Second Kazan Summer School on Chemoinformatics

Book of Abstracts
and Scientific Program

July 6-9, 2015
Kazan, Russia

P.Sidorov[1,2]
H.Gaspar[1]
G.Marcou[1]
D.Horvath[1]
A.Varnek[1]

MAPPABILITY OF DRUG-LIKE SPACE: TOWARDS A POLYPHARMACOLOGICALLY COMPETENT MAP OF DRUG-RELEVANT COMPOUNDS

[1] Laboratoire de Chémoinformatique, UMR 7140 CNRS - Université de Strasbourg, 1 rue Blaise Pascal, Strasbourg 67000, France;
[2] Laboratory of Chemoinfomatics, Butlerov Institute of Chemistry, Kazan Federal University, Kazan, Russia
pavel.sidorov@unistra.fr

This work attempts to address the question whether a "Universal model" of the Chemical Space exists and propose a representation of it. A universal model is intended as a probability distribution of compounds that could be set-independent. The probabilistic model is build as a Generative Topographic Map (GTM). The claim of "universality" is quantitatively justified, with respect to all the structure-activity information available so far. To this purpose, an evolutionary map growth and selection procedure considered both the choice of meta-parameters (poling molecule sets, descriptor types) and map-specific parameters (size, RBF function controls, etc) as degrees of freedom. It was associated to a fitness function measuring the polypharmacological performance of the map, with respect to a multi(144)-target quantitative affinity prediction challenge. Under the pressure of Darwinian selection, the emerging maps were pushed to find (a) the best descriptor type, out of the proposed substructural molecular fragments descriptors schemes, and (b) the specific non-linear "recipe" of generating a model GTM probability distribution which enhances the information contained in certain descriptor elements, but suppresses descriptor "noise". The fittest manifolds were seen to "grow" in rather low-resolution molecular descriptor spaces: pharmacophore- or force-field-type colored atom pairs and triplets rather than more specific sequence or circular fragment counts included in the pool of competing ISIDA descriptor types.

Obtained maps were perfectly suited to solve classification problems: on the overall, more than 80% of the more than 600 distinct and varied classification problems, chosen such as to cover a maximum of exploitable SAR data, were successfully solved. This justifies, in our view, the claim of "Universality" of the constructed GTMs.

In addition, intuitive 2D representations were shown to provide an insightful analysis of drug-like space, and provide huge perspectives for target- and therapeutic range-related compound collections. Due to quantitative validation, the user may gain confidence in the rendered visual patterns, and draw very meaningful conclusions on their behalf.

P.Sidorov thanks the Program of Competitive Growth of Kazan Federal University for the support

7e Journées de la Société Française de Chémoinformatique

8-9 octobre 2015

Université de Nice-Sophia Antipolis
France

Mappability of Drug-like Space: towards a polypharmacologically competent map of drug-relevant compounds

Pavel Sidorov [1,2], Héléna Gaspar [1], Gilles Marcou [1], Dragos Horvath [1], Alexandre Varnek [1]

[1] Laboratoire de Chémoinformatique, UMR 7140 CNRS - Université de Strasbourg, 1 rue Blaise Pascal, Strasbourg 67000, France
[2] Laboratory of Chemoinfomatics, Butlerov Institute of Chemistry, Kazan Federal University, Kazan, Russia

pavel.sidorov@unistra.fr

Intuitive, visual rendering – mapping – of high-dimensional Chemical Spaces (CS), is an important topic in chemoinformatics. This work addresses the question whether a general, compound set-independent map can be generated, and the claim of "universality" quantitatively justified, with respect to all the structure-activity information available so far.
A universal model is intended as a probability distribution of compounds that could be set-independent. The probabilistic model is build as a Generative Topographic Map (GTM). The claim of "universality" is quantitatively justified, with respect to all the structure-activity information available so far. To this purpose, an evolutionary map growth and selection procedure considered both the choice of meta-parameters (poling molecule sets, descriptor types) and map-specific parameters (size, RBF function controls, etc) as degrees of freedom. It was associated to a fitness function measuring the polypharmacological performance of the map, with respect to a multi(144)-target quantitative affinity prediction challenge. Under the pressure of Darwinian selection, the emerging maps were pushed to find (a) the best descriptor type, out of the proposed substructural molecular fragments descriptors schemes, and (b) the specific non-linear "recipe" of generating a model GTM probability distribution which enhances the information contained in certain descriptor elements, but suppresses descriptor "noise". The fittest manifolds were seen to "grow" in rather low-resolution molecular descriptor spaces: pharmacophore- or force-field-type colored atom pairs and triplets rather than more specific sequence or circular fragment counts included in the pool of competing ISIDA descriptor types.
Obtained maps were perfectly suited to solve classification problems: on the overall, more than 80% of the more than 600 distinct and varied classification problems, chosen such as to cover a maximum of exploitable SAR data, were successfully solved. This justifies, in our view, the claim of "Universality" of the constructed GTMs.
In addition, intuitive 2D representations were shown to provide an insightful analysis of drug-like space, and provide huge perspectives for target- and therapeutic range-related compound collections. Due to quantitative validation, the user may gain confidence in the rendered visual patterns, and draw very meaningful conclusions on their behalf.

J Comput Aided Mol Des (2015) 29:1087–1108
DOI 10.1007/s10822-015-9882-z

Mappability of drug-like space: towards a polypharmacologically competent map of drug-relevant compounds

Pavel Sidorov[1,2] · Helena Gaspar[1] · Gilles Marcou[1] · Alexandre Varnek[1,2] · Dragos Horvath[1]

Received: 5 October 2015 / Accepted: 6 November 2015 / Published online: 12 November 2015

Abstract Intuitive, visual rendering—mapping—of high-dimensional chemical spaces (CS), is an important topic in chemoinformatics. Such maps were so far dedicated to specific compound collections—either limited series of known activities, or large, even exhaustive enumerations of molecules, but without associated property data. Typically, they were challenged to answer some classification problem with respect to those same molecules, admired for their aesthetical virtues and then forgotten—because they were set-specific constructs. This work wishes to address the question whether a general, compound set-independent map can be generated, and the claim of "universality" quantitatively justified, with respect to all the structure–activity information available so far—or, more realistically, an exploitable but significant fraction thereof. The "universal" CS map is expected to project molecules from the initial CS into a lower-dimensional space that is neighborhood behavior-compliant with respect to a large panel of ligand properties. Such map should be able to discriminate actives from inactives, or even support quantitative neighborhood-based, parameter-free property prediction (regression) models, for a wide panel of targets and target families. It should be polypharmacologically competent, without requiring any target-specific parameter fitting. This work describes an evolutionary growth procedure of such maps, based on generative topographic mapping, followed by the validation of their polypharmacological competence. Validation was achieved with respect to a maximum of exploitable structure–activity information, covering all of *Homo sapiens* proteins of the ChEMBL database, antiparasitic and antiviral data, etc. Five evolved maps satisfactorily solved hundreds of activity-based ligand classification challenges for targets, and even in vivo properties independent from training data. They also stood chemogenomics-related challenges, as cumulated responsibility vectors obtained by mapping of target-specific ligand collections were shown to represent validated target descriptors, complying with currently accepted target classification in biology. Therefore, they represent, in our opinion, a robust and well documented answer to the key question "What is a good CS map?"

Electronic supplementary material The online version of this article (doi:10.1007/s10822-015-9882-z) contains supplementary material, which is available to authorized users.

✉ Dragos Horvath
dhorvath@unistra.fr

[1] Laboratoire de Chémoinformatique, UMR 7140, CNRS-Univ. Strasbourg, 1 rue Blaise Pascal, 67000 Strasbourg, France

[2] Laboratory of Chemoinformatics, Butlerov Institute of Chemistry, Kazan Federal University, Kazan, Russia

ChemPubSoc Europe

DOI: 10.1002/chem.201403703

CHEMISTRY
A European Journal
Full Paper

Redox Chemistry

Electrochemical Properties of Substituted 2-Methyl-1,4-Naphthoquinones: Redox Behavior Predictions

Mourad Elhabiri,[a] Pavel Sidorov,[b, c] Elena Cesar-Rodo,[a] Gilles Marcou,[b] Don Antoine Lanfranchi,[a] Elisabeth Davioud-Charvet,*[a] Dragos Horvath,[b] and Alexandre Varnek*[b, c]

Abstract: In the context of the investigation of drug-induced oxidative stress in parasitic cells, electrochemical properties of a focused library of polysubstituted menadione derivatives were studied by cyclic voltammetry. These values were used, together with compatible measurements from literature (quinones and related compounds), to build and evaluate a predictive structure–redox potential model (quantitative structure–property relationship, QSPR). Able to provide an online evaluation (through Web interface) of the oxidant character of quinones, the model is aimed to help chemists targeting their synthetic efforts towards analogues of desired redox properties

Introduction

The quinone structure, which is common to numerous natural products with important biological activities, is known for its ability to accept one and/or two electrons in redox processes.[1] The electron-acceptor properties of quinones, causing the formation of radical semiquinone anion or dihydroquinone dianion species responsible for in vivo oxidative stress,[2] can be modulated by the electron-withdrawing or -donating substituents of the electroactive core. The molecular basis of quinone toxicity is the enzyme-catalyzed reduction of the quinone to semiquinone radicals, which then reduce O_2 to superoxide anion radicals and hydrogen peroxide through 1 e- or 2 e-transfer reactions thereby regenerating the quinone. This futile redox cycling and concomitant oxygen activation leads to increased levels of reactive oxygen species (ROS) and glutathione disulfide.[3,4] A well-known example is menadione (2-methyl-1,4-naphthoquinone or vitamin K3), which is a redox-cycler or a "subversive substrate" for numerous flavoproteins acting through a one-electron reduction mechanism, for example, the nicotinamide adenine dinucleotide phosphate (NADPH)-dependent glutathione reductase,[5,6] the NAD(P)H dehydrogenase, lipoamide dehydrogenase (LipDH),[7,8] the trypanothione reductase,[7–9] the thioredoxin reductase[5,10] or the thioredoxin-glutathione reductase.[11] Anecdotally, lipoamide dehydrogenase was named menadione reductase in earlier times.

Menadione and its 5-hydroxylated analogue (plumbagin), are important examples of the broad family of 1,4-naphthoquinones (1,4-NQs), largely distributed in nature (Figure 1). Menadione is the parent core of vitamins K1 and K2. Vitamin K1 (phylloquinone, phytomenadione, or phytonadione) is only

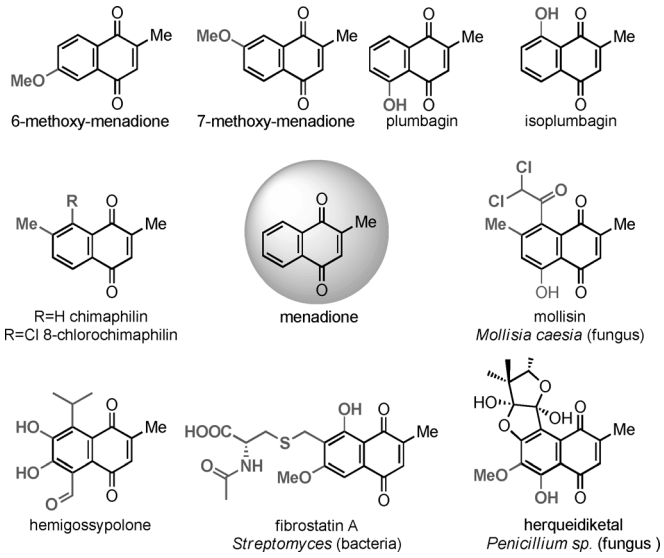

Figure 1. Natural menadione derivatives polysubstituted at the aromatic ring including menadione (2-methyl-1,4-naphthoquinone) and plumbagin (5-hydroxy-menadione).

[a] *Dr. M. Elhabiri, E. Cesar-Rodo, Dr. D. A. Lanfranchi, Dr. E. Davioud-Charvet*
Laboratoire de Chimie Bioorganique et Medicinale
UMR7509 CNRS-Université de Strasbourg
European School of Chemistry, Polymers and Materials (ECPM)
25 Rue Becquerel, F-67087 Strasbourg (France)
Fax: (+33) 3-68-85-27-42
E-mail: elisabeth.davioud@unistra.fr

[b] *P. Sidorov, Dr. G. Marcou, Dr. D. Horvath, Prof. A. Varnek*
Laboratoire de Chemoinformatique
UMR 7140 CNRS-Université de Strasbourg
1 rue Blaise Pascal, Strasbourg 67000 (France)
E-mail: varnek@unistra.fr

[c] *P. Sidorov, Prof. A. Varnek*
Butlerov Institute of Chemistry
Kazan Federal University, Kazan (Russia)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201403703.