Predicting Molecular Mechanisms of Action Associated with Binary Endpoints

A. Valencia, M. Carbó and J. Prous

[1]*Prous Institute for Biomedical Research, Rambla de Catalunya 135, 3-2, Barcelona, 08008 Spain*

CBI Annual Meeting 2014

Oct 28-30 ,2014, Tokyo

Abstract

The in silico identification of mechanisms of action related to drug candidates can help elucidate therapeutic targets and discover off-target activities that could be linked to adverse effects, providing information to better understand the underlying pathways involved in the efficacy and adverse effects of drugs. Additionally, predicting the mechanism of action of chemicals is of utmost importance to determine their environmental fate, and to predict the effects of long term exposure.

To address these needs, we have used Prous Institute's software solution Symmetry [1] *to train a model that predicts 644 mechanism of action based on a training set extracted from literature and patent analysis. The training set contains approximately 1.5 million Structure-Activity Relationships (SAR) with a ratio of ~1.6 SAR per training set structure, and it is continuously updated. The prediction quality of the Mechanism of Action (MoA) model has been assessed in a 10% hold-out external validation yielding an average recall of 93%.*

One interesting application of Symmetry's MoA model is to predict the mechanisms of action associated with a binary data set and using that information to find differentially expressed mechanisms, i.e. mechanisms associated with positives but unrelated to negatives for a given endpoint. A methodology is proposed to find the relevant mechanisms of action associated with a binary training set, and avoid spurious MoA relationships.

As a case study, a binary data set of brain cancer cell line SNB-78 tumor growth inhibitors has been analyzed and MoAs related to this endpoint have been highlighted. The resulting MoAs associated with the anti-cancer activity in the SNB-78 cell line are compared with those described in the literature.

I. Objectives

- To develop a Mechanism of Action (MoA) model in SYMMETRY to assess the probability of chemical compounds having any of *644* different mechanisms of action.
- To design a probabilistic approach to find the mechanisms of action relevant to any binary endpoint and to distinguish chance relationships from evidence-supported associations between the selected MoAs and the positives of the binary endpoint.
- To show that the problem of finding the relevant MoAs for a binary endpoint, i.e. finding the MoAs that maximize P(MoA|+)/P(MoA|-), is mathematically equivalent and can be mapped into the problem of finding the MoAs that are the better predictors of a positive outcome of the binary endpoint, i.e. maximizing P(+|MoA).
- To illustrate in a case study the identification of relevant mechanisms of action associated with a binary endpoint using SYMMETRY's MoA model, we apply our approach to predict the MoAs associated with the anti-tumoral activity of small molecules in the SNB-78 brain tumor cell line.

II. Predicting MoAs Associated with Chemical Compounds

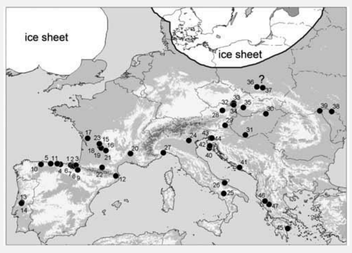
Motivation:

Data Challenges for Creating a MoA Model

MoA data biases:

a) **Inhomogeneity:** MoA data is not homogeneously distributed in chemical space (molecular descriptor space). There is more data in regions of past/present scientific interest than in other regions.
b) **Imbalance:** Data for different mechanisms of action can be very imbalanced. 40 vs. 27,000 compounds per MoA)
c) **Incompleteness:** Known Structure-Activity Relationships (SAR) for mechanisms of action are very incomplete. We only know experimentally a few of the probable SARs for each training set compound.

To create a MoA model unbiased by how past scientific interests have modulated the data that is available today we need to handle the inhomogeneity, imbalance and incompleteness issues.



MoA data, like archeological data, is mostly available for certain 'sites' (regions of chemical space) that have been studied up to this time.

FIGURE 1. Large-mammal archeological sites for the *last glacial maximum* [2].

Methods:

Prous Institute SYMMETRY

SYMMETRY is a computational platform aiming to replicate *in silico* the processes through which new drugs are discovered, developed and approved. A local server installation of SYMMETRY offers scientists a collaborative environment for drug discovery and toxicity screening through an intuitive web interface, a shared database and a reporting system. SYMMETRY is ready for high-throughput computation and optimized for managing and predicting large data sets of chemical compounds through multi-thread execution. Has modeling capabilities that enable construction of customized models and prediction of chemicals using *in silico* characterization of the molecular structures. Models in SYMMETRY can be grouped in 'suites' which yield a consensus prediction (metaclassifiers) that can then be used to support decision-making in the drug discovery process and regulatory review.

Symmetry's MoA model

SYMMETRY's new global Mechanism of Action (MoA) model has been trained with 1.5 million structure-activity relationships and covers 644 mechanisms of action with approx. 1.6 SAR per training set compound. It includes pharmacological targets of therapeutic interest and those related to toxicities or adverse effects; and can be used to predict both on- and off-target interactions of small molecules.

The training set is derived from manually curated data extracted from literature, congresses and patent analysis. The model uses a multi-label classification algorithm developed at Prous Institute and is based on molecular descriptors. SYMMETRY predicts a ranked list of possible MoAs indicating in each case the probability that a test structure interacts with a given target and produces certain action.

The MoA model prediction quality has been assessed in external 10% hold-out validation yielding satisfactory results (93% recall), and obtaining a reasonable number of MoAs per test set structure (5.2 predictions on average).

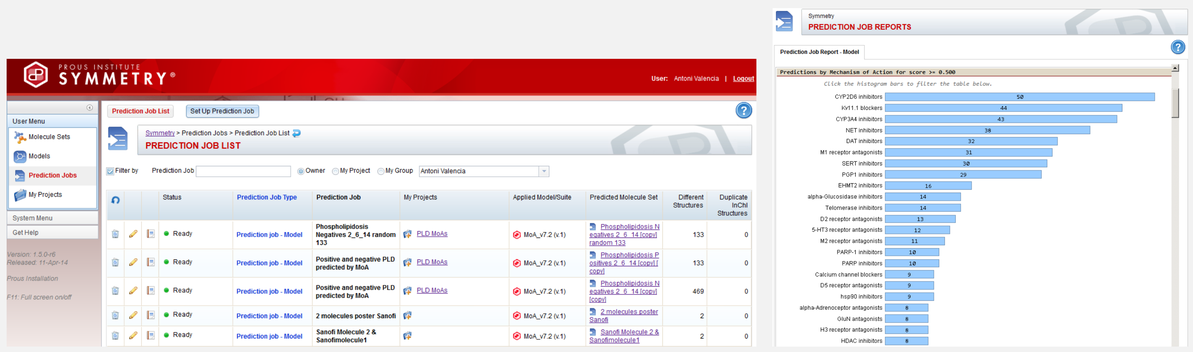


FIGURE 2. SYMMETRY's GUI and MoA prediction report screenshots.

II. Predicting MoAs Associated with Binary Endpoints

Methods:

One interesting application of the MoA model is to predict the mechanisms of action associated with a data set corresponding to a binary endpoint and using that information to find differentially expressed mechanisms, i.e. mechanisms highly associated with positives (compared to negatives) of the given endpoint.

Differentially expressed MoAs are characterized by a ratio P(MoA|+)/P(MoA|-)>1. Using Bayes theorem of conditional probability, one can show that

$$\frac{P(MoA|+)}{P(MoA|-)} = \frac{P(+|MoA)(1-P(+))}{(1-P(+|MoA))P(+)} \quad . \qquad (1)$$

P(MoA|+) and P(MoA|-): *Conditional probability of having certain MoA for a compound which is positive/negative for the binary endpoint.*
P(+|MoA): *Conditional probability of being positive for the binary endpoint for a compound that has certain MoA.*
P(+): *Probability that a compound is positive for the binary endpoint (0.5 for a balanced data set).*

Since Eq. (1) is a growing function of P(+|MoA), if we find the MoAs that maximize P(+|MoA), we automatically solve the problem of finding the differentially expressed MoAs associated with the positives of the binary endpoint. Additionally, since a confidence interval of the binomial distribution can be used to determine the threshold of randomness for P(+|MoA), we can then use Eq. (1) to translate that threshold to the problem of finding differentially expressed MoAs for the same endpoint.

Case Study:

The treatment of malignant primary brain tumors (e.g. glioblastoma) still remains palliative and encompasses surgery, radiotherapy and chemotherapy. Despite the extensive research, novel pharmacological agents for the treatment of nervous system tumors are urgently needed [3]. Phenotyping models have empowered drug discovery in oncology, and the tumoral growth inhibition induced by novel chemotherapy agents in cancer cell lines has been used as an index of pharmacological potency (IC_{50}) against tumors [4]. SNB-78 is a representative human tumor cell line of gliobastoma multiforme with enhanced expression of proteins involved in its invasive, tumorigenic and metastatic potential [5].

A binary data set of drugs with antitumoral activity in SNB-78 cell line (actives and inactives) has been analyzed and the MoAs associated with this endpoint have been predicted. The resulting MoAs have been compared with those described in the literature as relevant to SNB-78. Drugs were labelled as active when their IC_{50} tumor growth inhibitory value was lesser than 1.1 μM and inactive if greater than 10 μM .

The input data set containts 555 actives and the same number of inactives collected from the National Cancer Institute Database (Developmental Therapeutics Program, NCI/NIH) [6], we have used the binomial distribution and statistical hypothesis testing to distinguish MoAs with a high P(MoA|+)/P(MoA|-) ratio, indicating mechanisms of action associated with antitumoral activity. See Figure 3 and 4.

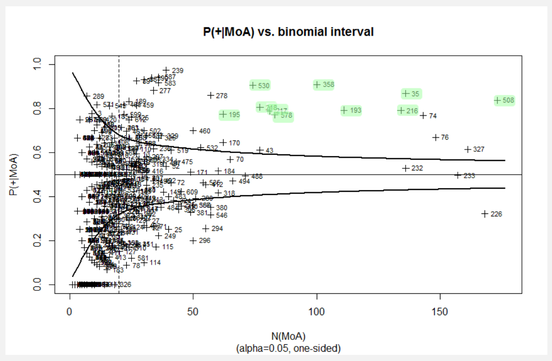


FIGURE 3. Finding MoAs that are good predictors of antitumoral activity in the SNB-78 cell line [green].

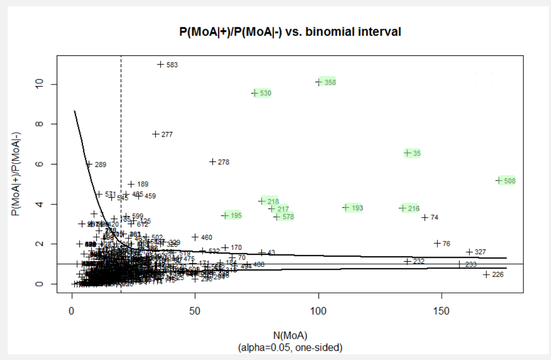


FIGURE 4. Finding MoAs that are differentially associated with antitumoral activity in SNB-78 cell line (compared to non-antitumorals).

MoA ID	MoA	P(+|MoA)	P(MoA|+)/P(MoA|-)	Compounds	References [PubMed]
358	KIT inhibitors	0.91	**10.11**	100	15671569, 12800187
530	Pregnane X receptor agonists	0.90	**9.57**	74	18839173, 18524938
35	ABCP inhibitors	0.86	**6.55**	136	24380367, 22401348
508	PGP1 inhibitors	0.83	**5.17**	173	11206006, 17387584, 24380367
218	DNA topoisomerase 2 alpha inhibitors	0.80	**4.13**	77	11263502, 18402387
193	CYP2D6 inhibitors	0.79	**3.82**	111	17504219, 7671227
216	DNA topoisomerase I inhibitors	0.79	**3.78**	134	18784279, 24021349, 11592782
217	DNA topoisomerase II inhibitors	0.79	**3.76**	81	11592782, 24690174
195	CYP3A4 inhibitors	0.77	**3.42**	62	18839173, 17504219
578	Telomerase inhibitors	0.77	**3.36**	83	21784756, 23329387

TABLE 1. Relevant MoAs predicted as associated with in vitro tumoral growth inhibition of SNB-78 cell line.

Key:
KIT: Mast/Stem cell growth factor receptor (SCFR), proto-oncogene c-kit , tyrosine protein kinase kit or CD117.
ABCP: ATP binding cassette subfamily G member, ABCG2.
PGP1: P-Glycoprotein 1, MDR1, ABCB1.
CYP: Cytochrome P450.

V. Discussion

A list of relevant MoAs predicted for in vitro tumoral growth inhibition in SNB-78 cells are shown in Table 1 with their corresponding P(+|MoA) and P(MoA|+)/P(MoA|-) values. A high value in the P(MoA|+)/P(MoA|-) ratio suggest that certain MoA (target/action) may be involved in the pharmacological activity in this cancer phenotype.

The predicted probability of in vitro antitumoral activity for compounds that are KIT inhibitors, P(+|MoA), is 91%. Additionally, a P(MoA|+)/P(MoA|-) ratio of 10.11 means that drugs with antitumoral activity in the phenotyping assay shows 10.11 times higher probability of being KIT inhibitors than non-active molecules. This results are in agreement with the published chemotherapeutic efficacy of KIT inhibitors in the treatment of neuroblastoma.

The mechanisms of action predicted as associated with chemotherapeutic activity in SNB-78 are, in general, in accordance with published literature (see PubMed references in the table above). DNA topoisomerase has been described as pharmacological targets for neuroblastoma treatment. DNA topoisomerase I inhibitors like irinotecan [PMID:18784279] or topotecan [PMID:24021349] alone or in combination with other chemotherapeutic agents have been used in the treatment of neuroblastoma. Telomerase inhibition has been used as a therapeutic target for the treatment of gliomas and anaplastic astrocytomas.

Because efflux glycoproteins (PGP1 and ABCP) and cytochrome P450 isozymes (2D6 and 3A4 isozymes) participate in the resistance and the metabolic clearance of chemotherapeutic agents, respectively. Thus, it is not surprising their presence as relevant MoAs associated with the antitumoral activity. In addition, Pregnane X receptor (PXR), a transcription factor involved with the induction of CYP3A4, has been related to clinically relevant drug-drug interactions. Chemotherapeutic agents like erlotinib, paclitaxel, tamoxifen or ifosfamide are able to activate PXR leading induction of CYP3A4 and an increase of their metabolic clearance [PMID:18839173].

VI. Conclusions & Outlook

- A statistically sound method has been developed to predict the differentially expressed MoAs between positives and negatives of a binary endpoint, and a threshold has been computed to distinguish relevant MoAs from chance relationships. This method is applicable to find the MoAs associated with any binary endpoint, starting from a balanced or imbalanced data set.
- The case study presented here provides insight for alternate pathways and potential targets involved in antitumoral activity of chemotherapeutic agents in tumor cell lines. The predicted MoAs have been proved as relevant in a literature analysis. Selected predicted MoAs will be experimentally validated in Prous Institute's lab facilities.
- This specific application using SYMMETRY's MoA model has been applied to predict the MoAs associated with other binary endpoints of clinical interest [7].

References

[1] Valencia, A., Prous, J., Mora, O., Nakissa, S., Valerio, L.G. (2013) *Toxicol. Appl. Pharmacology,* **273***, 427-434*.
[2] Sommer, R.S. and Nadachowski, A. (2006) *Mammal Review,* **36***, 251-265.*
[3] Preusser, M., de Ribaupierre, S., Wöhrer, A., Erridge, S.C., Hegi, M., Weller, M., Stupp, R. (2011) *Ann. Neurol.*, **70**, *9-21*.
[4] Moffat, J.G., Rudolph, J., Bailey D. (2014) *Nature Rev. Drug Discovery,* Jul. 18. doi: 10.1038/nrd4366.
[5] Kishi, Y., Okudaira, S., Tanaka, S., Hama, K., Shida, D., Kitayama, J., Yamori, T., Aoki, J., Fijimaki, T., Arai, H. (2006) *J. Biol. Chem.*, **281**,*17492-17500*.
[6] DTP World Wide Web, http://dtp.nci.nih/gov
[7] Valencia, A., Carbó, M., Prous, J. (2014) 16th International Workshop on QSAR in Environmental and Health Sciences. Milan.