spermidine and spermine are ubiquitous polycationic compounds present in significant amounts in nearly every prokaryotic and eukaryotic cell type. Spermidine and spermine primarily exist in aqueous solution at pH 7.4 as fully protonated polycations. Such ubiquitous chemical entities play an important role in cell growth and proliferation, in the synthesis of proteins and nucleic acids, in both normal and cancer cells. Preliminary structure based (SB) studies through the AutoDock suite were performed on 25 among natural polyamines and newly synthesized and biologically assayed polyamine analogs in order to clarify their binding modes. Further investigations through a combined approach of docking and 3-D QSAR and COMBINE procedures, named 3-D QSAutogrid/R and COMBINEr respectively, are in due course to rationalize in a multi-informative scenario the different activity profiles and derive a useful pharmacophoric frame able to weight the different ligand-residues interactions magnitudes. Such approach will be useful for the development of novel compounds endowed of both higher potency and selectivity. As future perspective, these molecules will be assayed

alone or in combination with BSAO on several cancer cells, with the aim to evaluate their cytotoxic effects that could be taken into consideration as new approach in anti-cancer therapy. Details and methodologies will be reported.

Protective effect of glutamic acid against oxidative stress in rats

N.O. Salyha

Institute of Animal Biology NAAS, V. Stus street, 38 Lviv, Ukraine 79034

Metabolic processes occurring in animal and human organisms at the times of stresses and diseases lead to the use of a large amount of L-glutamic acid. White male Wistar rats (3 months old), 200–220 g body weight, were divided in three groups. Each group consisted of 10 animals. Each animal received 20 g of food per day. Animals of all groups were healthy. The rats were sacrificed under anesthesia 4 weeks after. This study was conducted to determine effects of supplementing L-glutamic acid to the standard rodent diet (containing 17 % crude protein) on activities of antioxidant enzymes (catalase, superoxide dismutase, glutathione peroxidase, glutathione reductase) and the intensity of lipid peroxidation in tissues of rats. The supplemental doses of L-glutamic acid were determined on the basis of crude protein content in the basal diet. Animals received dietary supplementation with 0 % (control group), 10 and 25 % of L-glutamic acid (second and third group respectively). The obtained data are treated statistically. To determine the probable differences between mean values using Student's test. Results indicated that dietary supplementation with 25 % of L-glutamic acid increased the concentrations of reduced glutathione and glutathione peroxidase activity in the liver, spleen and kidneys, while decreasing the concentrations of lipid hydroperoxides and TBA-active products in these tissues. Supplementation with 10 and 25 % L-glutamic acid enhanced catalase and superoxide dismutase activities in erythrocytes. In all the measured variables, L-glutamic acid supplementation elicited dose-dependent responses. Collectively, dietary supplementation has beneficial effects on antioxidative system, thereby reducing lipid peroxidation.

Ageing: a public health concern and an opportunity for ICAAS

Luc Cynober

Dept of Clinical Chemistry, Cochin and Hôtel-Dieu Hospitals, AP-HP, Paris, and

Dept of Experimental, Metabolic and Clinical Biology, Laboratory of Biological Nutrition EA 4466,

Faculty of Pharmacy, Paris Descartes University, Paris, France

In France, a third of the population will be aged over 60 in 2050 *versus* 20 % in 2005. In the USA, the number of people aged ≥65 is forecast to double between 2010 and 2050 (from 40 to 88 million). At least a third of this population is very healthy, and 25 % get regular physical exercise. Those who are healthy or relatively healthy want to stay that way, and so are frequently dietary supplement users, although the benefits and risks remain uncertain, particularly at high intake levels. Among the dietary supplements available, amino acids (AAs) are of special interest in this sub-population, as they target functions like muscular and cognitive function that typically decline with age, and some AAs have

Acknowledgements

During my doctoral experience I had the pleasure to meet and work together with many people whom I am infinitely grateful for their friendship, support and contribution, making my PhD a memorable experience

First of all, I would like to express all my gratitude to my supervisor, Professor Rino Ragno whose expertise, understanding, support, guidance, enthusiasm, and (a lot of) patience were of fundamental importance for my doctoral experience and especially for my inner growth.

A very special thanks goes out to Professor Gilbert Kirsch and Professor Garland R. Marshall for giving me the opportunity to work in their research groups, for their invaluable scientific support, and for letting me spend many unforgettable moments. Certainly, I was very lucky to work with them.

I acknowledge my gratitude to Drs. Charlène Gadais, Zhanjie Xu, Stéphanie Hesse, Peter Harris, Jayne Marasa, Chris Ho, Nandarapu Damodara Reddy for their friendship, and invaluable support: it was really nice and very interesting to work with them.

I am very grateful to Alexandros Patsilidakos, Adele Pirolli, Alberto De Petris, Richard (Tripp) Wickersham, and all the RCMDers: they provided a friendly and cooperative atmosphere at work and also useful feedback and insightful comments on my work.

I am thankful to Professor Enzo Agostinelli for giving me the opportunity to introduce myself in the field of natural polyamines, and to participate in many international scientific conferences.

I would like to acknowledge all my friends and my family who supported me during these three years.

I would especially like to thank my wife, Emanuela, without whose love, support, and patience all of that would not have been possible.