Data and test mining

DRUDIT, Web-Based DRUgs DIscovery Tools to Design Small Molecules as Modulators of Biological Targets

Antonino Lauria[1]*, Salvatore Mannino[1], Carla Gentile[1], Giuseppe Mannino[1], Annamaria Martorana[1], and Daniele Peri[2]

[1]Dipartimento di Scienze e Tecnologie Biologiche Chimiche e Farmaceutiche "STEBICEF" – University of Palermo, Viale delle Scienze – Ed. 17 -90128 Palermo, Italy, [2]Dipartimento di Ingegneria, University of Palermo, Viale delle Scienze Ed. 6, I-90128 Palermo, Italy.

*To whom correspondence should be addressed.

Abstract

Motivation: New in silico tools to predict biological affinities for input structures are presented. The tools are implemented in the DRUDIT (DRUgs DIscovery Tools) web service. The DRUDIT biological finder module is based on molecular descriptors that are calculated by the MOLDESTO (MOLecular DEScriptors TOol) software module developed by the same authors, which is able to calculate more than one thousand molecular descriptors. At this stage, DRUDIT includes 250 biological targets, but new external targets can be added. This feature extends the application scope of DRUDIT to several fields. Moreover, two more functions are implemented: the Multi-Target and On/Off-Target tasks. These tools applied to input structures allow for predicting the polypharmacology and evaluating the collateral effects.**

Results: The applications described in the paper show that DRUDIT is able to predict a single biological target, to identify similarities among biological targets, and to discriminate different target isoforms. The main advantages of DRUDIT for the scientific community lie in its ease of use by worldwide scientists and the possibility to be used also without specific, and often expensive, hardware and software. In fact, it is fully accessible through the WWW from any device to perform calculations. Just a click or a tap can start tasks to predict biological properties for new compounds or repurpose drugs, lead compounds, or unsuccessful compounds. To date, DRUDIT is supported by 4 servers each able to execute 8 jobs simultaneously.**

Availability: The web service is accessible at the www.drudit.com URL and its use is free of charge.**

Contact: antonino.lauria@unipa.it

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

In the last decades the *in silico* approach to design and optimization of lead compounds, such as a new drugs, and even for repurposing old drugs (i.e., repositioning) or unsuccessful lead compounds, has attracted the interest of the scientific community.

Even if this approach has been adopted in a wide number of medicinal chemistry studies and it is rather consolidated, to date the robustness of *in silico* methods is still a crucial issue. The high complexity of both the

intracellular and the extracellular environment in animal tissues requires that multiple parameters and their interaction be considered. These are the minimum conditions to predict biological activity with acceptable reliability. In previous studies we devised different *in silico* protocols to predict the biological properties of several molecules. There are different methods reported in literature that propose tools based on descriptors of biological targets (González-Díaz,H. et al., 2011), or that encode the experimental conditions in transformed molecular descriptors, or even that mixed ligand and target information into new descriptors (Ferreira da Costa, J. et al., 2018). Our approaches are based only on molecular

descriptors of ligands, without using descriptors based on the experimental conditions. This allows to perform analysis with no request of experimental data and to run calculations efficiently. Among them, in 2011, the Virtual Lock and Key (VLAK) protocol was proposed as a theoretical method to foresee the biological target of input molecules exploiting the *in silico* adaption of the well-known Fisher model (Lauria et al., 2011). In that protocol, molecular descriptors were used to building a template simulating a biological-target "lock". A subsequent matching between molecular structures and the virtual biological locks gave an account of their binding properties.

The above-mentioned chemoinformatic means are ligand-based types and, for this reason, they considerably reduce the computation time with respect to structure-based studies or the cases in which it is necessary to analyze a huge number of structures.

The protocols that can be applied, through this approach, span from the simple and traditional *in silico* screening on single target to the more sophisticate and fashionable on-target/off-target and multitarget approaches (Lauria et al., 2016). These methods are developed in the present work through the **Biotarget Predictor Module**, implemented in the new **DRUDIT** (**DRU**g **DI**scovery **T**ools, https://www.drudit.com)

web service. The main features of DRUDIT are its ease of use, accessibility to researchers worldwide (by personal computers, tablets, or smartphones) and the possibility to perform in silico biological evaluations by common personal devices (e.g., personal computers, tablets, or smartphones) instead of expensive hardware and software. The motto that can be found in the web home page is "just a click or a touch and we make your compound a lead compound". The proposed tools are based on molecular descriptors and represent the evolution of the previous methods (Lauria et al., 2011, Lauria et al., 2014) with online availability. The **MOL**ecular **DES**criptor **TO**ol (**MOLDESTO**) that we implemented in DRUDIT is able to calculate more than one thousand molecular descriptors (Supplementary material, S1).

2 Methods

This section describes the **Biotarget Predictor Module**, as currently available in DRUDIT (the user manual is available on the DRUDIT login page, https://www.drudit.com). The tool is meant to predict the biological affinity of the given input structure **Xi** against the biological target **Ti**. Prediction is carried out following the steps depicted in Figure 1.
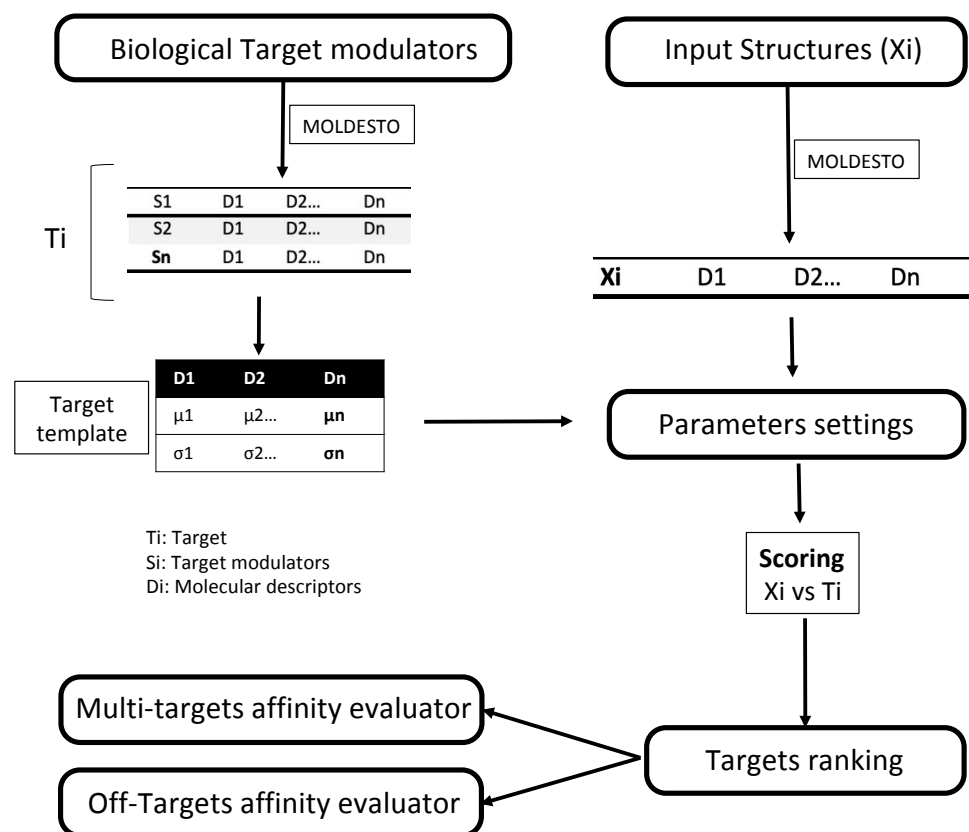

Fig. 1. Flowchart of the Biotarget Predictor Module as available in **DRUDIT**.

The set of known biological target modulators is processed by MOLDESTO to build the biological target template. The full set of molecular descriptors is calculated, and then from the output matrix (modulators *versus* molecular descriptors) mean (μ_i) and standard deviation (σ_i) values for each molecular descriptor are computed.

The pair of values μ_i and σ_i defines the biological target template **Di** (Figure 1). Independently, the input structures are submitted to **MOLDESTO** to calculate the full set of molecular descriptors. The molecular descriptors sequence is matched to the template of the biological target as reported below.

The **D**RUDIT **A**ffinity **S**core (**DAS**) of the input structures versus the biological targets is then computed according to the choices for the three

input parameters **N**, **Z**, and **G** which, respectively, control the number of dynamically selected molecular descriptors (**N**), the maximum allowed percentage of unavailable values (zeros) per molecular descriptor (**Z**), and the Gaussian smoothing function (**G**).

In details, **N** captures the relevance of each molecular descriptor **Di** to the biological target set. The weight is assigned by considering if the descriptor value for a structure, belonging to the biological target set, is in the range □□□□□ ± □□□□□. In this case the assigned score is 1, otherwise is 0. The normalized value of the scores for all the structures gives a measure of the importance of the molecular descriptor **Di** for the

biological target template. Then, the molecular descriptors **Di** are ranked and the top scored **N** are considered. Each molecular descriptor **Di** is considered only if available and if the percentage of zero values in its evaluation for all the structures of the target is above **Z**.

The **G** parameter defines the Gaussian approximation used to score the descriptor values. The matching of the molecular descriptor value **Di** of the input structure **Xi** with the range centered on □i(Di) can be chosen as "large" (**G**=a), "medium" (**G**=b), and "small" (**G**=c). The assigned scores to parameter **G** are reported in Figure 2.
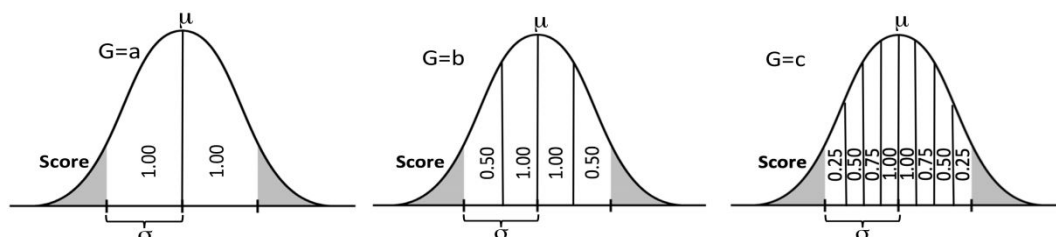

Fig. 2. Score assignation for G parameter.

The **D**RUDIT **A**ffinity **S**core (**DAS**) is finally assigned by weighting the scores (the sum of all values divided by the number of molecular

descriptors) assigned by the chosen G function only for the molecular descriptor selected according to the **Z** and **N** parameters (Figure 3).
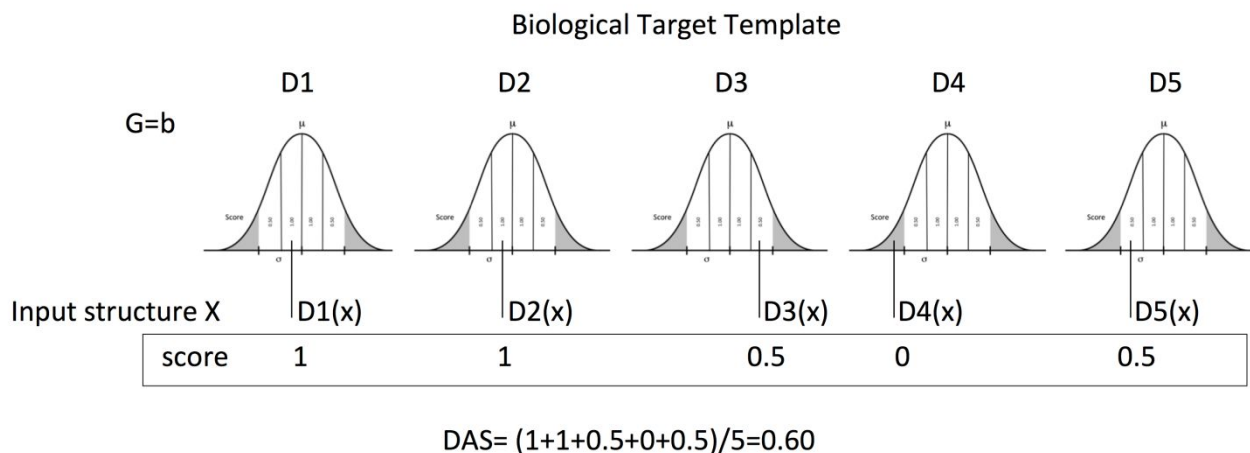

Fig. 3. Example of DRUDIT Affinity Score (**DAS**) calculation on a template of 5 molecular descriptors, previously filtered according to the **Z** and **N** parameters.

The **DRUDIT** parameters are crucial to address the biological target selectivity. In fact, two different actions can be carried out by adjusting **N** and **Z**, and choosing **G**. The first one consists in guiding the search in the entire biological targets database to obtain broad-spectrum results for the input structures. The second one is **DRUDIT** parameters optimization for specific sets of biological targets. The latter action is useful when either multi-target or on/off-targets processing is performed as explained below.

2.1 Implementation and validation of Biotarget Predictor Tools

Currently, the tools include the 250 biological targets (supplementary material, S2) that were used to validate the protocol. The structures of the modulators for the selected biological targets were downloaded from BindingDB (Liu et al. 2007) and filtered as described in the Materials and Methods section.

2.2.1 Cross-validation and biological target similarity identification

In order to validate the **Biotarget Predictor Tools** and to search the ligand-based biological targets similarity, all the sets of modulators for the included biological targets (supplementary material, S3) were submitted to the DRUDIT Biotarget Predictor tool to find the average **DAS** for each of them. To optimize the search, all the biological targets modulators were calculated by setting the parameters (**N**, **Z**, and **G**) as reported in Table 1.

Table 1. Number of biological targets properly assigned by varying DRUDIT parameters (**N**, **Z**, and **G**).

		G		
Z	**N**	a	b	c
50	**200**	125	120	70

	500	133	121	88
	800	99	111	80
	200	113	99	65
100	500	115	119	75
	800	93	95	74

The resulting 18 matrices reported the average **DAS** values for all the biological targets *versus* the biological target templates. The average **DAS** calculated for each biological target template was considered for each set of DRUDIT parameters. The selection was carried out by identifying the set of DRUDIT parameters for which the number of correctly assigned biological targets was higher (Table 1, Z=50, N=500, G=a).

Analyzing the final matrix of the results obtained with the selected DRUDIT parameters (Z=50, N=500, G=a), it was possible, besides checking the consistency of DRUDIT in assigning the right biological target, to correlate biological targets showing higher similarity. For example, from the full matrix (supplementary material, S4) it emerged that cross-validation gave the best or the highest scores, in almost all the cases, along the diagonal (as expected). A few cases showed that specific biological targets are strictly related (for example in supplementary material – S4: Ribosomal protein S6 kinase alpha-2, Stearoyl-CoA desaturase-1, Tyrosine-protein kinase BTK). This aspect can be successfully used in polypharmacology evaluations[2] especially when these biological targets play in the same cascade/pathways.

It is thus evident that when the cross-averaged **DAS** differs significantly from the mean of all the DAS values of dynamically selected molecular descriptors, the biological target can be selectively modulated (for example in supplementary material, S4: Cytochrome P450 2A6, Hepatitis C virus non-structural 3/4A protein, Carbonic Anhydrase II, Metabotropic Glutamate Receptor 5). This gives only an account on the features and capabilities of DRUDIT that can be used to analyze large amounts of data producing output that can be further analyzed by structure- based methods, although, as shown below, these not always give the best results.

2.1.2 Validation through external data

A set of 63 known drugs (supplementary material, S5) was submitted to the DRUDIT Biological Affinity Tools. The output results are in quite good agreement with their known biological activities. The DRUDIT parameters were assessed as reported above. The selected drugs are known to be modulators of Adenosine Receptors, Aromatase, Cannabinoid Receptors, CDKs, Dopamine Receptors, EGFRs, Estrogen Receptor, GSK-3, HDACs, p38-MAPK, PI3K, and TGF-beta. The optimization of the parameters was thus focused on such biological targets. The results are reported in Table 1.

Table 2. **DAS** calculation on a set of known drugs.

Biological target	Modulator	DAS	averaged DAS	Biological target	Modulator	DAS	averaged DAS
Adenosine Receptors	Istradefylline	0.94	0.94	Estrogen Receptors	4-Hydroxytamoxifen	0.995	0.94
	Reversine	0.945			Diethylstilbestrol	0.975	
	SCH58261	0.92			Equol	0.895	
Aromatase	Aminoglutethimide	0.88	0.90		Estradiol valerate	0.87	
	Anastrozole	0.87			Estriol	0.915	
	Letrozole	0.96			Hexestrol	0.975	
Cannabinoid Receptors	AM251	0.675	0.72	GSK-3	BIO	0.785	0.85
	BML-190	0.7			BIO-acetoxime	0.875	
	Rimonabant	0.775			SB-216763	0.88	
CDKs	AZD5438	0.815	0.82	HDACs	LMK-235	0.93	0.90
	BMS-265246	0.82			M344	0.89	
	ON123300	0.85			RG2833	0.935	
	PHA-767491	0.79			RGFP966	0.875	
	PHA-793887	0.795			Pracinostat	0.875	
	PHA-848125	0.8			Scriptaid	0.875	
	Ribociclib	0.915			Panobinostat	0.935	
	SU9516	0.785		p38-MAPK	Losmapimod	0.865	0.84
	AT7519	0.845			Skepinone-L	0.815	
Dopamine Receptors	Azaperone	0.975	0.98		SB202190	0.9	
	Domperidone	0.965			VX-702	0.79	

	Droperidol	0.995			PI3K	A66	0.83	0.87
	Haloperidol	0.985				AZD6482	0.91	
	Perphenazine	0.965				AZD8835	0.87	
EGFRs	AZD9291	0.965	0.94			GDC-0032	0.82	
	CL-387785	0.89				GNE-317	0.845	
	Erlotinib	0.9				HS-173	0.88	
	Gefitinib	0.98				Idelalisib	0.905	
	Icotinib	0.91				YM201636	0.86	
	PD168393	0.88			TGF-beta	Galunisertib	0.92	0.91
	Varlitinib	0.94				GW788388	0.915	
	WZ3146	0.995				SB-505124	0.9	
	WZ8040	0.985						

The results show a robust prediction capability of the protocol. In particular, the best performance is obtained for the modulators of Adenosine Receptors, Dopamine Receptors, EGFRs, and Estrogen-progestogen-Receptors, with DAS scores in the range 0.94-0.98. On the other hand, the worst results showed DAS values below the threshold of 0.8. This is the case of cannabinoid receptor ligands, showing an averaged DAS of 0.72. The success of the method, as above underlined, is strictly dependent on both the quality of the selected dataset and the selection of DRUDIT parameters. Thus, in this case, the results can be improved by acting on the modulators that were used in building the target template, or by working only on a specific target with a more focused choice of the input parameters to DRUDIT.

2.2 Multi-Target and On-Target/Off-Targets evaluators

The MultiTarget-Directed Ligand (MTDL) strategy is based on the concept that a single molecular entity can be designed to hit multiple targets that cooperate in the network of the disease. This paradigm has been the focus of increasing research over the last decade (Morphy and Rankovic 2009, Rampa et al. 2011, Leòn et al. 2013). On the other hand, it was estimated that each existing drug binds, on average, to 6.3 biological targets (Mestres et al. 2008). Identification of the off-targets will provide the molecular basis for a new kind of therapies, but it can also lead to better understanding of potential drug side-effects. Moreover, it can suggest drug repurposing in the treatment of different conditions with respect to those originally intended, leading to possible consideration of the studied structures in a polypharmacological drug design environment (Martorana et al. 2016).

Additionally, **DRUDIT** can be used to perform predictions focused on multi-target and on/off-target strategies. By flagging the Multi-Target or the On/Off-Target button the user can select the biological targets to be included in the selected tool. The multi-target and on-target/off-targets evaluations assign the affinity score to each input structure through weighting functions (see DRUDIT user manual at https://www.drudit.com).

3   Results

3.1 Application of Multi-target and On/Off-target tools

DRUDIT Multi and On/Off target tools can be used also to address the issue of selectivity in order to investigate drug molecules able to discriminate two or more isoforms of several biological targets. To this aim we validated the DRUDIT Biological Target Affinity calculator against a set of known Non-Steroidal Anti-Inflammatory Drugs (NSAIDs).

NSAIDs are the most commonly used molecules for the treatment of several inflammatory diseases. The cyclooxygenase enzymes (COXs) are the main biological targets of NSAIDs. COXs catalyze the rate-limiting step in the production of the soluble inflammatory mediators, i.e. prostaglandins. The two isoforms, COX-1 and COX-2, encoded by two different genes, are structurally and functionally close, but they vary in their expression and distribution (Tazawa et al. 1994). COX-1, ubiquitously and constitutively expressed in mammalian tissues, including kidney, stomach, and vascular endothelium, is a housekeeping enzyme involved in physiological adaptation. In particular, its role in the vascular hemostasis and in maintaining the integrity of the gastrointestinal epithelium and the normal renal function is known. On the contrary, COX-2 isoform is expressed at very low levels in physiological conditions but is highly induced by pro-inflammatory stimuli. The different distribution and expression of the two COX isoforms suggests that a selective inhibition of the COX-2 isoform leads to the inhibition of inflammation without interfering with the COX-1 dependent protective effects in the gastrointestinal tissue, kidney, and in blood coagulation (Zarghi and Arfaei 2011).

Within the DRUDIT Biotarget Affinity module, the templates of COX-1 and COX-2 were built using the set of known modulators, as available in bindingDB (Liu et al. 2007). Then, the selected 16 known modulators of COX-1 and COX-2 (Figure 4) were submitted to calculation.
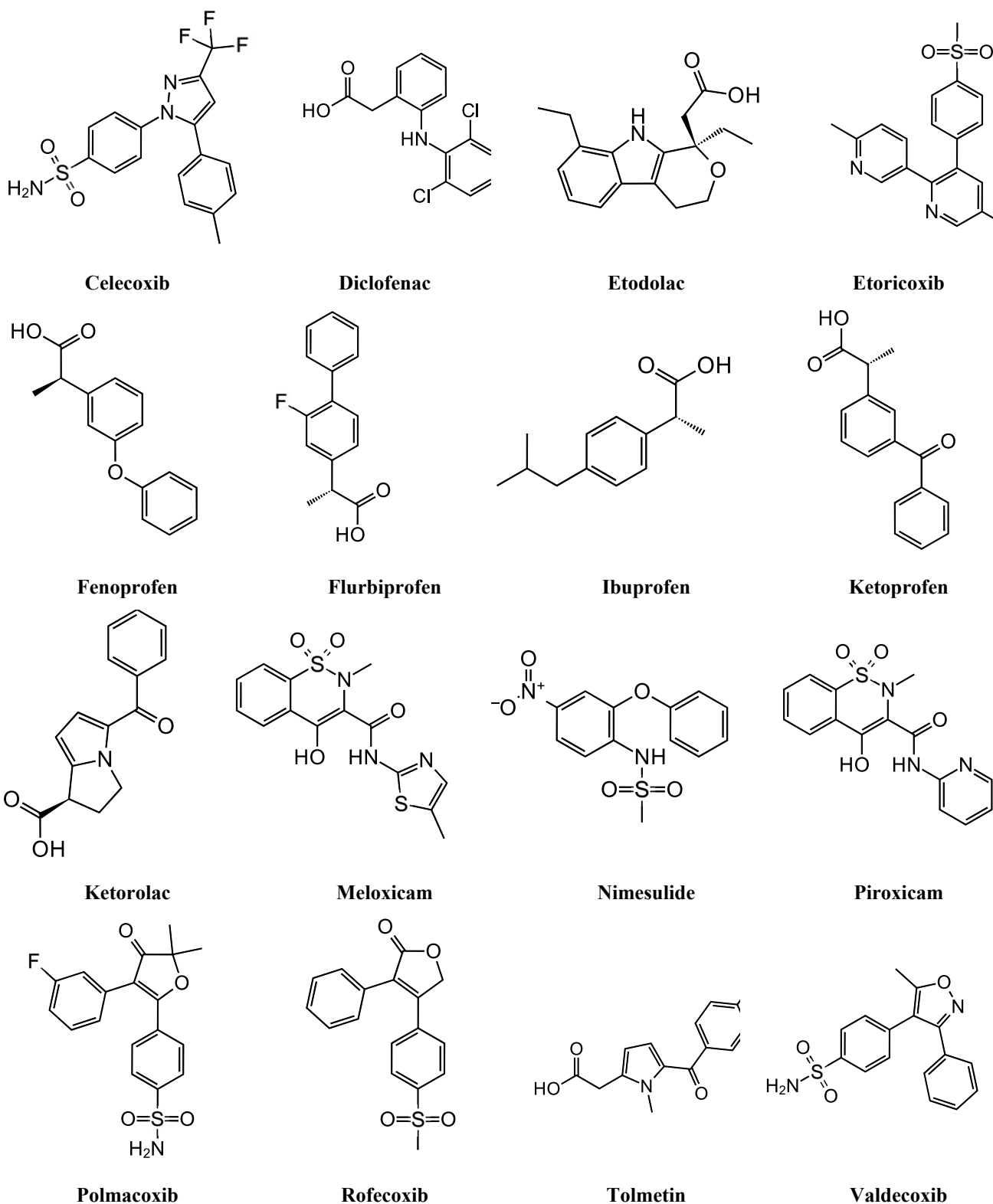
Fig. 4. Selected COXs inhibitors.

Table 3 reports the *in silico* ligand-based evaluations produced by DRUDIT. In particular, calculations were performed using three approaches: On/Off-Target mode, where the On-target is COX-1 and Off-target is COX-2 (Column 1 in table 3); On/Off-Target mode, where the On-target is COX-2 and Off-target COX-1 (Column 2 in table 3); multi-target mode where the biological targets are COX-1 and COX-2. The scores give the affinity of each drug with respect to the combinations reported above (●, high affinity/selectivity; ○, low affinity/selectivity). The COX-1/COX-2 column highlights COX-1

selectivity whereas the COX-2/COX-1 column the COX-2 selectivity. Higher values in the MM column are related to a lower selectivity for the analyzed drugs.

Table 3. On/Off Target and Multi-Target results.

Drug	selectivity	COX-1/COX-2	COX-2/COX-1	MM
Flurbiprofen	COX-1	●1.80	○0.56	●0.36
Tolmetin	COX-1	●1.70	○0.59	●0.30
Ketorolac	COX-1	●1.70	○0.59	●0.29
Ketoprofen	low	●1.62	○0.62	●0.33
Fenoprofen	low	●1.70	○0.59	●0.32
Diclofenac	COX-1	●1.66	○0.60	⊙0.18
Ibuprofen	low	⊙1.36	○0.74	⊙-0.16
Nimesulide	low	⊙1.22	○0.82	⊙-0.07
Etodolac	low	⊙1.10	○0.91	⊙-0.25
Piroxicam	COX-2	⊙1.01	⊙0.99	⊙-0.23
Valdecoxib	low	⊙1.06	⊙0.94	⊙-0.07
Rofecoxib	COX-2	⊙0.93	⊙1.07	⊙-0.17
Meloxicam	COX-2	⊙0.93	⊙1.07	○-0.33
Etoricoxib	COX-2	○0.66	●1.51	○-0.37
Polmacoxib	COX-2	○0.66	●1.51	○-0.44
Celecoxib	COX-2	○0.56	●1.79	○-0.52

● score higher than 70%; ⊙ score in the 30-70% range; ○ score lower than 30%.

The analysis of the scores produced by DRUDIT is widely in agreement with clinical experimental data reported in the literature for the NSAIDs (Al-Hourani et al. 2011). The higher on/off scores clearly confirmed the affinity of the selected drugs with their own COX isoform. The well-known COX-2 inhibitors, such as Celecoxib, Polmacoxib, and Etoricoxib, are the top scored in COX-2/COX-1 column (1.79, 1.51, and 1.50 respectively). Also in the cases of the COX-1 drugs Ketorolac, Tolmetin, and Flurbiprofen the results reported in the COX-1/COX-2 column (1.7, 1.7, and 1.8 respectively) are in good agreement with the experimental data.

Moreover, for such drugs, these results are further confirmed by the low scores in the MM column. In the case of the non-selective drugs Nimesulide and Ibuprofen, the tool was able to successfully predict their expected affinity. These results proved the ability of DRUDIT in identifying the selective behavior of a set of known or unknown molecular structures against a panel of several target isoforms.

With the aim to compare the DRUDIT Biotarget predictor results with those arising from classical structure-based methods in predicting selectivity of NSAIDs, the drugs (Figure 4) were processed by Induced Fit Docking (IFD). In particular, PDB-IDs 5GMM and 5GMN (Kim et al. 2016) were selected, from Protein Data Bank (RCSB Protein Data Bank, https://www.rcsb.org/), as 3D structures of COX-1 and COX-2 isoforms respectively, and were submitted to the IFD protocol as described in the Experimental Section. The IFD output results are reported in Table 4.

Table 4. Induced Fit Docking results.

Drug	DS COX-1	DS COX-2	IFD-Score COX-1	IFD-Score COX-2
Etoricoxib	⊙-8.06	⊙-7.73	●-505.8	●-511.7
Valdecoxib	●-8.72	●-9.09	●-504.4	●-510.3
Fenoprofen	⊙-8.22	○-7.2	⊙-503.8	●-509.1
Tolmetin	⊙-8.21	⊙-8.16	⊙-503.2	●-509
Piroxicam	○-7.65	⊙-7.46	○-501	●-508.8
Nimesulide	○-7.77	⊙-7.33	⊙-503.4	⊙-508.7
Etodolac	●-8.66	⊙-8.07	●-506	⊙-508.6
Flurbiprofen	○-7.34	●-8.64	○-501.1	⊙-507.9
Celecoxib	●-9.00	⊙-8.04	⊙-502.2	⊙-507.9
Meloxicam	○-7.53	○-6.68	○-500.4	⊙-507
Ketorolac	○-7.81	○-7.23	○-500.2	⊙-507
Diclofenac	⊙-8.32	⊙-7.99	⊙-503.6	⊙-506.9
Ketoprofen	●-8.55	⊙-7.42	○-501.4	⊙-506.6
Rofecoxib	⊙-8.08	○-6.57	⊙-503.2	⊙-506.4
Ibuprofen	○-7.67	○-7.14	○-501.6	⊙-506.3
Polmacoxib	⊙-7.96	●-8.44	●-505.9	○-501.8

● score higher than 70%; ⊙ score in the 30-70% range; ○ score lower than 30%.

IFD failed in predicting the appropriate isoforms for many structures. Only in the case of Etoricoxib (Figure 5, left) and Valdecoxib (Figure 4, right) the IFD scores are in agreement with experimental data.
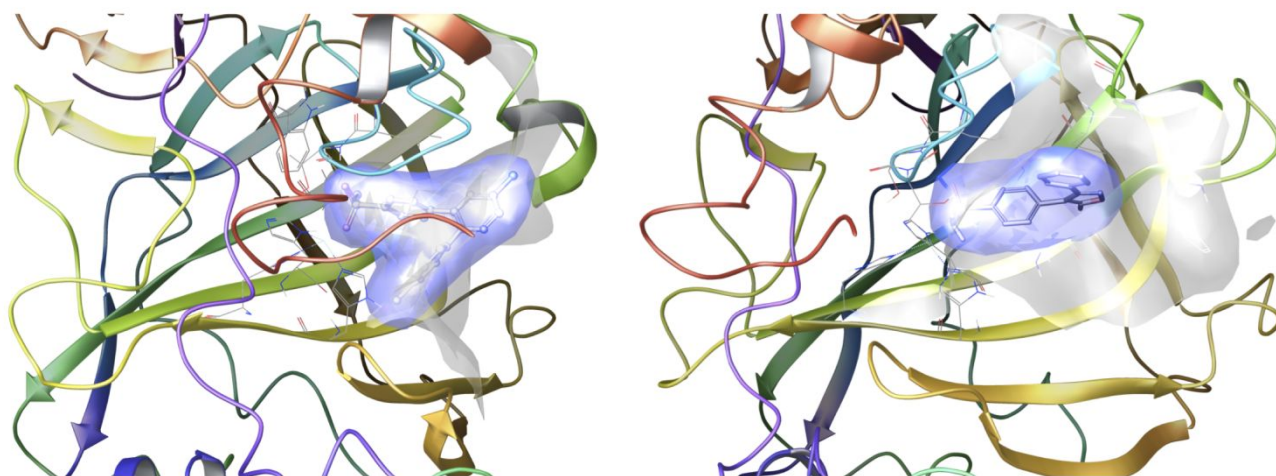

Fig. 5. IFD poses of Etorixoxib (left) and Valdecoxib (right) in the COX-2 binding site.

These results show that the molecular descriptors approach (DRUDIT), in this case of study, was able to identify the selectivity of NSAIDs better than the structure-based method (IFD). This can be attributed to the fact that the structure-based methods do not take into account multiple factors such as pharmacokinetics and pharmacodynamics that are notoriously not efficaciously evaluated. DRUDIT instead, relying on models built from drugs whose pharmacokinetics and pharmacodynamics processes are known, has the ability to overcome these issues.

3.2 External Biological Targets implementation

DRUDIT allows the implementation of new biological targets in local databases (see DRUDIT user manual, available on the DRUDIT login page, https://www.drudit.com). The application on HSP90 is here reported. In a previous work (Lauria et al. 2013) we proposed new amino-cyanopyridines **1a-k** as modulator of HSP90 (Table 5). Thus, here we used DRUDIT to evaluate these structures by building a template with HSP90 known inhibitors, as available in BindingDB (Supplementary Material, S6). The amino-cyanopyridines submitted to the HSP90 template gave the results reported in Table 5. These results were obtained by setting DRUDIT parameters after the optimization process (N= 700; Z= 50; G=b, Supplementary Material, S7).

Table 5. DRUDIT results on amino-cyanopyridines **1** as modulator of HSP90.
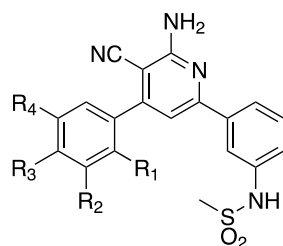
1a-k

Cmd	R_1	R_2	R_3	R_4	*Hsp90 binding	DAS
1a	Cl	H	H	H	950 ± 35	0.834
1b	H	H	NMe_2	H	371 ± 28	0.832
1c	H	F	H	H	$>10^4$	0.722
1d	OMe	H	H	H	634 ± 22	0.844
1e	H	OMe	OMe	H	867 ± 27	0.81
1f	H	H	OMe	H	541 ± 29	0.848
1g	H	H	OH	H	$>10^4$	0.786
1h	OH	OMe	H	H	252 ± 23	0.816
1i	H	OMe	H	H	$>10^4$	0.798
1j	H	H	NEt_2	H	783 ± 19	0.824
1k	Br	H	H	H	$>10^4$	0.752

The DRUDIT output is coherent with the experimental data. In fact, the most active derivatives are **1b**, **1f**, **1h**, **1e**, whereas the least active derivatives are **1c**, **1g**, and **1k**. In general, the most active compounds fall in the range of the 90% top scores in the full biological targets results (Supplementary Material, S7).

4 CONCLUSIONS

DRUDIT implements a protocol as a combination of different algorithms working on biological targets databases. The robustness of the calculation is deeply related to the quality of the biological data used as a training set. The great advantage of the WEB service is its ease of use and World Wide Web, 24/7 availability. Every scientist, even with no

molecular modeling background, can use it through a simple WEB browser without investing in demanding hardware and expensive software. The theoretical aspects that are focused on biological targets can be applied also to other fields. In fact, the "external biological target" could belong to any area (either biological or not) for which molecular structures are known to be the modulators. The applications reported in this work showed that DRUDIT is able not only to evaluate the affinity for the input structure against biological targets, but also to approach the polypharmacology (Multi-Target mode) and the opposite effects (On/Off-Target mode) aspects. Moreover, we demonstrated that the Biological Target module is able to predict similarities among biological targets. This, in turn, can support predicting the collateral effects or identify biological targets set useful in polypharmacology of candidate drugs.

5 MATERIALS AND METHODS
5.1 Hardware
The DRUDIT WEB service runs on 4 servers that are automatically selected according to the number of jobs and online availability. Each server can support up to 10 simultaneous jobs, while the exceeding jobs are placed in a queue. **Software**: DRUDIT consists in several software modules implemented in C and JAVA running on MacOS Mojave.

5.2 MOLDESTO: a new MOLecular DEScriptors TOol
DRUDIT is based on molecular descriptors and represents the evolution of previous automated and on-line available tools (Lauria et al. 2011, Lauria et al. 2014). The MOLecular DEScriptor TOol (MOLDESTO) that we implemented in DRUDIT is a new tool currently able to deal with more than one thousand and four hundred molecular descriptors (Supplementary Material, S3).

MOLDESTO is able to read common molecules file formats, such as SMILES, SDF, Inchi, Mdl, Mol2, to optimize structures, and is provided with a caching system to boost the calculation speed of previously submitted structures. Input structures can be drawn in the web application or uploaded to the server as external files. In either case, structures are optimized by MOPAC before being processed by MOLDESTO.

5.3 Databases
Binding database (BindingDB): It focuses on Ki, Kd, IC50 and EC50 values, related to a well-defined protein target. BindingDB contains 1,142,124 binding data for 7,032 protein targets and 495,498 small molecules. In particular 2,291 protein-ligand crystal structures, with BindingDB affinity measurements for proteins with 100% sequence identity, and 5,816 crystal structures, matching proteins to 85% sequence identity, are included in the database. A large amount of these data derives from open databases as ChEMBL (Gaulton et al. 2012) and PubChem (Wang et al. 2009). DRUDIT includes 250 biological targets from BindingDB. Modulators for each target were selected by considering the IC50 cut-off of 0.1 □M.

5.4 Induced fit docking
Induced fit docking simulation was performed using the IFD application as available (Sherman et al. 2006) in the Schrödinger software suite (Maestro, version 9.3, Schrödinger, LLC, New York, NY, 2012), which was demonstrated to be an accurate and robust method to account for both ligand and receptor flexibility (Zhong et al. 2009). The atomic co-ordinates for the isoforms of COXs were downloaded from the Protein Data Bank (PDB id 5GMM and 5GMN, COX1 and COX2 isoforms respectively) and submitted to the Protein Preparation Wizard module in Schrödinger as follows: adding hydrogen, assigning partial charges (using the OPLS-2001 force field) and protonation states. All crystal waters were removed.

The IFD protocol was carried out as follows (Wanga et al. 2008, Luo et al. 2013): the ligands were docked into the rigid receptor models with scaled-down van der Waals (vdW) radii. The Glide Standard Precision (SP) mode (Friesner et al. 2004) was used for the docking, and 20 ligand poses were retained for protein structural refinements. The docking boxes were defined to include all amino acid residues within the dimensions of 25 Å × 25 Å × 25 Å from the center of the original ligands; the induced-fit protein–ligand complexes were generated using the Prime software (Jacobson et al. 2004). The 20 structures from the previous step were submitted to side chain and backbone refinements. All residues with at least one atom located within 5.0 Å of each corresponding ligand pose were included in the refinement by Prime. All the poses generated were then hierarchically classified, refined and further minimized into the active site grid before being finally scored using the proprietary GlideScore function, defined as: GScore= 0.065*vdW + 030*Coul + Lipo + Hbond + Metal + BuryP + RotB + Site, where: vdW is the van der Waals energy term, Coul is the Coulomb energy, Lipo is a Lipophilic contact term which rewards favorable hydrophobic interactions, Hbond is a H-bonding term, Metal is a metal-binding term (where applicable), BuryP is a penalty term applied to buried polar groups, RotB is a penalty for freezing rotatable bonds and Site is a term used to describe favourable polar interactions in the active site.

Finally, IFD score (IFD score = 1.0 Glide_Gscore + 0.05 Prime_Energy), which accounts for both protein–ligand interaction energy and total energy of the system, was calculated and used to rank the IFD poses. The more negative is the IFDscore, the more favorable is the binding.

SUPPLEMENTARY MATERIAL

S1, molecular descriptors list implemented in MOLDESTO; S2, selected biological targets; S3, modulators of the selected targets, in MOL format (due to file-size constraints not included as supplementary material files) can be downloaded at www.drudit.com/S3.zip; S4, full biological targets matrix results; S5, selected known drugs in MOL file format; S6, HSP90 modulators in MOL file; S7, DAS output for amino-cyanopyridine derivatives.

References

Al-Hourani,B.J. et al. (2011) Cyclooxygenase-2 inhibitors: a literature and patent review (2009 – 2010). *Expert Opin Ther Pat.* **21**, 1339-1432.

Ferreira da Costa, J. et al. (2018) Perturbation Theory/Machine Learning Model of ChEMBL Data for Dopamine Targets: Docking, Synthesis, and Assay of New l-Prolyl-l-leucyl-glycinamide Peptidomimetics. *ACS Chem. Neurosci.* **9(11)**, 2572-2587.

Friesner,R.A. et al. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–1749.

Gaulton,A. et al. (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–1107.

González-Díaz,H. et al. (2011) NL MIND-BEST: a web server for ligands and proteins discovery--theoretic-experimental study of proteins of Giardia lamblia and new compounds active against Plasmodium falciparum. *J. Theor. Biol.* **276**, 229-249.

Jacobson, M.P. et al. (2004) A hierarchical approach to all-atom protein loop prediction. *Proteins* **55**, 351–367.

Kim,H.T. et al. (2016) Structural insight into the inhibition of carbonic anhydrase by the COX-2-selective inhibitor polmacoxib (CG100649). *Biochem Biophys Res Commun.* **478**, 1-6.

Lauria,A. et al. (2011) Virtual Lock-and-Key Approach: The In Silico Revival of Fischer Model by Means of Molecular Descriptors. *Eur. J. Med. Chem.* **46**, 4274-4280.

Lauria,A. et al. (2013) Synthesis and Biological Activities of a New Class of Heat Shock Protein 90 Inhibitors, Designed by Energy-Based Pharmacophore Virtual Screening. *J. Med. Chem.*, **56**, 3424-3428.

Lauria,A. et al. (2014) Multivariate analysis in the identification of biological targets for designed molecular structures: the BIOTA protocol. *Eur. J. Med. Chem.* **75**, 106-110.

Lauria,A. et al. (2016) Drugs Polypharmacology by In Silico Methods: New Opportunities in Drug Discovery. *Curr. Pharm. Des.* **22**, 3073-3081.

Leòn,R. et al. (2013) Recent advances in the multitarget-directed ligands approach for the treatment of Alzheimer's disease. *Med. Res. Rev.* **33**, 139–189.

Liu,T. et al.(2007) BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **35**, 198-201.

Luo,H.J. et al. (2013) Induced-fit docking and binding free energy calculation on furostanol saponins from Tupistra chinensis as epidermal growth factor receptor inhibitors. *Med. Chem. Res.* **22**, 4970–4979.

Martorana,A. et al. (2016) The Repurposing of Old Drugs or Unsuccessful Lead Compounds by in Silico Approaches: New Advances and Perspectives. *Curr. Top. Med. Chem.* **16**, 2088-2106.

Mestres,J. et al. (2008) Data completeness--the Achilles heel of drug-target networks. Nat. Biotechnol. **26**, 983–984.

Morphy,R. and Rankovic,Z. (2009) Designing multiple ligands - medicinal chemistry strategies and challenges. *Curr. Pharm. Des.* **15**, 587–600.

Rampa,A. et al. (2011) A. Hybrid-based multi-target ligands for the treatment of Alzheimer's disease. *Curr. Top. Med. Chem.* **11**, 2716–2730.

Sherman,W. et al. (2006) Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects *J. Med. Chem.* **49**, 534–553.

Tazawa,R. et al. (1994) Biochemical characterization of the genomic structure, chromosomal location and promoter of human prostaglandin h synthase-2 gene. *Biophys. Res. Commun.* **203**, 190–199.

Wang,Y. et al. (2009) PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res. **37**, W623–W633.

Wanga,H. et al. (2008) Induced-fit docking of mometasone furoate and further evidence for glucocorticoid receptor 17α pocket flexibility. *J. Mol. Graph. Model.* **27**, 512–521.

Zarghi,A. and Arfaei, S. (2011) Selective COX-2 Inhibitors: A Review of Their structure-Activity Relationships. *Iran J. Pharm. Res.* **10**, 655–683.

Zhong,H. et al. (2009) Induced-fit docking studies of the active and inactive states of protein tyrosine kinases. *J. Mol. Graph. Model.* **28**, 336–346.