Steinmetz, FP, Mellor, CL, Meinl, T and Cronin, MTD

Screening Chemicals for Receptor-Mediated Toxicological and Pharmacological Endpoints: Using Public Data to Build Screening Tools within a KNIME Workflow

Molecular Informatics

Screening chemicals for receptor-mediated toxicological and pharmacological endpoints: Using public data to build screening tools within a KNIME Workflow

F.P. Steinmetz[1], C.L. Mellor[1], T. Meinl[2], M.T.D. Cronin[1]*

[1] School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, England

[2] KNIME.com AG, Technoparkstr. 1, 8005 Zurich, Switzerland

* Corresponding author: Tel. +44 151 231 2402; e-mail address: M.T.Cronin@ljmu.ac.uk (Mark Cronin)

Abstract

Assessing compounds for their pharmacological and toxicological properties is of great importance for industry and regulatory agencies. In this study an approach using open source software and open access databases to build screening tools for receptor-mediated effects is presented. The retinoic acid receptor (RAR), as a pharmacologically and toxicologically relevant target, was chosen for study. RAR agonists are used in the treatment of a number of dermal conditions and specific types of cancer, such as acute promyelocytic leukemia. However, when administered chronically, there is strong evidence that RAR agonists cause hepatosteatosis and liver injury. After compiling information on ligand-protein-interactions, common substructures and physico-chemical properties of ligands were identified manually and coded into SMARTS strings. Based on these SMARTS strings and calculated physico-chemical features, a rule-based screening workflow was built within the KNIME platform. The workflow was evaluated on two datasets: one with RAR agonists exclusively and another large, chemically diverse dataset containing only a few RAR agonists. Possible modifications and applications of screening workflows, dependent on their purpose, are presented.

1. Introduction

Predicting and understanding the properties of new chemical entities is not trivial, whether in the development of novel pharmaceuticals or in assessing potential toxicity. However, *in silico*, Quantitative Structure-Activity Relationship ((Q)SAR) and read-across approaches provide a means of rapidly obtaining information (Blackburn and Stuard, 2014; Patlewicz et al., 2013; Cronin et al., 2013). These can be supported by, or developed from, mechanistic understanding (Zhu et al., 2014). Additionally the concept of the Adverse Outcome Pathway (AOP), i.e. describing a sequence of causally linked events at different biological levels, is increasingly being applied to investigate adverse effects (Vinken et al., 2013). Models may be developed from knowledge of the first key event of an AOP, the molecular initiating event (MIE). In AOP terminology the MIE is followed by cellular and organ responses, which may ultimately result in an adverse effect to an organ, organism or population (Ankley et al., 2010). The MIE represents the initial interaction between molecule and the target. Examples of MIEs include covalent binding to DNA and, of relevance for this study, receptor binding (Gutsell and Russell, 2013; Allen et al., 2014). In pharmacology the mode of action, similar to an AOP, incorporates a MIE which describes how a compound interacts with specific proteins, e.g. receptors, carriers and enzymes. Instead of an adverse effect, the aim in pharmacology is, to achieve a beneficial effect, such as the prevention or treatment of a disease (OECD, 2012; FDA, 2013).

Toxicity may also be brought about by interactions with specific proteins, such as receptors. Endocrine disruptors, for example, are a class of toxicants known to cause their effects by receptor-mediated mechanisms. As such, models for endocrine disruption are usually built around knowledge of receptor interactions, e.g. binding to the oestrogen receptor. For instance, one approach to modelling these effects has been proposed recently by Kolšek et al. (2014) who developed a tool to identify nuclear receptor ligands based on AutoDock Vina; a freeware to investigate ligand-protein-interactions (Molecular Graphics Laboratory, 2014). Limitations of this type of approach are associated with several of the typical issues of docking. First, nuclear receptors, particularly the

non-steroid receptors, are considered to be flexible (Nettles et al., 2007). An inflexible docking model, such as AutoDock Vina, is unlikely to cope with the diversity of ligands (cf. full, partial, inverse agonists and antagonists). The second limitation, when docking is applied on its own, is that kinetics are systemically ignored, which might be vital for *in vivo* biological activity, such as target-organ-toxicity (Campbell, 1983; Davis and Riley, 2004).

The current study focuses on the retinoic acid receptor (RAR), a target relevant for pharmacology and toxicology in equal measure. The RAR is a nuclear receptor which can be divided into three subtypes, RAR-α, RAR-β and RAR-γ. Bound together with the retinoid X receptor (RXR) as a heterodimer, RAR regulates genetic expression. All three subtypes of the RAR are activated by *all-trans* retinoic acid and *9-cis* retinoic acid, which are derivatives of vitamin A (Liu et al., 2014). Ligands are used in the treatment of dermal diseases, e.g. *Acne vulgaris*, *Psoriasis vulgaris*, *Keratosis pilaris* and specific types of cancer, such as acute promyelocytic leukemia (Alizadeh et al., 2014; Leyden et al., 2005; Allen and Bloxham, 1989; Dicken, 1984). The toxicological effects of RAR agonists include changes in lipid metabolism, which may cause hepatosteatosis and leading to liver inflammation, fibrosis and eventually liver failure. Teratogenic effects and neural disorders, such as nausea and headache, have been also reported from retinoids (Moya et al. 2010; Adams, 1993; Biesalski, 1989; Shalita, 1988). There is, therefore, a great need to develop tools to identify these compounds which show these effects.

There are many open source software applications and open access databases supporting modern life sciences and informatics. A number of these open access/source technologies can be utilised to develop tools and approaches for predictive and/or computational toxicology. Some technologies relevant to this study are described below.

The KoNstanz Information MinEr (KNIME) technology is freely available software to analyse and mine data, as well as to build and evaluate predictive models. The software is based on a graphical user interface utilising so called

"nodes" as key units to alter and process data in a "workflow". The basic KNIME workflow technology, as well as many nodes and add-ons for chemo-informatics, is available from www.knime.org. Many types of data can be handled, including chemical formats, such as the Simplified Molecular Input Line Entry System (SMILES) and SMiles ARbitrary Target Specification (SMARTS) strings (Daylight, 2014). KNIME has a strong community of developers building additional nodes for chemo-informatics applications (amongst others), to edit data, calculate physico-chemical properties, analyse structural features etc. It has been shown to be useful in developing workflows for screening tools (Saubern et al., 2011) in the context of predictive toxicology (KNIME, 2013). Furthermore, many other programming languages, such as R, Python or Perl, can be used within a KNIME workflow (KNIME, 2014; Richarz et al., 2013; Berthold et al., 2007).

With regard to biological activity, there are an increasing number of resources available. For instance, ChEMBL is a database of bioactive molecules comprising over 1.5 million compounds and over 9,000 biological targets. Activity values are reported for a variety of endpoints including K_i, K_d, AC_{50}, IC_{50}, and EC_{50}. The database is curated manually and maintained by the European Molecular Biology Laboratory (ChEMBL, 2014). A good example of the application of ChEMBL and the utilisation of its resources was published by Czodrowski (2013). In that study, a detailed analysis of ChEMBL hERG assay data was used to build classification models relevant for drug development and demonstrated the applicability of these data for modelling and value that may result from data mining (Czodrowski, 2013).

Another valuable resource is the Protein Data Bank (PDB) which contains over 100,000 crystallographic structures of proteins such as receptors, transporters and enzymes. A quarter of these protein structures are of human origin, the other structures are from other mammals (mainly rodents) and bacteria. For some proteins, such as the RAR, there are data for several subtypes, species and ligands (PDB, 2014; Berman et al., 1999). Besides the linked publications for every entry, ligand-protein-interactions can be investigated with specific software, for example the freely available PyMOL (PyMOL, 2014). Visualisation

of protein structures of targets, such as receptors, transporters and enzymes, and their corresponding ligands helps to understand ligand-protein-interactions, e.g. hydrogen bonds between ligand and ligand-binding-domain of the protein.

Whilst there is a growing number of computational resources, some of which have been developed for computational toxicology, up until now there has been little effort, and few publications demonstrating the utility of these disparate sources of information and techniques. The aim of this investigation, therefore, was to present a hands-on approach to develop screening tools applicable for many pharmacological and toxicological challenges. The methods applied are based firstly on gathering publically available data on RAR ligands (from ChEMBL and PDB) and secondly extracting information on physico-chemical space and structural features that are relevant to activity. Thirdly, this information was used to build a rule-based screening tool within KNIME. The purpose of the screening tool in this study was to identify potential RAR ligands. RAR is only one example target, i.e. this approach was designed to provide a framework that can, in principle, be used to create screening tools for other receptors should sufficient data be available.

2. Methods

The RAR and its ligands were investigated solely using freeware (N.B. PyMOL is free for academic users only) and open access databases.

2.1 Analysis of RAR ligands using the PDB

The PDB 3.3 (PDB, 2014) was searched for human RAR structures, i.e. RAR-α, RAR-β and RAR-γ. The structures obtained were investigated visually with regard to their ligand-protein-interaction within PyMOL 1.3. Common structural features of the ligands, particularly when apparently responsible for similar ligand-protein-interactions, were extracted. The extracted structural features combined information about molecular distances and electric forces, which may be responsible for hydrogen bonding or the occupation of lipophilic pockets. Subsequently the structural features were coded manually into SMARTS strings. These SMARTS strings were later used in the rule-based workflow to predict potential RAR ligands.

2.2 Extracting data from ChEMBL

The ChEMBL_19 (ChEMBL, 2014) database was searched for the target "RAR". Human data from compounds with K_i (binding affinity), K_d (dissociation constant), AC_{50} (50% activity in molar units) and EC_{50} (50% effect concentration in molar units) values towards RAR-α, RAR-β and RAR-γ were downloaded, combined and sorted by the pChEMBL value. The pChEMBL value is an approach to standardise different types of activity values (Bento et al., 2013). Every compound with a value of not less than five was regarded as being active. This is consistent with the activity interpretations of the ChEMBL database.

2.3 Physico-chemical property calculation

The physico-chemical properties of RAR ligands were calculated using the CDK node for molecular properties within KNIME 2.9.4 (incl. community

contributions). Ranges (i.e. minimum and maximum values) for different types of calculated descriptors for the active ligands were studied including: vertex adjacency information magnitude (VAIM) for structural complexity, number of rotational bonds (RB) for flexibility, molecular weight (MW) for molecular size and the logarithm of the water-octanol partition coefficient (XLogP) for lipophilicity. These four descriptors and their calculated property ranges were utilised to give an insight into the physico-chemical applicability domain (i.e. space) of active RAR ligands.

2.4 Building rules for the screening workflow

The analysis of the PDB has provided structural features coded as SMARTS strings; whilst the analysis of the ChEMBL dataset provided physico-chemical property ranges. Both describe the necessary features for compounds to be active RAR ligands. These features can be interpreted as rules, where compliance and violation will distinguish between RAR ligands and non-ligands respectively. These rules, characterising the physico-chemical space (CDK node for molecular properties) and structural features (Indigo substructure matcher), were written into a KNIME workflow. When executed, this KNIME "screening workflow" will identify potential RAR ligands.

2.5 Testing the screening workflow

The RAR ligands, identified from the ChEMBL dataset, were used to test if all active compounds were identified by the "screening workflow". Since no external validation dataset was available, the dataset of hepatotoxicants provided by Fourches et al. (2010) was screened. The Fourches dataset is a large, chemically diverse dataset (951 compounds), which contains hepatotoxic and non-hepatotoxic drug molecules, including several RAR ligands (Fourches et al., 2010). As the number of RAR ligands is unknown, the performance statistics (sensitivity, specificity etc.) of the screening workflow cannot be calculated, and the predictions for the Fourches dataset are for illustration only. This approach

cannot be considered a full validation as the Fourches data could include liver damage by a number of mechanism not restricted to RAR binding.

3. Results

This study utilised a number of data sources, such as the PDB for ligand-protein-interactions and the ChEMBL database for chemical structures of active compounds against RAR.

3.1 Ligand-protein-interaction in RAR

20 human RAR protein structures bound to different ligands were retrieved from the PDB. These were 4JYG, 4JYH, 4JYI, 4DQM, 4DM6, 4DM8, 3KMR, 3KMZ, 1XAP, 1FD0, 1FCX, 1FCY, 1FCZ, 1DSZ, 1EXA, 1EXX, 3LBD, 4LBD, 2LBD and 1HRA (PDB, 2014). Independent of receptor subtype and ligand, as proposed by Klaholz et al. (2000) the hydrogen bond between an oxygen (most often from a carboxylic group) and the arginine R278 was found to be of great importance for the ligand-protein-interaction. Figure 1, for example, indicates the carboxylic acid of retinoic acid binding to amino acid R278.

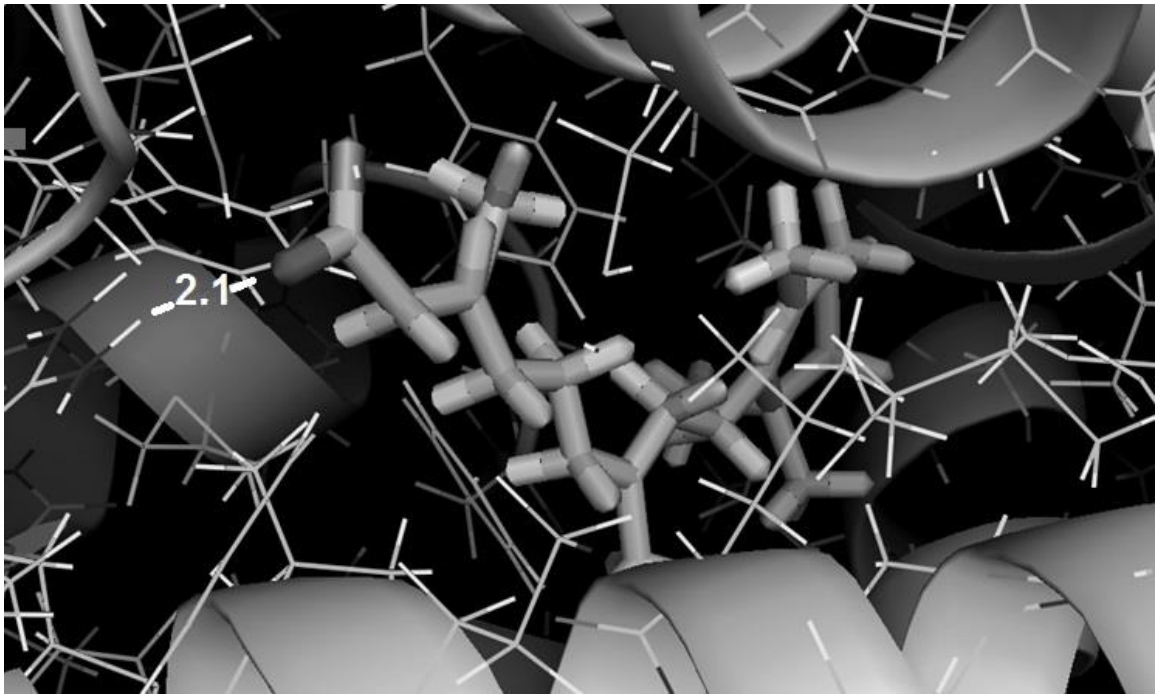

Figure 1: Retinoic acid binding to human RAR gamma (3LBD), highlighting the distance of 2.1 Å between R278 and an oxygen of the carboxylic group of retinoic acid (investigated with PyMOL 1.3)

3.2 Substructures extracted from the ChEMBL database

251 active RAR ligands (pChEMBL ≥ 5) were identified from the ChEMBL database and recorded in the supplementary information. Common structural features to the ligands, as identified from analysis of the chemical properties and visual appearance, were flexibility, a lipophilic scaffold and a terminal hydrogen acceptor (e.g. the carbonyl of a carboxylic group). This information about essential molecular substructures and properties was coded in SMARTS strings, as shown in Table 1. The first rule is for a carboxylic group, an amide or a ring structure derived from these structures, e.g. 1,2,4-oxadiazol-5-one, that has to be at the end of a predominately aliphatic chain. Specific aromatic-containing scaffolds are possible too (cf. Fig. 3), which are still recognised by the substructures from Table 1. Regarding the second rule, the ring structure, e.g. cyclohexene in retinoic acid, can be methylated or halogenated, as the ChEMBL dataset of active RAR ligands revealed.

Table 1: Structural features of ligands converted to rules for the KNIME workflow

Rule	SMARTS string	Structural feature
1. Arginine (R278) binder	*~*~*~*~*~*~*~*~*~*~*~[#6](=O)~[#8]	
	or	
	~~*~*~*~*~*~*~*~*~*~[#6](=O)~[#7]	
and		
2. Methylated or halogenated ring-system	*1~*([F,Cl,Br,I,C])~*~*~*~*~1	

"A" or "*" is a wild card, i.e. it could represent any heavy atom

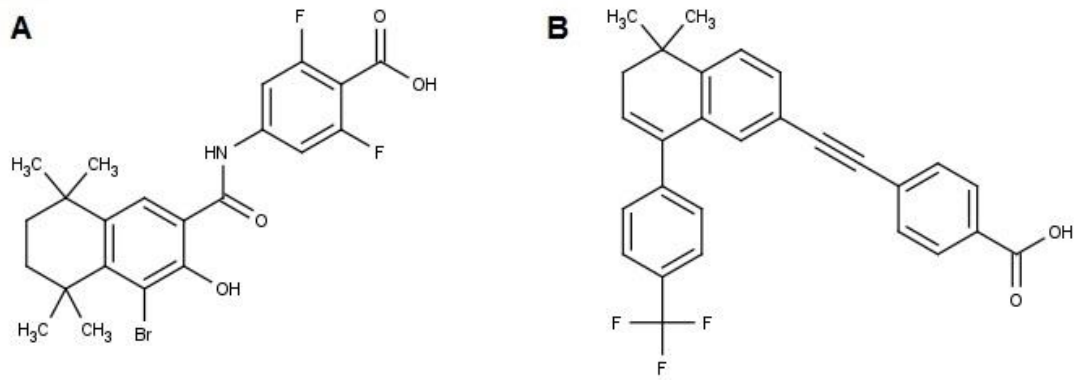

Figure 2: Structures of 4-{[(4-Bromo-3-hydroxy-5,5,8,8-tetramethyl-5,6,7,8-tetrahydro-2-naphthalenyl)carbonyl]amino}-2,6-difluorobenzoic acid (A) and 4-({5,5-Dimethyl-8-[4-(trifluoromethyl)phenyl]-5,6-dihydro-2-naphthalenyl}ethynyl)benzoic acid (B) illustrating the flexible nature, lipophilic character and terminal hydrogen bonding group of two chemically diverse potent RAR ligands

3.3 Physico-chemical properties

The ranges of the physico-chemical properties calculated for the 251 ChEMBL-derived RAR ligands are shown in Table 2. The ranges were converted into rules which can be used as exclusion critera, i.e. if a compound has a MW of greater or equal to 500 Da, then it is, according to the retrieved data, unlikely to be a RAR ligand. The rules have some structural basis, i.e. VAIM and MW express the size and the complexity of the molecule respectively, and the XLogP describes the overall molecular lipophilicity. Beside this basic information the RB indicates the required flexibility of the (lipophilic) chain. Generally speaking, the chemical space covers small, lipophilic molecules with certain degrees of flexibility within the lipophilic scaffold. This is constraint with our understanding of the properties of the ligands and their impact on receptor binding. When dealing with continuous data, margins of error have been applied to the rules, e.g. lower limit for XLogP being 2.00 instead of 2.03 (cf. Table 2). Whilst these are arbitrary, they provide a usable buffer.

Table 2: Physico-chemical property ranges of the RAR ligands and derived rules

Descriptor	Min	Max		Rule
RB:	4	23	➔	≥ 4
VAIM:	5.46	6.40	➔	5 to 6.5
MW:	278.13	488.25	➔	< 500
XLogP:	2.03	10.18	➔	≥ 2.00

3.4 Building the KNIME workflow

A KNIME workflow, which can be downloaded from the supplementary information, was created combining structural features based on the information from PDB and physico-chemical rules based on the ChEMBL dataset. The workflow is shown diagrammatically in Figure 3. The workflow takes the compound of interest through molecular input, implementation of physico-chemical and structural rules in turn, resulting in an output of whether the compound is in or out of "binding space". In more detail, the chemical structure of interest is imported as a SMILES string. Subsequently physico-chemical properties are calculated and the exclusion criteria (cf. Tab. 2) are applied. Following this, the structural rules from Table 1 are applied. In this part of the workflow, the input SMILES strings, which have already passed the physico-chemical rules, are run against a set of SMARTS strings, looking for matches regarding rule 1, the arginine binder, and rule 2, the methylated/halogenated ring-system (cf. Table 1). If a compound's calculated physico-chemical properties is within the defined ranges (cf. Table 2), i.e. it lies within the applicability domain, and contains the relevant structural features (cf. Table 1), then the compound is classed as having the possibility of being an active RAR ligand. If a compound is outside the calculated physico-chemical ranges of Table 2 or does not contain the structural features (cf. Table 1), it is classified as being inactive towards RAR. Finally the workflow, as it is built in Figure 3, exports a csv-file gathering the potential RAR ligands.

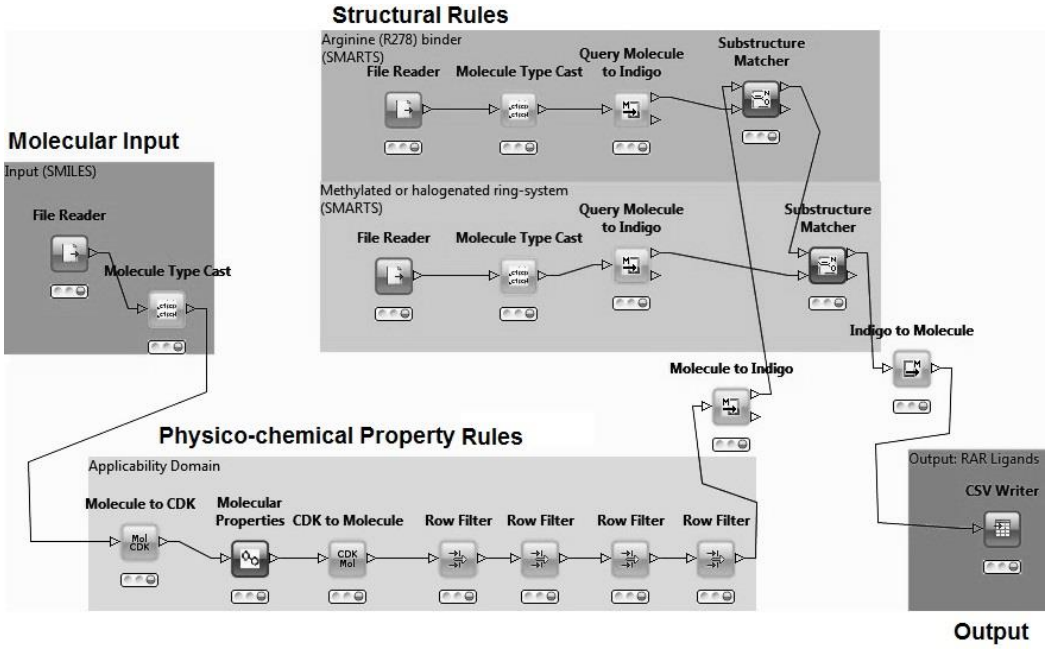

Figure 3: KNIME workflow to screen for RAR ligands indicating the different

3.5 Evaluating the workflow: Screening two datasets

The workflow was used to screen the active 251 compounds from the ChEMBL dataset and all compounds were identified as RAR ligands. 109 of 951 compounds in the Fourches dataset (Fourches et al., 2010) were identified as RAR ligands. Beside retinoids and retinoid-similar structures, some steroids and structurally diverse drugs, such as amineptine (tricyclic antidepressant) and cocaine (tropane alkaloid) were identified as potential RAR binders. The Fourches dataset does not contain information on RAR activity, so performance statistics, such as Cooper statistics (Cooper et al., 1979), i.e. false positive ratio, sensitivity etc., are not meaningful in this context.

4. Discussion

Extrapolation of chemistry to pharmacology or toxicology is a non-trivial, often even impossible, task. However, it is recognised that assessing chemicals for their pharmacological and toxicological properties is of great importance for industry and regulatory agencies. The AOP framework is increasingly seen as providing useable information for modelling as it describes the linkage between the (bio)chemistry of the MIE and the potential adverse effect on individuals and populations (Gutsell and Russell, 2013). A key challenge remains in the prediction of chronic toxicity, particularly modes of action relating to organ level toxicity. New technologies have the potential to exploit the wealth of data that will be delivered from modern database approaches such as ChEMBL and increasing reporting of information from molecular biology. To exploit thse data, tools and strategies, such as data mining, knowledge extraction techniques and (chemo-)informatics tools, are required. Particularly in risk assessment, the identification, characterization and application of chemistry from the MIE of an AOP is increasingly commonly used method to "group" or form categories of similar categories (Vinken et al., 2013; Ankley et al., 2010). Grouping is a crucial element of the further use of predictive toxicology approaches, such as read-across or QSAR and is best undertaken from mechanistic standpoint (Blackburn and Stuard, 2014; Patlewicz et al., 2013; Cronin et al., 2013; OECD, 2012). One of the key challenges for grouping compounds is the definition of similarity. The mechanistic framework provided by the AOP paradigm gives a rational basis to developing chemistry based alerts (from the MIE) for grouping and ultimately confirming group membership using data from assays representing key event.

This study has applied innovative methods to obtain structural information relating to an important MIE. This has been achieved by investigating protein-ligand binding data. Thus, screening a toxicity dataset with the RAR ligand workflow may help to identify compounds acting by the same mechanism and therefore belonging in the same group. For such a group of compounds it is more likely to develop mechanistically valid, robust QSARs (OECD, 2014; Patlewicz et al., 2013; Enoch et al.; OECD 2012; 2011). In drug design, there is an interest in

identifying potent RAR agonists to address several types of cancer and skin diseases (Alizadeh et al., 2014; Leyden et al., 2005; Allen and Bloxham, 1989; Dicken, 1984). The interest may lie in advances towards the receptor-specificity (Vaz and de Lera, 2012; Schinke et al., 2010), i.e. significant activity for certain receptor subtypes, or pharmacokinetics (el Mansouri et al., 1995), e.g. targeted drug localisation. Both strategies may lead to RAR agonists with fewer side effects or better risk-benefit ratios.

In this study information from a set of 251 active RAR ligands from ChEMBL and 20 crystal structures of ligand-protein-interactions from the PDB was extracted and investigated to build a screening workflow prediction potential RAR ligands. The set of active RAR ligands is based on K_i, K_d, AC_{50} and EC_{50} values, that means beside agonists, the dataset is also likely to contain antagonists. However, structural and physico-chemical information on antagonists is regarded as beneficial to predict agonists, as both share many chemical features. The disadvantage of this procedure is a higher likelihood to predict false positives, i.e. predicting antagonists as being active. However as a result of the precautionary nature of this approach, potential drug candidates in drug discovery and potential toxicants should be identified the screening workflow.

As proposed by Klaholz et al. (2000), and confirmed by this study, all ligands are small, flexible compounds with lipophilic (mostly aliphatic) scaffolds and a (more or less) terminal polar functional group, for example, an amide or a carboxylic acid, which creates a hydrogen bond with arginine R278 (PDB, 2014; Klaholz et al., 2000). Potent ligands contain at least one ring structure in the aliphatic scaffold. Furthermore, ring structures may be halogenated, as this does not decrease lipophilicty, such as the compounds illustrated in Figure 2, which are highly potent RAR-α binders (Beard et al., 2002; Johnson et al., 1999).

Figure 2 also illustrates the lipophilic (mostly aliphatic) scaffold. As long as flexibility and lipophilicity are not greatly impaired, compounds with aromatic rings and amides within their scaffold are potential ligands. This explains the large number of wild cards within the SMARTS strings (cf. Table 1). These wild

cards, which are expressed with a "*", represent any heavy atom and the wild card bond expressed with a "~" represents any type of bond. On its own the SMARTS strings developed seem not to be very specific, however due to the rule-based combination of SMARTS strings and the applicability domains defined by physico-chemical attributes, the RAR ligands can be identified with a certain degree of specificity. The exact degree of specificity cannot be calculated, but when observing the predictions for the Fourches dataset (Fourches et al., 2010), where 109 potential RAR ligands out of 951 drug-like compounds were predicted, the outcome implies a certain degree of specificity – or better, selectivity. According to the analysis of the Fourches, 85 compounds of the 109 predicted RAR ligands are hepatotoxic. The RAR actives from the ChEMBL dataset were all correctly predicted, what indicates high sensitivity.

A screening workflow, as designed as in this study, is assumed to be more sensitive than specific, according to the terminology of Cooper et al. (1979), but as "conservativeness" is relative. It should be pointed out that KNIME allows for the easy adjustment of workflows – without mastering computer language; parameters, thresholds and alerts can be changed intuitively. Furthermore it shall be pointed out that the purpose of these kind of screening tools is not to replace *in vitro* assays or any other *in silico* investigation. The main application lies in tasks, such as prioritisation, or as a valuable part of an elaborated consensus model (cf. integrated testing strategy) and it can also assist in the rational grouping of compounds assisting in read-across to predict activity and fill data gaps. It is noted that placing this knowledge in the context of the AOP framework allows for the grouping and read-across to be supported with evidence from assay for other key events (Tollefson et al., 2014).

5. Conclusions

A novel approach to build screening tools solely with freeware (at least for academia) and open access databases has been described. The flexible design within KNIME allows for adjustment and combination of workflows individually regarding their purpose and their specific endpoints. Furthermore a prediction tool for RAR ligands, as an example for toxicology and pharmacology in equal measure, is presented, which may help to identify potential new drugs and toxicants.

Acknowledgement

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) COSMOS Project under grant agreement n° 266835 and Cosmetics Europe.

References

Adams, 1993
J. Adams. Structure-activity and dose-response relationships in the neural and behavioral teratogenesis of retinoids. Neurotoxicol Teratol 15 (1993), 193-202

Alizadeh et al., 2014
F. Alizadeh, A. Bolhassani, A. Khavari, S.Z. Bathaie, T. Naji, S.A. Bidgoli. Retinoids and their biological effects against cancer. Int Immunopharmacol 18 (2014), 43-49

Allen and Bloxham, 1989
J.G. Allen, D.P. Bloxham. The pharmacology and pharmacokinetics of the retinoids. Pharmac Ther 40 (1989), 1-27

Allen et al., 2014
T.E.H. Allen, J.M. Goodman, S. Gutsell, P.J. Russell. Defining Molecular Initiating Events in the Adverse Outcome Pathway Framework for Risk Assessment Chem Res Toxicol. Article in Press (Web Publication Date: October 29, 2014)

Ankley et al., 2010
G.T. Ankley, R.S. Bennett, R.J. Erickson, D.J. Hoff, M.W. Hornung, R.D. Johnson, D.R. Mount, J.W. Nichols, C.L. Russom, P.K. Schmieder, J.A. Serrrano, J.E. Tietge, D.L. Villeneuve DL. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29 (2010), 30-41

Beard et al., 2002
R.L. Beard, T.T. Duong, M. Teng, E.S. Klein, A.M. Standevan, R.A.S. Chandraratna. Synthesis and biological activity of retinoic acid receptor-α specific amides. Bioorg Med Chem Lett 12 (2002), 3145–3148

Bento et al., 2013
A.P. Bento, A. Gaulton, A. Hersey, L.J. Bellis, J. Chambers, M. Davies, F.A. Krüger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos, J.P. Overington. The ChEMBL bioactivity database: an update. Nucleic Acids Res 42 (2013), 1-8

Berman et al., 1999
H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. The Protein Data Bank. Nucleic Acids Res 28 (1999), 235-242

Berthold et al., 2007
M. Berthold, N. Cebron, F. Dill, T. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, B. Wiswedel. Knime: The Konstanz Information Miner, studies in classification, data analysis, and knowledge organization (GfKL), Freiburg, Germany, 2007, Springer-Verlag, pp. 1-8

Biesalski, 1989
H.K. Biesalski. Comparative assessment of the toxicology of vitamin A and retinoids in man. Toxicology 57 (1989), 117-161

Blackburn and Stuard (2014)
K. Blackburn, S.B. Stuard. A framework to facilitate consistent characterization of read across uncertainty. Regul Toxicol Pharmacol 68 (2014), 353-62

Campbell, 1983
T.J. Campbell. Importance of physico-chemical properties in determining the kinetics of the effects of Class I antiarrhythmic drugs on maximum rate of depolarization in guinea-pig ventricle. Br J Pharmac 80 (1983), 33-40

ChEMBL, 2014
ChEMBL www.ebi.ac.uk/chembl, accessed 07.07.2014

Cooper et al., 1979
J.A. Cooper, R. Saracci, P. Cole. Describing the validity of carcinogen screening tests. Br J Cancer 39 (1979), 87-89

Cronin et al., 2013
M.T.D. Cronin, J.C. Madden, A.-N. Richarz, M.D. Nelms, F.P. Steinmetz. Applying read-across for quantitative chronic toxicity prediction. Abstract of 49th congress of the European Societies of Toxicology (EUROTOX), Toxicol Lett 221 (2013), p. 52

Czodrowski, 2013
Paul Czodrowski. hERG Me Out. J Chem Inf Model 53 (2013), 2240–2251

Davis and Riley, 2004
A.M. Davis, R.J. Riley. Predictive ADMET studies, the challenges and the opportunities. Curr Opin in Chem Biol 8 (2004), 378–386

Daylight, 2014
Daylight, www.daylight.com, accessed 07.07.2014

Dicken, 1984
C.H. Dicken. Retinoids: A review. J Am Acad Dermatol 11 (1984), 541-552

Enoch et al., 2011
S.J. Enoch, C.M. Ellison, T.W. Schultz, M.T.D. Cronin. A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. Crit Rev Toxicol 41 (2011), 783-802

FDA, 2013
US Food and Drug Administration, code of federal regulations title 21 (April 2013), www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=3.2, accessed 03.08.2013

Fourches et al., 2010
D. Fourches, J.C. Barnes, N.C. Day, P. Bradley, J.Z. Reed, A. Tropsha. Cheminformatics analysis of assertions mined from literature that describe drug-induced liver injury in different species. Chem Res Toxicol 23 (2010), 171-183

Gutsell and Russell, 2013
S. Gutsell, P. Russell. The role of chemistry in developing understanding of adverse outcome pathways and their application in risk assessment. Toxicol Res 2 (2013), 299-307

Johnson et al., 1999
A.T. Johnson, L. Wang, S.J. Gillett, R.A.S. Chandraratna. High affinity retinoic acid receptor antagonists: Analogs of AGN 193109. Bioorg Med Chem Lett 9 (1999), 573-576

Klaholz et al., 2000
B.P. Klaholz, A. Mitschler, D. Moras. Structural basis for isotype selectivity of the human retinoic acid nuclear receptor. J Mol Biol 302 (2000), 155-170

KNIME, 2014
KNIME, www.knime.org, accessed 07.07.2014

KNIME, 2013
KNIME newsletter, September 2013, www.knime.com/files/knimenewsletter_vol3no3_2013.pdf, accessed 08.08.2014

Kolšek et al., 2014
K. Kolšek, J. Mavri, M. Sollner Dolenc, S. Gobec, S. Turk. Endocrine disruptome - an open source prediction tool for assessing endocrine disruption potential through nuclear receptor binding. J. Chem. Inf. Model. 54 (2014), 1254–1267

Leyden et al., 2005
J.J. Leyden, A. Shalita, D. Thiboutot, K. Washenik, G. Webster. Topical retinoids in inflammatory acne: A retrospective, investigator-blinded, vehicle-controlled, photographic assessment. Clin Ther 27 (2005), 216-224

Liu et al., 2014
Z. Liu, Q. Hu, M.G. Rosenfeld. Complexity of the RAR-mediated transcriptional regulatory programs. Subcell Biochem 70 (2014), 203-225

el Mansouri et al., 1995
S. el Mansouri, M. Tod, M. Leclerq, O. Petitjean, G. Perret, M. Porthault. Time- and dose-dependent kinetics of all-trans-retinoic acid in rats after oral or intravenous administration(s). Drug Metab Dispos February 23 (1995), 227-231

Molecular Graphics Laboratory, 2014
Molecular Graphics Laboratory, www.autodock.scripps.edu, accessed 07.07.2014

Moya et al., 2010
M. Moya, M.J. Gómez-Lechón, J.V. Castell, R. Jover. Enhanced steatosis by nuclear receptor ligands: A study in cultured human hepatocytes and hepatoma cells with a characterized nuclear receptor expression profile. Chem Biol Interact 184 (2010), 376-387

Nettles et al., 2007
K.W. Nettles, J.B. Bruning, G. Gil, E.E. O'Neill, J. Nowak, Y. Guo, Y. Kim, E.R. DeSombre, R. Dilis, R.N. Hanson, A. Joachimiak, G.L. Greene. Structure plasticity in the oestrogen receptor ligand-binding domain. EMBO reports 8 (2007), 563-568

OECD, 2012
Organisation for Economic Co-operation and Development, Collection of working definitions, appendix 1 (2012), www.oecd.org/chemicalsafety/testing/49963576.pdf, accessed 03.08.2014

OECD, 2014
Organisation for Economic Co-operation and Development, Guidance on grouping of chemicals, 2nd edition (2014), www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2014)4&doclanguage=en

Patlewicz et al., 2013
G. Patlewicz, N. Ball, E.D. Booth, E. Hulzebos, E. Zvinavash, C. Hennes. Use of category approaches, read-across and (Q)SAR: General considerations. Regul Toxicol Pharmacol 67 (2013), 1-12

PDB, 2014
Protein Data Bank, www.rcsb.org, accessed 07.07.2014

PyMOL, 2014
PyMOL, http://www.pymol.org, accessed 07.07.2014

Richarz et al., 2013
A.-N. Richarz, S. Enoch, M. Hewitt, J. Madden, M. Nelms, K. Przybylak, C. Yang, M. Berthold, T. Meinl, P. Ohl, M. Cronin. *In silico* workflows for toxicity prediction implemented into KNIME. Abstract of the 49th congress of the European Societies of Toxicology (EUROTOX), Toxicol Lett 221 (2013), S. 81

Sarawat et al., 2014
D. Saraswat, S. Nehra, K.K. Chaudhary, C.V.S. Siva Prasad. *In-silico* screening and *in-vitro* validation of vascular endothelial growth factor receptor-2 (VEGFR-2) inhibitors. Bioinformation 10 (2014), 273-280

Saubern et al., 2011
S. Saubern, R. Guha, J. Baell. KNIME Workflow to assess PAINS filters in SMARTS format. Comparison of RDKit and Indigo cheminformatics libraries. Mol Inform 30 (2011), 847-850

Schinke et al., 2010
C. Schinke, S. Goel, T.D. Bhagat, L. Zhou, Y. Mo, R. Gallagher, G.W. Kabalka, L.C. Platanias, A. Verma, B. Das. Design and synthesis of novel derivatives of all-trans retinoic acid demonstrate the combined importance of acid moiety and conjugated double bonds in its binding to PML–RAR-α oncogene in acute promyelocytic leukemia. Leuk Lymphoma 51 (2010), 1108-1114

Shalita, 1988
A.R. Shalita. Lipid and teratogenic effects of retinoids. J Am Acad Dermatol 19 (1988), 197-198

Tollefsen et al., 2014
K.E. Tollefsen, S. Scholz, M.T. Cronin, S.W. Edwards, J. de Knecht, K. Crofton, N. Garcia-Reyero, T. Hartung, A. Worth, G. Patlewicz. Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA), Regul Toxicol Pharmacol (in press; available online 27 September 2014)

Vaz and de Lera, 2012
B. Vaz, Á.R. de Lera. Advances in drug design with RXR modulators. Expert Opin Drug Discov 7 (2012) 1003-1016

Vinken et al., 2013
M. Vinken, B. Landesmann, M. Goumenou, S. Vinken, I. Shah, H. Jaeschke, C. Willett, M. Whelan, V. Rogiers. Development of an adverse outcome pathway from drug-mediated bile salt export pump inhibition to cholestatic liver injury. Toxicol Sci 136 (2013), 97-106

Zhu et al., 2014
H. Zhu, J. Zhang, M.T. Kim, A. Boison, A. Sedykh, K. Moran. Big data in chemical toxicity research: The use of High-Throughput Screening assays to identify potential toxicants. Chem Res Toxicol 27 (2014), 1643-1651