

This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

Using Molecular Initiating Events to Generate 2D Structure Activity Relationships for Toxicity Screening

Journal:	<i>Chemical Research in Toxicology</i>
Manuscript ID	tx-2016-00101y.R2
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Allen, Timothy; University of Cambridge, Department of Chemistry Liggi, Sonia; University of Cambridge, Department of Chemistry Goodman, Jonathan; University of Cambridge, Department of Chemistry Gutsell, Steve; Unilever Safety and Environmental Assurance Centre Russell, Paul; Unilever Safety and Environmental Assurance Centre

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27

Using Molecular Initiating Events to Generate 2D Structure Activity Relationships for Toxicity Screening

28
29
30
31
32
33

Timothy E. H. Allen,[†] Sonia Liggi,[†] Jonathan M. Goodman,^{†} Steve Gutsell,[‡] and Paul J.
Russell[‡]*

34
35
36
37
38

[†]Centre for Molecular Informatics, Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge, CB2 1EW, United Kingdom

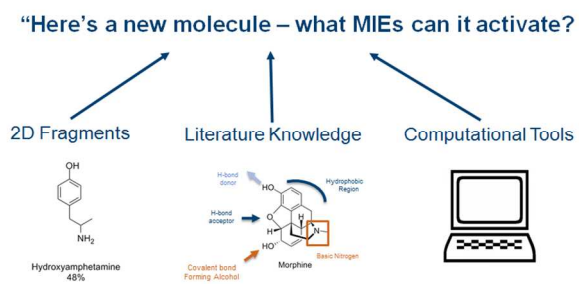
39
40
41
42
43
44

[‡]Unilever Safety and Environmental Assurance Centre, Colworth Science Park, Sharnbrook,
Bedfordshire, MK44 1LQ, United Kingdom

45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

KEYWORDS: Molecular Initiating Event (MIE), Structure Activity Relationship (SAR),
Adverse Outcome Pathway (AOP), Human Toxicology, Risk Assessment.

Table of Contents Graphic



1
2
3 ABSTRACT
4
5
6

7 Molecular initiating events (MIEs) can be boiled down to chemical interactions. Chemicals that
8 interact must have intrinsic properties that allow them this behavior, be these stereochemical,
9 electronic or otherwise. In an attempt to discover some of these chemical characteristics we have
10 constructed structural alert-style structure activity relationships (SARs) to computationally
11 predict MIEs. This work utilizes chemical informatics approaches, searching the ChEMBL
12 database for molecules that bind to a number of pharmacologically important human toxicology
13 targets, including G-protein coupled receptors (GPCRs), enzymes, ion channels, nuclear
14 receptors, and transporters. By screening these compounds to find common 2D fragments, and
15 combining this approach with a good understanding of the literature, bespoke 2D structural alerts
16 have been written. These SARs form the beginning of a tool for screening novel chemicals to
17 establish the kind of interactions they may be able to make in humans. These SARs have been
18 run through an internal validation to test their quality and the results of this are also discussed.
19 MIEs have proven to be difficult to find and characterize but we believe we have taken a key
20 first step with this work.
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 INTRODUCTION
44
45
46

47 The molecular initiating event (MIE) has been defined as a key chemical event in toxicity
48 leading to adverse outcomes through adverse outcome pathways (AOPs).^{1,2} With its basis
49 strongly founded in chemistry, the MIE makes a good starting point for the development of *in*
50 *silico* tools, such as structure activity relationships (SARs). In order for a chemical (or
51 metabolite/breakdown product) to cause an effect *via* an MIE it must conform to specific
52
53
54
55
56
57
58
59
60

1
2
3 chemical characteristics: those which allow it to bind to a receptor, inhibit an enzyme, or modify
4 a protein. An MIE is essentially a chemical interaction. The links between these chemical
5 characteristics and the MIE will undoubtedly be stronger than links to a toxicological endpoint,
6 which is much further down the pathway.
7
8
9
10
11

12
13
14 The AOP framework for risk assessment brings together chemical and biological understanding
15 in an attempt to develop predictive methods for human and environmental toxicology.^{3,4} AOPs
16 span multiple levels of biological organization, and as such a large amount of knowledge is
17 required to make an AOP risk assessment. This will consist of knowledge of the exposure of an
18 organism to a chemical, followed by an understanding of how that chemical is absorbed,
19 distributed, metabolized and excreted and hence how much of the chemical gets to the active
20 site. Once the chemical reaches the active site how, and to what extent, does it bind to, or interact
21 with, a target? How does this interaction lead to a disturbance of the biosystem within this
22 organism and, finally, how does this disturbance leads to effects at a measurable, level of
23 biological organization? This is indeed a large undertaking. However, the AOP provides a
24 framework that, once these pieces are in place, can provide a genuine alternative to *in vivo*
25 testing.⁵⁻⁷
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 MIEs are already being used as gateways to the development of predictive tools and mechanistic
44 understanding. The MIE has been used in quantitative structure activity relationship (QSAR)
45 development⁸ and category formation and read across⁹ within *in silico* tools. Molecular
46 modelling is also finding a use for the MIE, in research leading to an increase in mechanistic
47 understanding of biological-chemical interactions.¹⁰ Once the potential for a compound to
48 activate an MIE has been established, this can be linked to *in vitro* studies of downstream
49 biomarkers associated with known AOPs, allowing risk assessment to be carried out.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 The development, evaluation and current trends in SAR and QSAR research have been well
4 reviewed.^{11,12} Despite advances in these powerful *in silico* techniques they are still treated with
5 some suspicion in the toxicology community. Their role in toxicology risk assessment is
6 generally confined to chemical screening.² Anchoring SARs and QSARs at the MIE removes
7 large amounts of biological complexity from the models, providing stronger links between
8 chemical characteristics and the prediction of adverse outcomes. The construction of clear
9 models with sound theoretical backing is of great importance, and these will illustrate the value
10 of the technique.
11
12

13
14
15 One such type of SAR in toxicity prediction relies on the use of structural fragments from within
16 molecules to distinguish active compounds from inactive ones. One such approach has been used
17 in the development of SAR “rules” for skin sensitization.¹³ Similar approaches are used within
18 computational tools, such as Derek Nexus, a knowledge-based expert system for qualitative
19 toxicity prediction based on chemical structure.¹⁴ The OECD QSAR toolbox
20 (<http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm>) provides a
21 further example of the use of chemical characteristics being linked to potential mechanism or
22 mode of action.¹⁵ The use of structural alerts can be associated with MIEs in predictive
23 toxicology and tools to make these predictions are already under development.^{16,17}
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45 The aim of this work is to build a structural alert-style SAR approach to predict MIEs *in silico*.
46 This study has been conducted with a number of pharmacologically important human targets for
47 systemic toxicity drawn from a paper by Bowes in 2012¹⁸, and several targets of interest from
48 our previous work.¹ In particular the Bowes receptors are highlighted as pharmacologically
49 important off-target interactions, often leading to the failure of new drug candidates in clinical
50 trials.¹⁸ These receptors are also of importance in the development of ingredients for consumer
51
52
53
54
55
56
57
58
59
60

1
2
3 goods products as they could lead to damaging systemic toxicity outcomes. As important drug
4 safety targets these receptors are also likely to have a good amount of biological *in vitro* data
5 available for analysis and the construction of models, providing a further advantage. In addition
6 to this these targets cover a wide variety of different biomolecules, and a good representation of
7 biological space. As such these receptors provide a good starting point for the development of
8 any *in silico* technique. These targets include G-protein coupled receptors (GPCRs), enzymes,
9 ion channels, nuclear receptors and transporters (Table 1).
10
11
12
13
14
15
16
17
18
19

20
21 These biological targets require MIEs to be described in a different way to the well-characterized
22 skin sensitization alerts based on chemical reactivity,¹⁹ because binders are likely to sit in a
23 sterically restricted receptor binding pockets or enzyme active sites and interact with the
24 biomolecule through hydrogen bonds, or charge-charge and hydrophobic interactions.
25
26
27

28
29 Fortunately structural and pharmacophore-style 2D fragments coded in SMILES and SMARTS
30 can still describe these alerts well.
31
32
33
34
35

36 2D fragment based approaches to SARs have several advantages. They are computationally
37 simple, allowing the rapid assessment of a large number of chemicals without the use of a large
38 amount of computational resources. 2D SARs are also mechanistically transparent, and can be
39 easily interpretable as to what it is about a molecule that causes it to activate an MIE. Despite
40 this they do have their limitations. For example, this strategy only locates common substructures
41 in the training data when a relatively large common molecular scaffold is available. Some
42 biological targets are promiscuous and only slight changes in the chemistry of binders results in a
43 large change in biological activity, which is sometimes difficult to describe with a chemical
44 fragment. These fragments also do not analyze the chemical characteristics of whole molecules,
45 and so important pieces of information can be missed if they are just outside of the located
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 fragment. Finally using 2D substructures to model 3D environments is considered an incomplete
4 way to predict biological activity. The shape and size of a molecule may play a bigger role in its
5 ability to fit into, and bind to, a biological target than its chemical features, particularly if the
6 target is a well-defined enzyme active site. This can only be assessed in 3D.
7
8
9

10
11
12
13
14 One target may have more than one MIE. From a chemical perspective it is also true that one
15 binding pocket may allow chemicals to bind to it through different interactions, and these may
16 lead to a diversity of alerts. The alerts in this paper are for a specific “mode of binding” (MOB)
17 and this term will be used in this study to avoid confusion. Without additional biological
18 understanding, it is sometimes not clear if two alerts apply to the same MIE, or to different ones.
19
20
21
22
23
24

25 26 MATERIALS AND METHODS

27
28
29
30 **Data Set.** ChEMBL (<https://www.ebi.ac.uk/chembl/>, version 19, extracted November 2014) was
31 used, containing more than a million annotated compounds, comprising more than twelve million
32 bioactivities covering more than 10 000 targets, all abstracted from the primary scientific
33 literature.²⁰ Compounds with a confidence score of 8 or 9 and with reported activities
34 ($K_i/K_d/IC_{50}/EC_{50}$) better than, or equal to, 10 μM against human protein targets were treated as
35 positives and used for model generation. These cutoffs were chosen to provide chemicals with a
36 pharmacologically-relevant activity at a specific, well-defined, human target. A cut-off of 10 μM
37 will ensure that the compounds have a good degree of biological activity and represents a trade-
38 off between activity and dataset size. A confidence score of 8 represents the assignment of
39 homologous single proteins, and 9 direct single protein interactions.²¹ These compounds are
40 binders irrespective of agonistic and antagonistic activity. Non-binders were omitted from the
41 generation of the models because of the small fraction of compounds found in ChEMBL with
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 reported activities below 10 μ M. Only receptors with at least twenty active compounds were
4
5 used to ensure enough information is available to construct appropriate *in silico* models. The data
6
7 set for each target was split randomly into 75 % training set and a 25 % test set using a function
8
9 in Pipeline Pilot,²² so internal validation could be provided. The data were uniquified to ensure
10
11 no duplicate data corrupted results. This was performed on the molecular structure of each
12
13 chemical based on their atomic connectivity, resulting in different tautomers, enantiomers and
14
15 salts being treated as different data points. The training sets were then combined and uniquified
16
17 to make an amalgamated training set, and the test sets were combined in an analogous fashion to
18
19 make an amalgamated test set. These give a representation of ChEMBL chemical space. The
20
21 amalgamated training set was used to provide fragments present across the whole set. It was not
22
23 used in the development of models. The amalgamated test set was used in the internal validation.
24
25 These compounds had unknown activity across each receptor but were used as anticipated
26
27 negatives to give an idea of chemical space. This is done to overcome bias in the ChEMBL set
28
29 towards positives, and allow model quality to be better assessed than through the use of
30
31 sensitivity alone. In total 51 179 activities across 30 349 unique compounds and 45 human
32
33 targets (Table 1) were extracted for model construction.
34
35
36
37
38
39
40
41

42 **Model Construction.** The compound's canonical SMILES were used to generate 2D fragment-
43
44 based structural alerts using custom scripting in Pipeline Pilot.²² Molecular fragments were
45
46 identified based on matching atom types and charges between molecules. Partial aliphatic rings
47
48 were allowed within fragments. The fragment size and frequency within the training set were
49
50 altered to sample a diversity of potentially active fragments. The outputs were viewed and
51
52 curated manually to identify key fragments which are associated with positive activity, rather
53
54 than just being common among medicinal chemicals. Text based literature searches were
55
56
57
58
59
60

1
2
3 conducted using SciFinder to find information on targets that could aid in the construction of
4
5 models. In some cases, crystal structures were found. In others, well-known molecules with
6
7 precisely understood binding behavior were found and the structures of these compared to the
8
9 found fragments. Based on the elucidated fragments, and understanding gained through the
10
11 literature searches, structural alerts were coded as SMILES strings or SMARTS and models were
12
13 constructed using a substructure filter.
14
15

16
17
18 **Internal Validation.** Test sets were used to calculate sensitivity (SE) for each structural alert.
19

$$SE = \frac{TP}{TP + FN}$$

20
21
22
23
24
25
26 In order to provide a negative test set for each human receptor target the test set for the target of
27
28 interest was subtracted from the amalgamated test set. As such every negative test set is not the
29
30 same. This is to give an idea of chemical space, but the chemicals are untested against the target
31
32 of interest and are treated as negatives only for the purpose of this analysis. This is done to
33
34 provide analysis and get a confidence score for models. As such these results must be treated
35
36 carefully, but should provide more guidance greater than sensitivity-only calculations.
37
38

39
40 Specificity (SP) is calculated from these results.
41
42

$$SP = \frac{TN}{TN + FP}$$

43
44
45
46
47
48 Overall quality (Q) is calculated based on the total number of correct predictions.
49

$$Q = \frac{TP + TN}{TP + TN + FP + FN}$$

To overcome the issue of having a larger negative test set than positive the Matthews correlation coefficient (MCC) was used;²³

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

RESULTS AND DISCUSSION

Amalgamated Set

In the search for common 2D fragments in the ChEMBL dataset it is inevitable that some fragments will be found that are simply common among medicinal chemicals within ChEMBL. In order to take this into account, and to prevent models being built using fragments that are simply common across the entire dataset of all receptors, a bespoke model building process is used. The fragments are subjected to human analysis, and appropriate fragments selected, rather than simply using a computer to select all fragments. In addition to this, a dataset was constructed containing all compounds across all test sets analyzed in this study. This training set was uniquified, and then run through the same protocols used to generate 2D fragments for each receptor. This generated a number of fragments that are simply common among medicinal chemicals of this type, and so should not be considered indicative of any particular receptor-ligand interaction, unless exceptionally high percentages of the training set include them. These are shown in Figure 1.

As may have been expected, some fragments, including benzene rings, aliphatic amines, and short carbon chains are commonly present.

Structural Alerts

1
2
3 Across the 45 human targets analyzed in this work, 126 structural alerts have been developed,
4 and these are shown in Table 2. These structural alerts include those found using maximal
5
6 common substructure searches using Pipeline Pilot, those found in typical ligands for the target,
7
8 those identified in previously accepted pharmacophores, those identified in existing crystal
9
10 structures, and combinations of these. Each alert is labelled in Table 2 to indicate its origin and
11
12 pharmacophores, crystal structures and typical binders found in the literature are referenced. The
13
14 SMILES and SMARTS that were coded for these structural alerts are provided in supporting
15
16 information. Table 2 also provides the fraction of positives identified by each alert in the training
17
18 set.
19
20
21
22
23

24
25 Results for the internal validation of structural alerts are shown in Table 3. A total of 126 alerts
26
27 have been produced across 45 human targets. In large part this has been successful, with 77
28
29 models across 39 targets producing MCC values of greater than 0.2 (representing 60 % correct
30
31 predictions for a balanced data set), 28 models across 21 targets giving MCC values greater than
32
33 0.4 (70 % correct predictions for a balanced data set), and seven models across seven receptors
34
35 scoring MCC values of greater than 0.6 (80 % correct predictions for a balanced data set) in
36
37 internal validation. These results must be treated with a note of caution, however, as the
38
39 remainder of the ChEMBL training set was used as a negative test set, meaning these compounds
40
41 have not been confirmed as true negatives. Despite this, these results show that the structural
42
43 alerts that have been developed do not tend to over-fire, and as such the results obtained by
44
45 running a novel chemical through these structural alerts should be of benefit to a toxicologist
46
47 wanting to be able to perform a pragmatic risk assessment when combined with other data, or
48
49 identify targets of interest for further investigation.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 This study has produced an average of just under three structural alerts (2.93) per target. It is
4
5 notable that a number of targets have only one structural alert covering a large proportion of their
6
7 active compounds (Dihydrofolate reductase, GAR transformylase, tyrosine-protein kinase, the
8
9 glutamate (NMDA) receptor, the HERG channel, and the voltage gated K channel subunit
10
11 Kv7.1). This suggests that these biological targets are quite specific in the compounds that they
12
13 accept and hence there is less structural variability in their binders. Only two targets have more
14
15 than five fragments, the adrenergic 2a receptor (6) and monoamine oxidase (7), indicating that
16
17 these are more promiscuous receptors with more structurally varied binders.
18
19
20
21
22

23 Combined Models

24
25
26 2D structural alerts, such as those developed in this work, can be used in toxicity screening, as a
27
28 hazard identification tool, or as a tool in risk assessment to support a decision on chemical safety.
29
30 In each of these cases the structural alerts will be used in different ways. In hazard identification
31
32 tools need to be calibrated to provide maximum safety; that is to minimize the number of false
33
34 negatives at the expense of the number of false positives predicted. In risk assessment tools will
35
36 be adjusted to provide maximum accuracy; to provide the best possible predictivity as measured
37
38 by a metric such as MCC.
39
40
41
42
43

44 To provide results on the structural alerts developed in this work for hazard identification, the
45
46 structural alerts for each target were combined into a model requiring a molecule to contain any
47
48 of the structural alerts to be predicted a positive. These models were tested against the positive
49
50 test set for each target to provide a sensitivity value in each case. These results are shown in
51
52 Table 4. This provides perspective on the proportion of binders likely to be predicted by these
53
54 models, in a hazard assessment exercise. 45 models were tested and their sensitivity values
55
56
57
58
59
60

1
2
3 calculated. 29 models scored sensitivity values of 50% or more, 13 scored 75% or more and five
4
5 scored 90% or more. These results provide a promising basis for the construction of a toxicity
6
7 screening tool for hazard identification based on these structural alerts predicting MIEs. 2D
8
9 structural alert methods such as this have an advantage in hazard assessment as they are
10
11 computationally quick and inexpensive, and often hazard assessment requires the processing of
12
13 large numbers of chemicals to screen out potentially toxic molecules at an early stage of
14
15 compound development.
16
17
18
19

20
21 Risk assessment for individual chemicals is very different to hazard identification, and as such
22
23 requires a different approach. Small numbers of chemicals are assessed on a case-by-case basis
24
25 to ensure their safety. This requires expert input and pragmatic decision making by an expert
26
27 scientist, and as such a batch test of these structural alerts is not an appropriate way to assess
28
29 their value. 2D structural alerts such as these may find use in this field, and if they do it is
30
31 important that the expert takes into account the results from our study to make an informed
32
33 decision as to whether a chemical is likely to bind to a given target or not. The statistical results
34
35 for sensitivity, specificity, overall quality and MCC values provided in Table 2 give guidance as
36
37 to which of the alerts are most predictive and which are likely to over-fire. Alerts with a high
38
39 sensitivity and MCC are very predictive and should be most useful in risk assessment. Alerts
40
41 with a low specificity are hitting a number of compounds that are not considered binders in
42
43 ChEMBL, and as such will produce a large number of false positives in risk assessment.
44
45
46
47
48

49
50 Further research will be required to expand our work to other key MIEs and improve it to the
51
52 point where it can quantitatively predict the amount of a toxicant required at a target to exhibit a
53
54 toxicological response. In addition to this a number of other tools will be required to accurately
55
56 predict the amount of a toxicant that is able to reach the site of the MIE, and the biological
57
58
59
60

1
2
3 response that will result from the activation of the MIE, in order to represent a complete risk
4
5 assessment.
6
7

8 9 Biological Relevance

10
11
12 The AOP Wiki represents the best current repository for AOP information, and as such a search
13 was performed to identify AOPs which could result from the binding of a molecule to the targets
14 examined in this work.²⁴ The majority of the AOPs that may be associated with this work are
15 currently under development (totaling 19 pathways) and the associated targets are listed below:
16
17

- 18 • Acetylcholinesterase
- 19
- 20 • Androgen Receptor
- 21
- 22 • Cyclooxygenase (5 pathways)
- 23
- 24 • Ether-a-go-go voltage gated potassium channel
- 25
- 26 • Glucocorticoid Receptor (2 pathways)
- 27
- 28 • Glutamate Receptor
- 29
- 30 • Histamine H2 Receptor
- 31
- 32 • Serotonin Transporter (2 pathways)
- 33
- 34 • Sodium Channel (3 pathways)
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

1
2
3 This shows much promise for the future of AOP-based toxicology studies, as the AOP Wiki is
4 currently in its infancy and will continue to grow over time, providing further scope for the
5 identification of potentially toxic pathways using work such as this.
6
7
8
9

10
11 One pathway from the AOP wiki is well developed for agonism of the androgen receptor leading
12 to reproductive dysfunction in adult female fish. While this pathway is not based on human
13 toxicology studies, with time the AOP Wiki will be able to provide such detailed pathways for
14 human toxicity pathways. In addition to this there are a number of MIEs, KEs and AOPs that
15 transcend species and as such this research may be relevant to human toxicology. In essence
16 agonism at the androgen receptor leads to a decrease in the concentrations of the hormones
17 gonadotropin, testosterone, estradiol and vitellogenin. A decrease in vitellogenin uptake leads to
18 impaired development of oocytes, decreased spawning rates and a population decline. A graphic
19 of this AOP is shown in Figure 2.
20
21
22
23
24
25
26
27
28
29
30
31

32
33
34 Further information on clinical pathways that may be affected by the agonism or antagonism of
35 the targets studied in this work is presented in the Bowes 2012 article.¹⁸
36
37
38

39 CONCLUSIONS

40
41
42

43 As toxicology moves away from animal based approaches and towards in silico and in vitro
44 methods, further understanding and new tools are required. The MIE is the first KE in the AOP,
45 the boundary between chemistry and biology where a chemical makes its first key interaction
46 with the body. By understanding the chemical properties of existing receptor-binding chemicals
47 we aim to be able to predict the MIEs of new molecules computationally. The focus of this work
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 was to utilize open source data to develop 2D structural alert-based SAR models for the
4
5 prediction of MIEs associated with pharmacologically important human targets.
6
7

8
9 We have produced a number of 2D SAR models in an attempt to describe the characteristic
10
11 fragments that allow molecules to cause effects *via* MIEs. A total of 126 alerts across 45 human
12
13 targets have been developed in this work. A number of these alerts performed well, with an
14
15 emphasis on being able to combine alerts into models for individual targets which will have high
16
17 levels of specificity and therefore will not over-fire. When the 45 combined models were tested,
18
19 29 models scored sensitivity values of 50% or more, 13 scored 75% or more and five scored 90%
20
21 or more, providing overall a good level of coverage. To prevent the models over-firing the whole
22
23 ChEMBL data set was analyzed to provide fragments that are common in chemicals in this set,
24
25 and as such should not be used as structural alerts. This information was used, along with
26
27 existing pharmacophores, crystal structures and an understanding of typical binders found in the
28
29 literature, to manually curate structural alerts provided by a maximal common substructure
30
31 searches and develop structural alerts. This has resulted in a number of structural alerts with high
32
33 specificity values: 95 of 126 alerts scored 95% or greater specificity values.
34
35
36
37
38
39

40
41 Understanding of the chemical characteristics that govern receptor MIEs will be a key step in the
42
43 development of AOP based tools for toxicity risk assessment. 2D fragment based approaches are
44
45 not the only way to attempt to answer this complex problem, and a number of approaches will
46
47 need to be combined to provide a quantitative risk assessment. In this first step we have begun to
48
49 explore this area, with sights set on the development of *in silico* screening tools. We believe that
50
51 these fragment alerts can provide useful information in compound development, regarding the
52
53 potential toxic effects of lead chemical compounds, and provide a basis for exciting new
54
55 developments in the understanding of receptor MIEs, and how they can be used in toxicology.
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 ASSOCIATED CONTENT
4
5

6
7 **Supporting Information.**
8

9
10 SMARTS/SMILES for the structural alerts developed in this work. This material is available free
11
12 of charge via the Internet at <http://pubs.acs.org>.
13
14

15
16 AUTHOR INFORMATION
17

18
19 **Corresponding Author**
20

21 jmg11@cam.ac.uk
22
23

24
25 **Author Contributions**
26

27 The manuscript was written through contributions of all authors. All authors have given approval
28
29 to the final version of the manuscript.
30
31

32
33 **Funding Sources**
34

35 The authors acknowledge the financial support of Unilever.
36
37

38
39 **Data Statement**
40

41
42 According to the University of Cambridge data management policy, all the data used in this
43
44 paper is available either in the paper or in the SI. A copy of the data is also available in the
45
46 University of Cambridge repository at: <https://www.repository.cam.ac.uk/>
47
48
49

50
51
52
53 ABBREVIATIONS
54
55
56
57
58
59
60

1
2
3 AA2aR, adenosine A2a receptor; AC, acetylcholinesterase; ADR, adrenergic receptor; AOP,
4
5 adverse outcome pathway; AR, androgen receptor; CCKAR, cholecystokinin A receptor; COX,
6
7 cyclooxygenase; CR, cannabinoid receptor; DHFR, dihydrofolate reductase; DR, dopamine
8
9 receptor; DT, dopamine transporter; E2, estradiol; ER ET-A, endothelin receptor ET-A; GART,
10
11 GAR transformylase; GPCR, G-protein coupled receptor; GR, glucocorticoid receptor; GRSZ1,
12
13 glutamate receptor subunit zeta 1; GtH, gonadotropin hormone; HDAC 1, histone deacetylase 1;
14
15 HERG; human ether-a-go-go related gene; HR, histamine receptor; LCK, tyrosin protein kinase;
16
17 MAO A, monoamine oxidase A; MAR, muscarinic acetylcholine receptor; MCC, Matthews
18
19 correlation coefficient; MIE, molecular initiating event; MOB, mode of binding; NMDA, N-
20
21 methyl-D-aspartate; NT, norepinephrine transporter; OR, opioid receptor; PDE,
22
23 phosphodiesterase; Q, overall quality; (Q)SAR, (quantitative)structure activity relationship;
24
25 SCV-A, sodium channel V subunit alpha; SE, sensitivity; SP, specificity; SR, serotonin receptor;
26
27 ST, serotonin transporter; T, testosterone; TS, thymidylate synthase; VGKC Kv7.1, voltage gated
28
29 K channel subunit Kv7.1; VTG, vitellogenin; VV1a, vasopressin V1a.

36 REFERENCES

- 37
38
39 (1) Allen, T. E. H., Goodman, J. M., Gutsell, S., and Russell, P. J. (2014) Defining molecular
40
41 initiating events in the adverse outcome pathway framework for risk assessment. *Chem.*
42
43 *Res. Toxicol.* 27, 2100–2112.
- 44 (2) Gutsell, S., and Russell, P. (2013) The role of chemistry in developing understanding of
45
46 adverse outcome pathways and their application in risk assessment. *Toxicol. Res.* 2, 299–307.
- 47 (3) OECD. Proposal for a Template, and Guidance on Developing and Assessing the
48
49 Completeness of Adverse Outcome Pathways, Appendix I, Collection of Working Definitions.
50
51 <http://www.oecd.org/chemicalsafety/testingofchemicals/49963576.pdf>.
- 52 (4) Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D.,
53
54 Mount, D. R., Nichols, J. W., Russom, C. L., Schmieder, P. K., Serrano, J. A., Tietge, J. E., and
55
56 Villeneuve, D. L. (2010) Adverse outcome pathways: A conceptual framework to support
57
58 ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.* 29, 730–741.
- 59 (5) Committee on Toxicity Testing and Assessment of Environmental Agents and National
60
61 Research Council. (2007) *Toxicology testing in the 21st century: A vision and a strategy*,
National Academies Press, Washington, DC.

- 1
2
3 (6) Hartung, T. (2009) A Toxicology for the 21st Century—Mapping the Road Ahead. *Toxicol.*
4 *Sci.* 109, 18–23.
- 5
6 (7) Burden, N., Sewell, F., Andersen, M. E., Boobis, A., Chipman, J. K., Cronin, M. T. D.,
7 Hutchinson, T. H., Kimber, I., and Whelan, M. (2015) Adverse Outcome Pathways can drive
8 non-animal approaches for safety assessment. *J. Appl. Toxicol.* 35, 971–975.
- 9
10 (8) Yang, C., Arvidson, K., Richard, A., Worth, A., Tarkhov, A., Ringeissen, S., Marusczyk, J.,
11 Gasteiger, J., Rathman, J., and Schwab, C. (2014) Chemotypes and chemotyper: A new
12 structural representation standard to include atomic/bond properties into structural alerts for
13 toxicity effects and mechanisms, [http://www.epa.gov/ncct/download_](http://www.epa.gov/ncct/download_files/sot/posters/2013/SOT2013_poster_Chemotype_final.pdf)
14 [files/sot/posters/2013/SOT2013_poster_Chemotype_final.pdf](http://www.epa.gov/ncct/download_files/sot/posters/2013/SOT2013_poster_Chemotype_final.pdf).
- 15
16 (9) Nelms, M. D., Ates, G., Madden, J. C., Vinken, M., Cronin, M. T. D., Rogiers, V., and
17 Enoch, S. J. (2014) Proposal of an in silico profiler for categorisation of repeat dose toxicity data
18 of hair dyes. *Arch. Toxicol.* 89, 733–741.
- 19
20 (10) Tsakovska, I., Al Sharif, M., Alov, P., Diukendjieva, A., Fioravanzo, E., Cronin, M. T. D.,
21 and Pajeva, I. (2014) Molecular modelling study of the PPAR γ receptor in relation to the mode
22 of action/adverse outcome pathway framework for liver steatosis. *Int. J. Mol. Sci.* 15, 7651–
23 7666.
- 24
25 (11) Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., Dearden,
26 J., Gramatica, P., Martin, Y. C., Todeschini, R., Consonni, V., Kuz'min, V. E., Cramer, R.,
27 Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., and Tropsha, A.
28 (2014) QSAR modeling: Where have you been? Where are you going to? *J. Med. Chem.* 57,
29 4977-5010.
- 30
31 (12) McKinney, J. D., Richard, A., Waller, C., Newman, M. C., and Gerberick, F. (2000) The
32 practice of structure activity relationships (SAR) in toxicology. *Toxicol. Sci.* 56, 8–17.
- 33
34 (13) Payne, M. P., and Walsh, P. T. (1994) Structure-activity relationships for skin sensitization
35 potential: development of structural alerts for use in knowledge-based toxicity prediction
36 systems. *J. Chem. Inf. Comput. Sci.* 34, 154–161.
- 37
38 (14) Marchant, C. A., Briggs, K. A., and Long, A. (2008) In Silico Tools for Sharing Data and
39 Knowledge on Toxicity and Metabolism: Derek for Windows, Meteor, and Vitic. *Toxicol. Mech.*
40 *Methods* 18, 177–187.
- 41
42 (15) The OECD QSAR Toolbox, <http://www.qsartoolbox.org/>.
- 43
44 (16) Madden, J. C., Rogiers, V., and Vinken, M. (2014) Application of in silico and in vitro
45 methods in the development of adverse outcome pathway constructs in wildlife. *Philos. Trans. R.*
46 *Soc. B Biol. Sci.* 369, 1–7.
- 47
48 (17) Tollefsen, K. E., Scholz, S., Cronin, M. T., Edwards, S. W., de Knecht, J., Crofton, K.,
49 Garcia-Reyero, N., Hartung, T., Worth, A., and Patlewicz, G. (2014) Applying Adverse
50 Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment
51 (IATA). *Regul. Toxicol. Pharmacol.* 70, 629–640.
- 52
53 (18) Bowes, J., Brown, A. J., Hamon, J., Jarolimek, W., Sridhar, A., Waldron, G., and
54 Whitebread, S. (2012) Reducing safety-related drug attrition: the use of in vitro pharmacological
55 profiling. *Nat. Rev. Drug Discov.* 11, 909–22.
- 56
57 (19) Schwöbel, J. A. H., Koleva, Y. K., Enoch, S. J., Bajot, F., Hewitt, M., Madden, J. C.,
58
59
60

- 1
2
3 Roberts, D. W., Schultz, T. W., and Cronin, M. T. D. (2011) Measurement and estimation of
4 electrophilic reactivity for predictive toxicology. *Chem. Rev.* *111*, 2562–2596.
- 5
6 (20) Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., Krüger, F. A.,
7 Light, Y., Mak, L., McGlinchey, S., Nowotka, M., Papadatos, G., Santos, R., and Overington, J.
8 P. (2014) The ChEMBL bioactivity database: An update. *Nucleic Acids Res.* *42*, 1083–1090.
- 9
10 (21) European Molecular Biology Laboratory–European Bioinformatics Institute, ChEMBL
11 database, <https://www.ebi.ac.uk/chembl/>.
- 12
13 (22) Accelrys Software Inc. Pipeline Pilot v.8.0.1.500.
- 14
15 (23) Matthews, B. W. (1975) Comparison of the Predicted and Observed Secondary Structure of
16 T4 Phage Lysozyme. *Biochim. Biophys. Acta* *405*, 442–451.
- 17
18 (24) European Commission Institute for Health and Consumer Protection. AOP Wiki,
19 https://aopkb.org/aopwiki/index.php/Main_Page.
- 20
21 (25) Park, H. M., Lee, J. H., Yaoyao, J., Jun, H. J., and Lee, S. J. (2011) Limonene, a natural
22 cyclic terpene, is an agonistic ligand for adenosine A2A receptors. *Biochem. Biophys. Res.*
23 *Commun.* *404*, 345–348.
- 24
25 (26) Russo, C., Arcidiacono, G., and Polosa, R. (2006) Adenosine receptors: Promising targets
26 for the development of novel therapeutics and diagnostics for asthma. *Fundam. Clin. Pharmacol.*
27 *20*, 9–19.
- 28
29 (27) Hodgson, R. A., Bertorelli, R., Varty, G. B., Lachowicz, J. E., Forlani, A., Fredduzzi, S.,
30 Cohen-williams, M. E., Higgins, G. A., Impagnatiello, F., Nicolussi, E., Parra, L. E., Foster, C.,
31 Zhai, Y., Neustadt, B. R., Stamford, A. W., Parker, E. M., Reggiani, A., and Hunter, J. C. (2009)
32 Characterization of the potent and highly selective A2A receptor antagonists preladenant and
33 SCH 412348 [7-[2-[4- 2,4-Difluorophenyl]-1-piperazinyl]ethyl]-2-(2-furanyl)-7H- pyrazolo[4,3-
34 e][1,2,4]triazolo[1,5-c]pyrimidin-5-amine] in rodent models of moveme. *J. Pharmacol. Exp.*
35 *Ther.* *330*, 294–303.
- 36
37 (28) Bylund, D. B. (1985) Heterogeneity of alpha-2 adrenergic receptors. *Pharmacol. Biochem.*
38 *Behav.* *22*, 835–843.
- 39
40 (29) Sanders, J., Miller, D. D., and Patil, P. N. (1975) Alpha adrenergic and histaminergic effects
41 of tolazoline-like imidazolines. *J. Pharmacol. Exp. Ther.* *195*, 362–371.
- 42
43 (30) Knepper, S. M., Buckner, S. A., Brune, M. E., DeBernardis, J. F., Meyer, M. D., and
44 Hancock, A. A. (1995) A-61 603 , a potent alpha1-adrenergic receptor agonist, selective for the
45 alpha1A receptor subtype. *J. Pharmacol. Exp. Ther.* *274*, 97–103.
- 46
47 (31) Li, G., Regunathan, S., Barrow, C. J., Eshraghi, J., Cooper, R., and Reis, D. J. (1994)
48 Agmatine: an endogenous clonidine-displacing substance in the brain. *Science (80-.).* *263*, 966–
49 969.
- 50
51 (32) Lemke, K. A. (2004) Perioperative use of selective alpha-2 agonists and antagonists in small
52 animals. *Can. Vet. J.* *45*, 475–480.
- 53
54 (33) Angeli, P., Volpin, R., Piovan, D., Bortoluzzi, A., Craighero, R., Bottaro, S., Finucci, G. F.,
55 Casiglia, E., Sticca, A., De Toni, R., Pavan, L., and Gatta, A. (1998) Acute effects of the oral
56 administration of midodrine, an alpha-adrenergic agonist, on renal hemodynamics and renal
57 function in cirrhotic patients with ascites. *Hepatology* *28*, 937–943.
- 58
59
60

- 1
2
3 (34) Wellman, P. J., and Davies, B. T. (1992) Reversal of cirazoline- and phenylpropanolamine-
4 induced anorexia by the alpha 1-receptor antagonist prazosin. *Pharmacol. Biochem. Behav.* 42,
5 97–100.
6
7 (35) Lasch, P., and Jakobs, K. H. (1979) Agonistic and antagonistic effects of various alpha-
8 adrenergic agonists in human platelets. *Naunyn. Schmiedebergs. Arch. Pharmacol.* 306, 119–
9 125.
10
11 (36) Kier, L. B. (1969) The preferred conformation of noradrenaline and a consideration of the
12 alpha-adrenergic receptor. *J. Pharm. Pharmacol.* 21, 93–96.
13
14 (37) Virtanen, R., Savola, J. M., Saano, V., and Nyman, L. (1988) Characterization of the
15 selectivity, specificity and potency of medetomidine as an alpha 2-adrenoceptor agonist. *Eur. J.*
16 *Pharmacol.* 150, 9–14.
17
18 (38) Arnsten, A. F., Cai, J. X., and Goldman-Rakic, P. S. (1988) The alpha-2 adrenergic agonist
19 guanfacine improves memory in aged monkeys without sedative or hypotensive side effects:
20 evidence for alpha-2 receptor subtypes. *J. Neurosci.* 8, 4287–4298.
21
22 (39) Tang, C., Biemond, I., and Lamers, C. B. (1996) Cholecystokinin receptors in human
23 pancreas and gallbladder muscle: a comparative study. *Gastroenterology* 111, 1621–1626.
24
25 (40) Feinle, C., Meier, O., Otto, B., D'Amato, M., and Fried, M. (2001) Role of duodenal lipid
26 and cholecystokinin A receptors in the pathophysiology of functional dyspepsia. *Gut* 48, 347–
27 355.
28
29 (41) Eberle-Wang, K., and Simansky, K. J. (1992) The CCK-A receptor antagonist, devazepide,
30 blocks the anorectic action of CCK but no peripheral serotonin in rats. *Pharmacol. Biochem.*
31 *Behav.* 43, 943–947.
32
33 (42) Zhang, J., Xiong, B., Zhen, X., and Zhang, A. (2008) Dopamine D1 receptor ligands: where
34 are we now and where are we going. *Med. Res. Rev.* 29, 272–294.
35
36 (43) Ishizuka, N., and Matsumura, K. (2002) Structure-activity relationships of a novel class of
37 endothelin-A receptor antagonists and discovery of potent and selective receptor antagonist, 2-
38 (benzo[1,3]dioxol-5-yl)-6-isopropoxy-4-(4-methoxyphenyl)-2H-chromene-3- carboxylic acid
39 (S-1255). 1. *Stud. J. Med. Chem.* 45, 2041–2055.
40
41 (44) Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., Katritch, V.,
42 Abagyan, R., Cherezov, V., Liu, W., Han, G. W., Kobayashi, T., Stevens, R. C., and Iwata, S.
43 (2011) Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475, 65–70.
44
45 (45) Langmead, C. J., Austin, N. E., Branch, C. L., Brown, J. T., Buchanan, K. A., Davies, C. H.,
46 Forbes, I. T., Fry, V. A. H., Hagan, J. J., Herdon, H. J., Jones, G. A., Jeggo, R., Kew, J. N. C.,
47 Mazzali, A., Melarange, R., Patel, N., Pardoe, J., Randall, A. D., Roberts, C., Roopun, A.,
48 Starr, K. R., Teriakidis, A., Wood, M. D., Whittington, M., Wu, Z., and Watson, J. (2008)
49 Characterization of a CNS penetrant, selective M1 muscarinic receptor agonist, 77-LH-28-1. *Br.*
50 *J. Pharmacol.* 154, 1104–1115.
51
52 (46) Vannucchi, M. G., Scali, C., Kopf, S. R., Pepeu, G., and Casamenti, F. (1997) Selective
53 muscarinic antagonists differentially affect in vivo acetylcholine release and memory
54 performances of young and aged rats. *Neuroscience* 79, 837–846.
55
56 (47) Baghdoyan, H. A., and Lydic, R. (1999) M2 muscarinic receptor subtype in the feline
57 medial pontine reticular formation modulates the amount of rapid eye movement sleep. *Sleep* 22,
58
59
60

835–847.

(48) Slack, B. E., Breu, J., Petryniak, M. A., Srivastava, K., and Wurtman, R. J. (1995) Tyrosine phosphorylation-dependent stimulation of amyloid precursor protein secretion by the m3 muscarinic acetylcholine receptor. *J. Biol. Chem.* *270*, 8337–8344.

(49) Haga, K., Kruse, A. C., Asada, H., Yurugi-Kobayashi, T., Shiroishi, M., Zhang, C., Weis, W. I., Okada, T., Kobilka, B. K., Haga, T., and Kobayashi, T. (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* *482*, 547–551.

(50) Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Mathiesen, J. M., Sunahara, R. K., Pardo, L., Weis, W. I., Kobilka, B. K., and Granier, S. (2012) Crystal structure of the μ -opioid receptor bound to a morphinan antagonist. *Nature* *485*, 321–326.

(51) Lemmens-Gruber, R., and Kamyar, M. (2006) Vasopressin antagonists. *Cell. Mol. Life Sci.* *63*, 1766–1779.

(52) Serradeil-Le Gal, C., Wagnon, J., Garcia, C., Lacour, C., Guiraudou, P., Christophe, B., Villanova, G., Nisato, D., Maffrand, J. P., Le Fur, G., Guillon, G., Cantau, B., Barberis, C., Trueba, M., Ala, Y., and Jard, S. (1993) Biochemical and pharmacological properties of SR 49059, a new, potent, nonpeptide antagonist of rat and human vasopressin V1a receptors. *J. Clin. Invest.* *92*, 224–231.

(53) Guillon, C. D., Koppel, G. A., Brownstein, M. J., Chaney, M. O., Ferris, C. F., Lu, S. F., Fabio, K. M., Miller, M. J., Heindel, N. D., Hunden, D. C., Cooper, R. D. G., Kaldor, S. W., Skelton, J. J., Dressman, B. A., Clay, M. P., Steinberg, M. I., Bruns, R. F., and Simon, N. G. (2007) Azetidiones as vasopressin V1a antagonists. *Bioorganic Med. Chem.* *15*, 2054–2080.

(54) Akaike, A., Takada-Takatori, Y., Kume, T., and Izumi, Y. (2010) Mechanisms of neuroprotective effects of nicotine and acetylcholinesterase inhibitors: Role of $\alpha 4$ and $\alpha 7$ receptors in neuroprotection. *J. Mol. Neurosci.* *40*, 211–216.

(55) Hawser, S., Lociuero, S., and Islam, K. (2006) Dihydrofolate reductase inhibitors as antibacterial agents. *Biochem. Pharmacol.* *71*, 941–948.

(56) Oefner, C., D'Arcy, A., and Winkler, F. K. (1988) Crystal structure of human dihydrofolate reductase complexed with folate. *Eur. J. Biochem.* *174*, 377–385.

(57) Miller, T. A., Witter, D. J., and Belvedere, S. (2003) Histone deacetylase inhibitors. *J. Med. Chem.* *46*, 5097–5116.

(58) Villar-Garea, A., and Esteller, M. (2004) Histone deacetylase inhibitors: understanding a new wave of anticancer agents. *Int. J. Cancer* *112*, 171–178.

(59) Wouters, J., Moureau, F., Evrard, G., Koenig, J. J., Jegham, S., George, P., and Durant, F. (1999) A reversible monoamine oxidase a inhibitor, befloxatone: Structural approach of its mechanism of action. *Bioorganic Med. Chem.* *7*, 1683–1693.

(60) Kan, J. P., Mouget-Goniot, C., Worms, P., and Biziere, K. (1986) Effect of the antidepressant minaprine on both forms of monoamine oxidase in the rat. *Biochem. Pharmacol.* *35*, 973–978.

(61) Lee, M. E., Markowitz, J., Lee, J. O., and Lee, H. (2002) Crystal structure of phosphodiesterase 4D and inhibitor complex. *FEBS Lett.* *530*, 53–58.

(62) Touroutoglou, N., and Pazdur, R. (1996) Thymidylate synthase inhibitors. *Clin. Cancer Res.*

1
2
3
4
2, 227–243.

5 (63) Lee, K. C., Ouwehand, I., Giannini, A. L., Thomas, N. S., Dibb, N. J., and Bijlmakers, M. J.
6 (2010) Lck is a key target of imatinib and dasatinib in T-cell activation. *Leukemia* 24, 896–900.

7
8 (64) Waldhauser, K. M., Brecht, K., Hebeisen, S., Ha, H. R., Konrad, D., Bur, D., and
9 Krähenbühl, S. (2008) Interaction with the hERG channel and cytotoxicity of amiodarone and
10 amiodarone analogues. *Br. J. Pharmacol.* 155, 585–595.

11 (65) Pearlstein, R. A., Vaz, R. J., Kang, J., Chen, X. L., Preobrazhenskaya, M., Shchekotikhin,
12 A. E., Korolev, A. M., Lysenkova, L. N., Miroshnikova, O. V., Hendrix, J., and Rampe, D.
13 (2003) Characterization of HERG potassium channel inhibition using CoMSiA 3D QSAR and
14 homology modeling approaches. *Bioorganic Med. Chem. Lett.* 13, 1829–1835.

15
16 (66) Ekins, S., Crumb, W. J., Sarazan, R. D., Wikel, J. H., and Wrighton, S. A. (2002) Three-
17 dimensional quantitative structure-activity relationship for inhibition of human ether-a-go-go-
18 related gene potassium channel. *J. Pharmacol. Exp. Ther.* 301, 427–434.

19 (67) Cavalli, A., Poluzzi, E., De Ponti, F., and Recanatini, M. (2002) Toward a pharmacophore
20 for drugs inducing the long QT syndrome: Insights from a CoMFA study of HERG K⁺ channel
21 blockers. *J. Med. Chem.* 45, 3844–3853.

22 (68) Kelley, S. P., Bratt, A. M., and Hodge, C. W. (2003) Targeted gene deletion of the 5-HT_{3A}
23 receptor subunit produces an anxiolytic phenotype in mice. *Eur. J. Pharmacol.* 461, 19–25.

24 (69) Mørk, A., Montezinho, L. P., Miller, S., Trippodi-Murphy, C., Plath, N., Li, Y., Gulinello,
25 M., and Sanchez, C. (2013) Vortioxetine (Lu AA21004), a novel multimodal antidepressant,
26 enhances memory in rats. *Pharmacol. Biochem. Behav.* 105, 41–50.

27 (70) Matias, P. M., Donner, P., Coelho, R., Thomaz, M., Peixoto, C., Macedo, S., Otto, N.,
28 Joschko, S., Scholz, P., Wegg, A., Bäsler, S., Schäfer, M., Egner, U., and Carrondo, M. A.
29 (2000) Structural evidence for ligand specificity in the binding domain of the human androgen
30 receptor. Implications for pathogenic gene mutations. *J. Biol. Chem.* 275, 26164–26171.

31 (71) Veldscholte, J., Berrevoets, C. A., Brinkmann, A. O., Grootegoed, J. A., and Mulder, E.
32 (1992) Anti-androgens and the mutated androgen receptor of LNCaP cells: differential effects on
33 binding affinity, heat-shock protein interaction, and transcription activation. *Biochemistry* 31,
34 2393–2399.

35 (72) Simard, J., Luthy, I., Guay, J., Bélanger, A., and Labrie, F. (1986) Characteristics of
36 interaction of the antiandrogen flutamide with the androgen receptor in various target tissues.
37 *Mol. Cell. Endocrinol.* 44, 261–270.

38 (73) Bledsoe, R. K., Montana, V. G., Stanley, T. B., Delves, C. J., Apolito, C. J., McKee, D. D.,
39 Consler, T. G., Parks, D. J., Stewart, E. L., Willson, T. M., Lambert, M. H., Moore, J. T., Pearce,
40 K. H., and Xu, H. E. (2002) Crystal structure of the glucocorticoid receptor ligand binding
41 domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* 110, 93–
42 105.

43 (74) Oomen, C. A., Mayer, J. L., De Kloet, E. R., Joëls, M., and Lucassen, P. J. (2007) Brief
44 treatment with the glucocorticoid receptor antagonist mifepristone normalizes the reduction in
45 neurogenesis after chronic stress. *Eur. J. Neurosci.* 26, 3395–3401.

46 (75) Honer, C., Nam, K., Fink, C., Marshall, P., Ksander, G., Chatelain, R. E., Cornell, W.,
47 Steele, R., Schweitzer, R., and Schumacher, C. (2003) Glucocorticoid receptor antagonism by
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 cyproterone acetate and RU486. *Mol. Pharmacol.* 63, 1012–1020.

4
5 (76) Ritz, M. C., Cone, E. J., and Kuhar, M. J. (1990) Cocaine inhibition of ligand binding at
6 dopamine, norepinephrine and serotonin transporters: a structure-activity study. *Life Sci.* 46,
7 635–645.

8
9 (77) Carroll, F. I., Lewin, A. H., Boja, J. W., and Kuhar, M. J. (1992) Cocaine receptor:
10 biochemical characterization and structure-activity relationships of cocaine analogues at the
11 dopamine transporter. *J. Med. Chem.* 35, 969–981.

12
13 (78) Beuming, T., Kniazeff, J., Bergmann, M. L., Shi, L., Gracia, L., Raniszewska, K., Newman,
14 A. H., Javitch, J. A., Weinstein, H., Gether, U., and Loland, C. J. (2008) The binding sites for
15 cocaine and dopamine in the dopamine transporter overlap. *Nat. Neurosci.* 11, 780–789.

16
17 (79) Florin, S. M., Kuczynski, R., and Segal, D. S. (1994) Regional extracellular norepinephrine
18 responses to amphetamine and cocaine and effects of clonidine pretreatment. *Brain Res.* 654, 53–
19 62.

20
21 (80) Segonzac, A., Raisman, R., Tateishi, T., Schoemaker, H., Hicks, P. E., and Langer, S. Z.
22 (1985) Tryptamine, a substrate for the serotonin transporter in human platelets, modifies the
23 dissociation kinetics of [3H]imipramine binding: Possible allosteric interaction. *J. Neurochem.*
24 44, 349–356.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FIGURES

Figure 1A

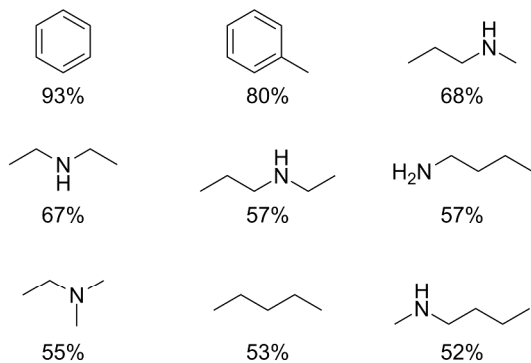


Figure 1B

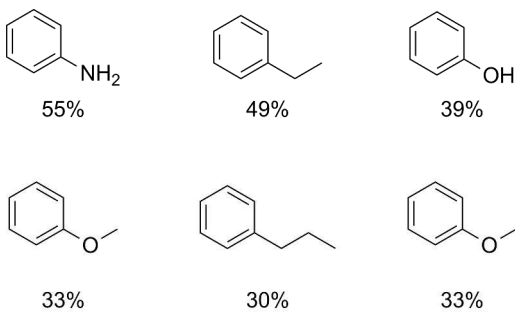


Figure 1C

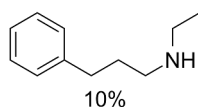


Figure 1. 2D fragments common within the entire dataset for this study. Part 1A shows fragments larger than four chemical bonds present within 50 % or greater of the data set. Part 1B shows fragments larger than seven chemical bonds present within 30 % or greater of the data set. Part 1C shows fragments larger than 12 chemical bonds present within 10 % or greater of the data set.

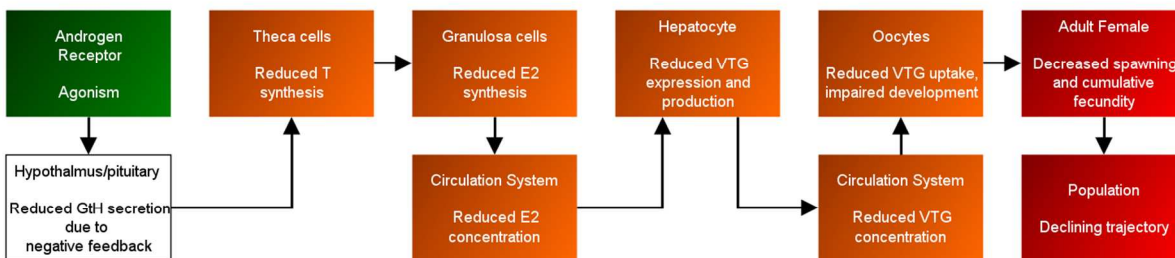


Figure 2. AOP for androgen receptor agonism leading to reproductive dysfunction. The MIE is shown in green, KEs in orange and adverse outcomes in red as per the AOP Wiki template. Reduced GtH secretion in the hypothalamus/pituitary is shown in white as there is uncertainty as to the specific mechanism through which androgen receptor agonism elicits a negative feedback response at this key event. E2: estradiol, GtH: gonadotropin hormone, T: testosterone, VTG: vitellogenin.

1
2
3 TABLES
4
5
6

7 Target

8 Binders Binders Binders Test Set Test Set

9 **G-Protein Coupled Receptors** Training Set (positive) (negative)

10

11 Adenosine A2a Receptor 2960 2150 810 10820

12

13

14 Alpha-1a Adrenergic Receptor 705 510 195 11435

15 Alpha-2a Adrenergic Receptor 380 269 111 11519

16 Beta-1 Adrenergic Receptor 694 505 189 11441

17 Beta-2 Adrenergic Receptor 770 540 230 11400

18

19

20

21 Cannabinoid CB1 Receptor 3738 2688 1050 10580

22 Cannabinoid CB2 Receptor 3405 2435 970 10660

23

24

25 Cholecystokinin Receptor A 255 177 78 11552

26

27 Dopamine D1 Receptor 453 322 131 11499

28 Dopamine D2 Receptor 2589 1831 758 10872

29

30

31 Endothelin Receptor A 100 72 28 11602

32

33

34 Histamine H1 Receptor 672 499 173 11457

35 Histamine H2 Receptor 191 141 50 11580

36

37

38 Muscarinic Acetylcholine Receptor M1 887 639 248 11382

39 Muscarinic Acetylcholine Receptor M2 620 456 164 11466

40 Muscarinic Acetylcholine Receptor M3 1067 772 295 11335

41

42

43 Delta Opioid Receptor 2550 1657 893 10737

44 Kappa Opioid Receptor 2347 1721 626 11004

45 Mu Opioid Receptor 2793 2056 737 10893

46

47

48 Serotonin 1A Receptor 1777 1273 504 11126

49 Serotonin 1B Receptor 396 282 114 11516

50 Serotonin 2A Receptor 1612 1162 450 11180

51 Serotonin 2B Receptor 750 555 195 11435

52

53

54 Vasopressin V1A Receptor 651 471 180 11450

55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60**Enzymes**

Acetylcholinesterase	1355	989	366	11264
Cyclooxygenase 1	379	274	105	11525
Cyclooxygenase 2	964	686	278	11352
Dihydrofolate Reductase	404	294	110	11520
GAR Transformylase	36	23	13	11617
Histone Deacetylase 1	1202	882	320	11310
Monoamine Oxidase A	533	382	151	11479
Phosphodiesterase 3A	132	94	38	11592
Phosphodiesterase 4D	385	267	118	11512
Thymidylate Synthase	239	178	61	11569
Tyrosine-Protein Kinase	568	414	154	11476

Ion Channels

Glutamate (NMDA) Receptor	25	18	7	11623
Potassium Voltage Gated Channel KQT-like Member 1	295	215	80	11550
Serotonin 3A Receptor	316	230	86	11544
Sodium Channel V Subunit Alpha	153	111	42	11588
Voltage Gated K Channel Subunit Kv7.1	28	21	7	11623

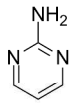
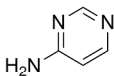
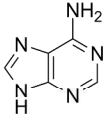
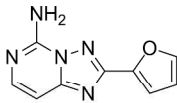
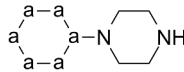
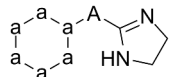
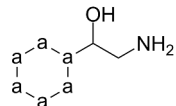
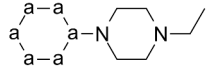
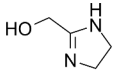
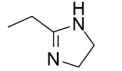
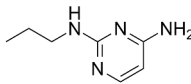
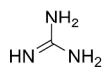
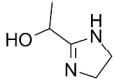
Nuclear Receptors

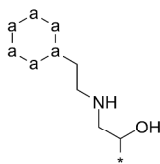
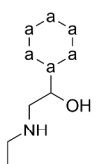
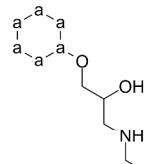
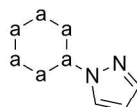
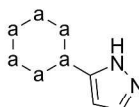
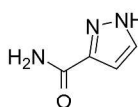
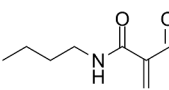
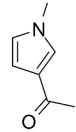
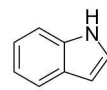
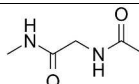
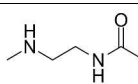
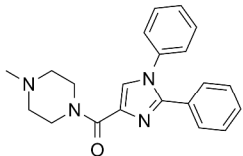
Androgen Receptor	1598	1186	412	11218
Glucocorticoid Receptor	2201	1622	579	11051

Transporters

Dopamine Transporter	1908	1414	494	11136
Norepinephrine Transporter	2616	1923	693	10937
Serotonin Transporter	3480	2578	902	10728

Table 1. Pharmacological targets analyzed in this work. Data was extracted from ChEMBL version 19. The total test set in each case was 11 630 compounds.

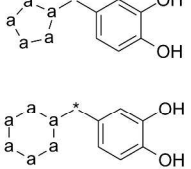
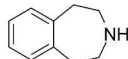
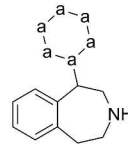
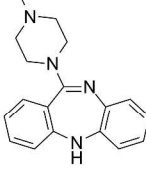


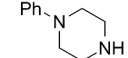
Receptor	MIE	Origin	Training Set Hits	Alerts
G-Protein Coupled Receptors				
Adenosine A2a Receptor				
	Alert AA2aR 1 - 2-Pyrimidine	S	842/2150	
	Alert AA2aR 2 - 4-Pyrimidine	S,L ^{25,26}	647/2150	
	Alert AA2aR 3 - Adenine	S,L ^{25,26}	349/2150	
	Alert AA2aR 4 - AA2aR Frag 1	S,L ²⁷	364/2150	
				Alert AA2aR 1 2-Aminopyrimidine
				Alert AA2aR 2 4-Aminopyrimidine
				Alert AA2aR 3 Adenine
				Alert AA2aR 4 AA2aR Frag 1
Adrenergic Receptors				
Alpha	Alert A-ADR 1 - Phenylpiperazine-like	S,L	207/758	
	Alert A-ADR 2 - Tolazoline-like	S,L ²⁸⁻³²	7/758	
				Alert A-ADR 1 Phenylpiperazine-like
				Alert A-ADR 2 Tolazoline-like
Alpha-1a	Alert A-1aADR 1 - Phenylethanolamine-like	L ^{29,33-36}	10/510	
	Alert A-1aADR 2 - 1-Ethyl-4-phenylpiperazine-like	S	158/510	
				Alert A-1aADR 1 Phenylethanolamine-like
				Alert A-1aADR 2 1-Ethyl-4-phenylpiperazine-like
Alpha-2a	Alert A-2aADR 1 - CID 5145436	S	55/269	
	Alert A-2aADR 2 - CID 13001	S	52/269	
	Alert A-2aADR 3 - A2aADR Frag 2	S	37/269	
	Alert A-2aADR 4 - 4-Benzyl-1H-imidazole-like	L ³⁷	3/269	
	Alert A-2aADR 5- Guanidine	L ^{31,38}	9/269	
	Alert A-2aADR 6 - 2-(1-Hydroxyl ethyl)-2-imidazole	S	50/269	
				Alert A-2aADR 1 CID 5145436 55/269
				Alert A-2aADR 2 CID 13001 52/269
				Alert A-2aADR 3 A2aADR Frag 2 37/269
				Alert A-2aADR 4 4-Benzyl-1H-imidazole-like 3/269
				Alert A-2aADR 5 Guanidine 9/269
				Alert A-2aADR 6 2-(1-Hydroxyl ethyl)-2-imidazole 50/269

Receptor	MIE	Origin	Training Set Hits	Alerts
Beta				
	Alert B-ADR 1 - 1-(Phenethylamino)propan-2-ol-like	S	383/774	
	Alert B-ADR 2 - 2-(Ethylamino)-1-phenylethanol-like	S	509/774	
	Alert B-ADR 3 - 1-(Ethylamino)-3-phenoxy-2-propanol-like	S	194/774	
Cannabinoid Receptor				
CB1				
	Alert CCB1R 1 - 2-Phenylpyrazole-like	S	391/2688	
	Alert CCB1R 2 - 5-Phenyl-1H-pyrazole-like	S	286/2688	
	Alert CCB1R 3 - Pyrazole-3-carboxamide	S	321/2688	
CB2				
	Alert CCB2R 1 - CCB2R Frag 1	S	298/2435	
	Alert CCB2R 2 - 1-(1-Ethyl-1H-pyrrol-3-yl)ethanone	S	217/2435	
	Alert CR 1 - Indole	S	489/3362	
Cholecystokin A Receptor				
	Alert CCKAR 1 - 2-Acetamido-N-methylacetamide	S,L ³⁹⁻⁴¹	72/177	
	Alert CCKAR 2 - N-[2-(methylamino)ethyl]acetamide	S,L ³⁹⁻⁴¹	159/177	
	Alert CCKAR 3 - CID 9957635	S	63/177	

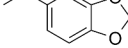
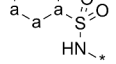
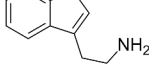
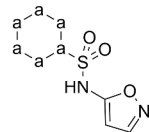
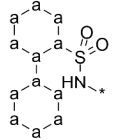
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Receptor	MIE	Origin	Training Set Hits	Alerts
----------	-----	--------	-------------------	--------

Dopamine Receptors

D1	Alert DD1R 1 - Dihydroxine-like	L ⁴²	21/322				
	Alert DD1R 2 - Benzazepine-like	L ⁴²	20/322				
	Alert DD1R 3 - Benzazepine-like with aromatic	L ⁴²	19/322				
	Alert DD1R 4 - CID 15288	S	84/322				
	Alert DD1R 1 - Dihydroxine-like						
	Alert DD1R 2 - Benzazepine-like						
	Alert DD1R 3 - Benzazepine-like with aromatic						
	Alert DD1R 4 - CID 15288						
D2	Alert DD2R 1 - Piperazine	S	1006/1831				
	Alert DD2R 2 - 1,4 Dimethylpiperazine	S	1000/1831				
	Alert DD2R 3 - Phenylpiperazine	S	727/1831				

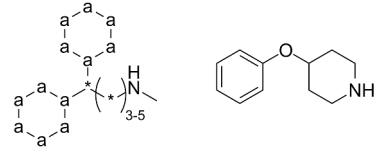
Endothelin Receptor A

	Alert ER ET-A 1 - 5-Ethyl-1,3-benzodioxole	S,L ⁴³	23/72			
	Alert ER ET-A 2 - N-Methylbenzenesulfonamide-like	S	53/72			
	Alert ER ET-A 3 - Tryptamine	S	25/72			
	Alert ER ET-A 4 - ER ET-A Frag 1-like	S	30/72			
	Alert ER ET-A 5 - N-Methyl-2-biphenylsulfonamide-like	S	25/72			
	Alert ER ET-A 4 - ER ET-A Frag 1-like					
	Alert ER ET-A 5 - N-Methyl-2-biphenylsulfonamide					

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Receptor	MIE	Origin	Training Set Hits	Alerts
----------	-----	--------	-------------------	--------

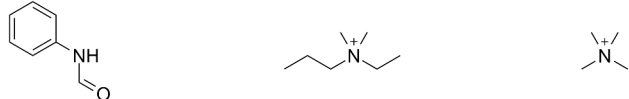
Histamine Receptors

H1	Alert HH1R 1 - Doxepine-like	S,X ⁴⁴	58/444	
	Alert HH1R 2 - 4-Phenoxypiperidine	S	66/444	

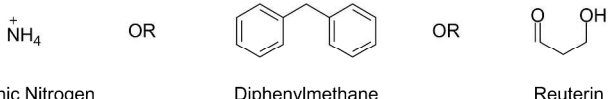
Alert HH1R 1
Doxepine-likeAlert HH1R 2
4-Phenoxypiperidine

H2	Alert HH2R 1 - Imidazole	S	36/126	
	Alert HH2R 2 - Guanidine	S	46/126	
	Alert HH2R 4 - Indole	S	36/126	

Alert HH2R 1
ImidazoleAlert HH2R 2
GuanidineAlert HH2R 4
Indole**Muscarinic Acetylcholine Receptors**

M1 and M2	Alert MAR 1 - Formanilide	S,L ^{45,46}	238/844	
M2 and M3	Alert MAR 2 - N-Ethyl-N,N-dimethylpropanaminium	S,L ^{47,48}	316/915	

Alert MAR 1 (M1 and M2)
FormanilideAlert MAR 2 (M2 and M3)
N-Ethyl-N,N-dimethylpropanaminiumAlert MAR 3 (M3)
Tetramethylamonium

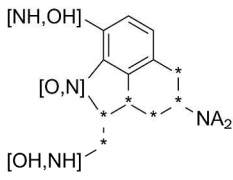
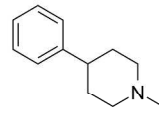
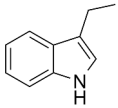
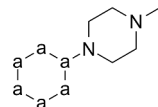
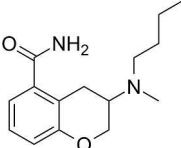
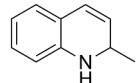
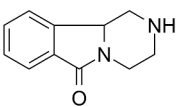
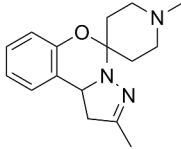
M3	Alert MAR 3 - Tetramethylamonium	S,L ^{47,48}	316/772	
All	Alert MAR 4 - MAR Pharmacophore	X ⁴⁹	393/1236	

Cationic Nitrogen

Diphenylmethane

Reuterin

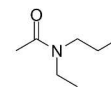
Alert MAR 4 (M1, M2 or M3)
MAR Pharmacophore

Receptor	MIE	Origin	Training Set Hits	Alerts	
Opioid Receptors					
	Alert OR 1 - Morphine-like (4 or more)	X,P ⁵⁰	164/1385		
	Alert OR 1 - Morphine-like (3 or more)	X,P ⁵⁰	459/1385		
	Alert OR 2 - 1-Methyl-4-phenylpiperidine	S	668/1385	Alert OR 1 Morphine-like	Alert OR 2 1-Methyl-4-phenylpiperidine
Serotonin Receptors					
All	Alert SR 1 - 3-Ethyl Indole	S	299/2633		
1A, 1B, 2A	Alert SR 1a1b2a 1 - 1-Methyl-4-Phenylpiperazine-like	S	908/2377	Alert SR 1 3-Ethyl Indole	Alert SR 1a1b2a 1 1-Methyl-4-phenylpiperazine
1A	Alert SR 1a 1 - SR 1a Frag 2	S	65/1208		
1B	Alert SR 1b 1 - 2-Methyl-1,2-dihydroquinoline	S	116/282	Alert SR 1a 1 SR 1a Frag 2	Alert SR 1b 1 2-methyl-1,2-dihydroquinoline
2B	Alert SR 2b 1 - CID 15206310	S	41/555		
	Alert SR 2b 2 - SR 2b Frag 2	S	31/555	Alert SR 2b 1 CID 15206310	Alert SR 2b 2 SR 2b Frag 2

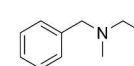
1
2
3
4 **Receptor** **MIE** **Origin** **Training Set Hits** **Alerts**

5
6
7
8 **Vasopressin V1a**

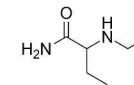
9 Alert VV1a 1 - VV1a Frag 1 S,L⁵¹⁻⁵³ 224/471
10
11 Alert VV1a 2 - n-benzyl-n-ethylmethylamine S 243/471
12
13 Alert VV1a 3 - VV1a Frag 3 S,L⁵¹⁻⁵³ 187/471
14
15 Alert VV1a 4 - VV1a Frag 4 S,L⁵¹⁻⁵³ 128/471



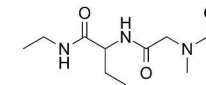
Alert VV1a 1
VV1a Frag 1



Alert VV1a 2
n-Benzyl-n-ethylmethylamine



Alert VV1a 3
VV1a Frag 3

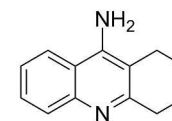


Alert VV1a 4
VV1a Frag 4

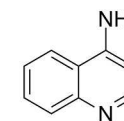
16
17 **Enzymes**

18
19
20
21 **Acetylcholinesterase**

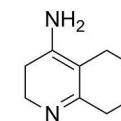
22
23
24 Alert AC 1a - Tacrine S,L⁵⁴ 184/989
25
26 Alert AC 1b - 4-Quinolinamine S,L⁵⁴ 192/989
27
28 Alert AC 1c - AC Frag 1 S,L⁵⁴ 245/989
29
30 Alert AC 1d - Dimethyl benzylamine S,L⁵⁴ 395/989
31
32 Alert AC 1e - Dimethyl benzylamine-like SMARTS S,L⁵⁴ 401/989



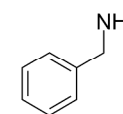
Alert AC 1a
Tacrine



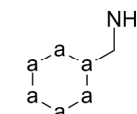
Alert AC 1b
4-Quinolinamine



Alert AC 1c
AC Frag 1



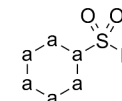
Alert AC 1d
Dimethyl benzylamine



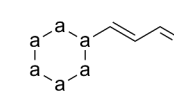
Alert AC 1e
Dimethyl benzylamine-like SMARTS

33
34
35 **Cyclooxygenases**

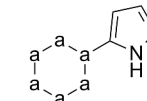
36
37
38 Alert COX 1 - Sulfonated aromatic rings S 392/816
39
40 Alert COX 2 - Cinnamaldehyde-like S 183/816
41
42 Alert COX 3 - 5-Phenyl-1H-pyrazole-like S 70/816



Alert COX 1
Sulfonated aromatic rings



Alert COX 2
Cinnamaldehyde-like



Alert COX 3
5-Phenyl-1H-pyrazole-like

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

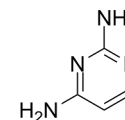
Receptor	MIE	Origin	Training Set Hits	Alerts
----------	-----	--------	-------------------	--------

Dihydrofolate Reductase

Alert DHFR 1 (SMILES) - Diaminopyrimidine

S,L⁵⁵,X⁵⁶

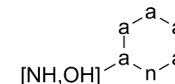
145/237

Alert DHFR 1 (SMILES)
Diaminopyrimidine

Alert DHFR 1 (SMARTS) - Diaminopyrimidine-like

S,L⁵⁵,X⁵⁶

158/237

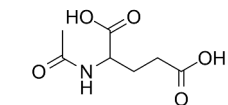
Alert DHFR 1 (SMARTS)
Diaminopyrimidine-like

GAR Transformylase

Alert GART 1 - N-Acetyl-DL-glutamic acid

S

23/23

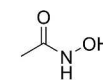
Alert GART 1
N-Acetyl-DL-glutamic acid

Histone Deacetylase 1

Alert HDAC 1 1 - Hydroxamic acids

S,L^{57,58}

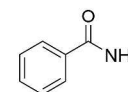
416/882

Alert HDAC 1
Acetohydroxamic acid

Alert HDAC 1 2a - Benzamide

S,L^{57,58}

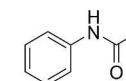
271/882

Alert HDAC 2a
Benzamide

Alert HDAC 1 2b - Benzamide-like SMARTS

S,L^{57,58}

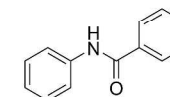
316/882

Alert HDAC 2c
Acetanilide

Alert HDAC 1 2c - Acetanilide

S,L^{57,58}

355/882

Alert HDAC 2e
Benzanilide

Alert HDAC 1 2d - Acetanilide-like SMARTS

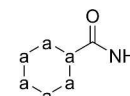
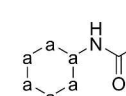
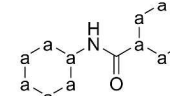
S,L^{57,58}

372/882

Alert HDAC 1 2e - Benzanilide

S,L^{57,58}

165/882

Alert HDAC 2b
Benzamide-likeAlert HDAC 2d
Acetanilide-likeAlert HDAC 2f
Benzanilide-like

Alert HDAC 1 2f - Benzanilide-like SMARTS

S,L^{57,58}

198/882

1
2
3
4 **Receptor** **MIE** **Origin** **Training Set Hits** **Alerts**

5
6
7
8 **Monoamine Oxidase**

9
10
11 Alert MAO A 1 - 3-Phenyl-2-oxazolidone

L^{59,60}

12/382



Alert MAO A 1
3-Phenyl-2-oxazolidone

12 Alert MAO A 2a - 4-Methylphenol

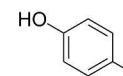
S

99/382

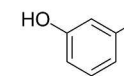
13 Alert MAO A 2b - 3-Methylphenol

S

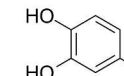
80/382



Alert MAO A 2a
4-Methylphenol



Alert MAO A 2b
3-Methylphenol



Alert MAO A 2c
Methyl-3,4-diphenol

14 Alert MAO A 2c - Methyl-3,4-diphenol

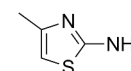
S

10/382

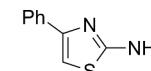
15 Alert MAO A 3a - Aminomethiazole

S

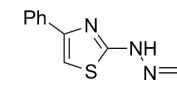
58/382



Alert MAO A 3a
Aminomethiazole



Alert MAO A 3b
4-Phenyl-1,3-thiazole-2-amine



Alert MAO A 3c
CID 7958070

16 Alert MAO A 3b - 4-Phenyl-1,3-thiazole-2-amine

S

57/382

17 Alert MAO A 3c - CID 7958070

S

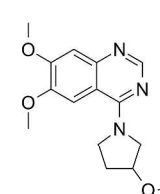
54/382

18
19
20
21
22
23
24
25 **Phosphodiesterases**

26
27
28
29 **3A** Alert PDE 3A 1 - PDE 3A Frag 2

S,X⁶¹

18/94

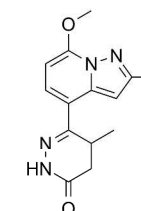


Alert PDE 3A 1
PDE 3A Frag 2

30 Alert PDE 3A 2a - PDE Frag 3

S,X⁶¹

11/94

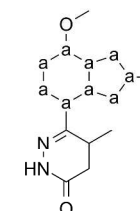


Alert PDE 3A 2a
PDE 3A Frag 3

31 Alert PDE 3A 2b - PDE Frag 3-like SMARTS

S,X⁶¹

15/94

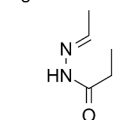


Alert PDE 3A 2b
PDE 3A Frag-like SMARTS

32 Alert PDE 3A 3 - PDE 3A Frag 4

S

34/94

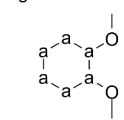


Alert PDE 3A 3
PDE 3A Frag 4

33 Alert PDE 3A 4 - Veratrol-like

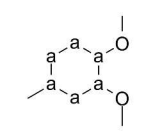
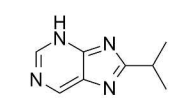
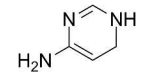
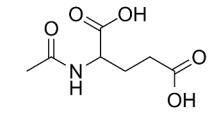

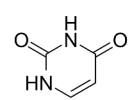
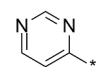
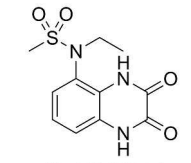
S,X⁶¹

36/94



Alert PDE 3A 4
Veratrol-like

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Receptor	MIE	Origin	Training Set Hits	Alerts
4D	Alert PDE 4D 1 - Homoveratrol-like	S,X ⁶¹	83/267	
	Alert PDE 4D 2 - 8-Isopropyl-1H-purine Frag	S	24/267	
	Alert PDE 4D 3 - PDE 4D Frag 3	S	74/267	
				Alert PDE 4D 1 Homoveratrol-like
				Alert PDE 4D 2 8-Isopropyl-1H-purine Frag
				Alert PDE 4D 3 PDE 4D Frag 3
Thymidylate Synthase				
	Alert TS 1 - N-Acetyl-DL-glutamic acid	S,L ⁶²	21/145	
	Alert TS 2 - Pyrimidine	L ⁶²	8/145	
	Alert TS 3 - Uracil	L ⁶²	2/145	
				Alert TS 1 N-Acetyl-DL-glutamic acid
				Alert TS 2 Pyrimidine
				Alert TS 3 Uracil
Tyrosine Protein Kinase				
	Alert LCK 1 - Pyrimidine-like	S,L ⁶³	212/383	
				Alert LCK 1 Pyrimidine
Ion Channels				
Glutamate (NMDA) Receptor				
	Alerts GRSZ1 1 - GRSZ1 Frag 1	S	18/18	
				Alert GRSZ1 1 GRSZ1 Frag 1

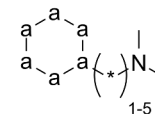
1
2
3
4 **Receptor** **MIE** **Origin** **Training Set Hits** **Alerts**

5
6
7
8 **Potassium Voltage Gated Channel Subunit H Member 2**

9
10
11 Alert HERG 1a - Flex Aromatic Amine 1-5

S,P⁶⁴⁻⁶⁷

167/215

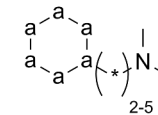


Alert HERG 1a
Flex Aromatic Amine 1-5

12 Alert HERG 1b - Flex Aromatic Amine 2-5

S,P⁶⁴⁻⁶⁷

147/215



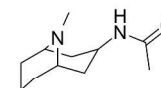
Alert HERG 1b
Flex Aromatic Amine 2-5

14
15
16
17
18 **Serotonin 3A Receptor**

19
20 Alert S3AR 1 - S3AR Frag 1

S,L⁶⁸

47/229

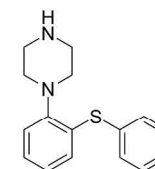


Alert S3AR 1
5HT 3a Frag 1

21 Alert S3AR 2 - 1-[2-(Phenylsulfanyl)phenyl]piperazine

S,L⁶⁹

27/229



Alert S3AR 2
1-[2-(Phenylsulfanyl)phenyl]piperazine

22
23
24
25
26
27 **Sodium Channel V Subunit alpha**

28
29
30 Alert SCV-A 1a - Pyrimidine-like aromatic

S

49/111



Alert SCV-A 1a
Pyrimidine-like
aromatic

31 Alert SCV-A 1b - Pyrimidine-like wildcard

S

49/111

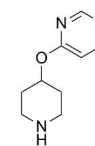


Alert SCV-A 1b
Pyrimidine-like
wildcard

32 Alert SCV-A 2 - 4-(4-Piperidineoxy)pyrimidine

S

30/111

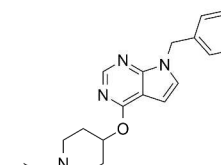


Alert SCV-A 2
4-(4-Piperidineoxy)pyrimidine

33 Alert SCV-A 3 - SCV Frag 1

S

28/111

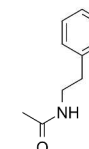


Alert SCV-A 3
SCV Frag 1

34 Alert SCV-A 4 - N-(2-Phenylethyl)acetamide

S

25/111



Alert SCV-A 4
N-(2-Phenylethyl)acetamide

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Receptor	MIE	Origin	Training Set Hits	Alerts
----------	-----	--------	-------------------	--------

Voltage Gated K Channel Sub Unit Kv7.1

Alert VGKC Kv7.1 1 - VGKV Kv7.1 Frag 1

S

21/21

**Nuclear Receptors****Androgen Receptor**

Alert AR 1a - Benzonitrile

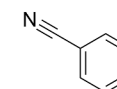
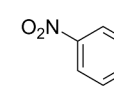
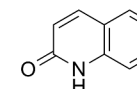
S,L⁷⁰

294/764

Alert AR 1b - Nitrobenzene

S,L⁷⁰⁻⁷²

73/764

Alert AR 1a
BenzonitrileAlert AR 1b
NitrobenzeneAlert AR 1c
Quinolone

Alert AR 1c - Quinolone

S,L⁷⁰

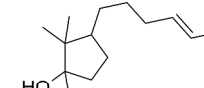
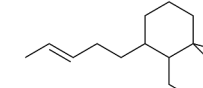
128/764

Glucocorticoid Receptor

Alert GR 1a - GR Frag 1

S,L⁷³⁻⁷⁵

228/1232

Alert GR 1a
GR Frag 1Alert GR 1b
tert-ButylcyclohexaneAlert GR 1c
GR Frag 4Alert GR 1d
GR Frag 5

Alert GR 1b - tert-Butylcyclohexane

S,L⁷³⁻⁷⁵

166/1232

Alert GR 1c - GR Frag 4

S,L⁷³⁻⁷⁵

201/1232

Alert GR 1d - GR Frag 5

S,L⁷³⁻⁷⁵

198/1232

1
2
3
4 **Receptor** **MIE** **Origin** **Training Set Hits** **Alerts**
5
6
7

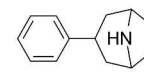
8 **Transporters**

9
10
11 **Dopamine Transporter**
12

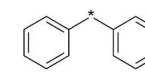
13
14 Alert DT 1 - 3-Phenyl-8-azabicyclo[3.2.1]octane

S,L^{76,77}

147/1159



Alert DT 1
3-Phenyl-8-azabicyclo[3.2.1]octane



Alert DT 2
Diphenylmethane-like

15 Alert DT 2 - Diphenylmethane-like

S,L⁷⁸

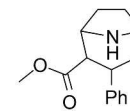
268/1159

16
17
18
19 **Norepinephrine Transporter**
20

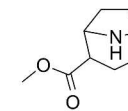
21
22 Alert NT 1a - Cocaine-like 1

S,L⁷⁹

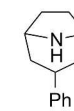
60/1573



Alert NT 1a
Cocaine-like 1



Alert NT 1b
Cocaine-like 2



Alert NT 1c
Cocaine-like 3



Alert NT 1d
Cocaine-like 4

23 Alert NT 1b - Cocaine-like 2

S,L⁷⁹

61/1573

24 Alert NT 1c - Cocaine-like 3

S,L⁷⁹

106/1573

25 Alert NT 1d - Cocaine-like 4

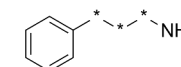
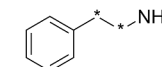
S,L⁷⁹

117/1573

26 Alert NT 2 - Amphetamines-like

S,L⁷⁹

1164/1573



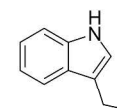
Alert NT 2
Amphetamines-like

27
28
29
30
31
32
33
34 **Serotonin Transporter**
35

36 Alert ST 1 - 3-Ethyl-indole

S,L⁸⁰

175/2166



Alert ST 1
3-Ethyl-indole

37 Alert ST 2 - DMEA

S,L⁸⁰

1438/2166

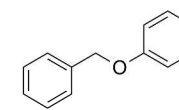


Alert ST 2
DMEA

38 Alert ST 3 - Benzyloxybenzene

S,L⁸⁰

172/2166

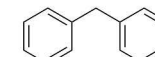


Alert ST 3
Benzyloxybenzene

39 Alert ST 4 - Diphenylmethane

S,L⁸⁰

192/2166



Alert ST 4
Diphenylmethane

1
2
3 **Table 2.** Structural alerts developed for each receptor with the fraction of hits within each training set shown. Alerts are labelled based
4 on their origin, S for fragments identified by maximal common substructure search using Pipeline Pilot, L for fragments found in
5 typical ligands found in the literature, P for fragments based on an existing pharmacophore and X for fragments derived from an
6 existing crystal structure. Alerts are shown pictorially as they have been coded in SMARTS or SMILES, with aromatic heavy atoms
7 depicted as (a), aliphatic heavy atoms depicted as (A), and wildcard atoms (any heavy atom) depicted as (*).
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Receptor	MIE	TP	FN	TN	FP	SE	SP	Q	MCC
G-Protein Coupled Receptors									
Adenosine A2a Receptor									
	Alert AA2aR 1 - 2-Aminopyrimidine	319	491	10591	229	39.4	97.9	93.8	0.448
	Alert AA2aR 2 - 4-Aminopyrimidine	251	559	10441	379	31.0	96.5	91.9	0.309
	Alert AA2aR 1 OR 2	505	305	10343	477	62.3	95.6	93.3	0.530
	Alerts AA2aR 1 AND 2	65	745	10689	131	8.0	98.8	92.5	0.135
	Alert AA2aR 3 - Adenine	137	673	10771	49	16.9	99.5	93.8	0.334
	Alert AA2aR 4 - AA2aR Frag 1	155	655	10820	0	19.1	100.0	94.4	0.425
Adrenergic Receptors									
Alpha-1a/-2a	Alert A-ADR 1 - Phenylpiperazine-like	69	234	10318	1009	22.8	91.1	89.3	0.076
	Alert A-ADR 2 - Tolazoline-like	6	297	11326	1	2.0	100.0	97.4	0.128
Alpha-1a	Alert A-1aADR 1 - Phenylethanolamine-like	6	189	11136	299	3.1	97.4	95.8	0.004
	Alert A-1aADR 2 - 1-Ethyl-4-phenylpiperazine-like	54	141	10705	730	27.7	93.6	92.5	0.109
Alpha-2a	Alert A-2aADR 1 - CID 5145436	24	86	11513	7	21.8	99.9	99.2	0.408
	Alert A-2aADR 2 - CID 13001	22	88	11507	13	20.0	99.9	99.1	0.351
	Alert A-2aADR 1 OR 2	24	86	11505	15	21.8	99.9	99.1	0.363
	Alert A-2aADR 3 - A2aADR Frag 2	6	104	11484	36	5.5	99.7	98.8	0.083
	Alert A-2aADR 4 - 4-Benzyl-1H-imidazole-like	3	107	11515	5	2.7	100.0	99.0	0.099
	Alert A-2aADR 5- Guanidine	10	100	11342	178	9.1	98.5	97.6	0.058
	Alert A-2aADR 6 - 2-(1-Hydroxyl ethyl)-2-imimidazole	22	88	11515	5	20.0	100.0	99.2	0.401

Receptor	MIE	TP	FN	TN	FP	SE	SP	Q	MCC
Beta	Alert B-ADR 1 - 1-(Phenethylamino)propan-2-ol-like	168	203	11154	105	45.3	99.1	97.4	0.515
	Alert B-ADR 2 - 2-(Ethylamino)-1-phenylethanol-like	242	129	11207	52	65.2	99.5	98.4	0.725
	Alert B-ADR 3 - 1-(Ethylamino)-3-phenoxy-2-propanol-like	102	269	11178	81	27.5	99.3	97.0	0.378
	Alert B-ADR 1 OR 2 OR 3	350	21	11025	234	94.3	97.9	97.8	0.742
Cannabinoid Receptor									
CB1	Alert CCB1R 1 - 2-Phenylpyrazole-like	170	880	10296	284	16.2	97.3	90.0	0.200
	Alert CCB1R 2 - 5-Phenyl-1H-pyrazole-like	117	933	10445	135	11.1	98.7	90.8	0.194
	Alert CCB1R 3 - Pyrazole-3-carboxamide	148	902	10533	47	14.1	99.6	91.8	0.305
	Alert CCB1R 1 OR 2	178	872	10210	370	17.0	96.5	89.3	0.182
	Alerts CCB1R 1 AND 2	109	941	10531	49	10.4	99.5	91.5	0.246
	Alerts CCB1R 1, 2 AND 3	86	964	10576	4	8.2	100.0	91.7	0.267
CB2	Alert CCB2R 1 - CCB2R Frag 1	111	859	10612	48	11.4	99.5	92.2	0.262
	Alert CCB2R 2 - 1-(1-Ethyl-1H-pyrrol-3-yl)ethanone	96	874	10621	39	9.9	99.6	92.1	0.246
CB1 & CB2	Alert CR 1 - Indole	207	1594	9043	786	11.5	92.0	79.5	0.045
Cholecystokinin A Receptor									
	Alert CCKAR 1 - 2-Acetamido-N-methylacetamide	35	43	11209	343	44.9	97.0	96.7	0.193
	Alert CCKAR 2 - N-[2-(methylamino)ethyl]acetamide	76	2	10595	957	97.4	91.7	91.8	0.256
	Alert CCKAR 3 - CID 9957635	30	48	11552	0	38.5	100.0	99.6	0.619

Receptor	MIE	TP	FN	TN	FP	SE	SP	Q	MCC
Dopamine Receptor									
D1	Alert DD1R 1 - Dihydroxine-like	10	121	11447	52	7.6	99.5	98.5	0.104
	Alert DD1R 2 - Benzazepine-like	9	122	11454	45	6.9	99.6	98.6	0.101
	Alert DD1R 3 - Benzazepine-like with aromatic	9	122	11499	0	6.9	100.0	99.0	0.261
	Alert DD1R 4 - CID 15288	32	99	11472	27	24.4	99.8	98.9	0.359
D2	Alert DD2R 1 - Piperazine	417	341	9678	1194	55.0	89.0	86.8	0.315
	Alert DD2R 2 - 1,4 Dimethylpiperazine	416	342	9928	944	54.9	91.3	88.9	0.355
	Alert DD2R 3 - Phenylpiperazine	321	437	10285	587	42.3	94.6	91.2	0.340
Endothelin Receptor A									
	Alert ER ET-A 1 - 5-Ethyl-1,3-benzodioxole	3	25	11573	29	10.7	99.8	99.5	0.098
	Alert ER ET-A 2 - N-Methylbenzenesulfonamide-like	18	10	11107	495	64.3	95.7	95.7	0.143
	Alert ER ET-A 3 - Tryptamine	8	20	11340	262	28.6	97.7	97.6	0.086
	Alert ER ET-A 4 - ER ET-A Frag 1-like	14	14	11602	0	50.0	100.0	99.9	0.707
	Alert ER ET-A 5 - N-Methyl-2-biphenylsulfonamide-like	13	15	11585	17	46.4	99.9	99.7	0.447
	Alert ER ET-A 4 OR 5	14	14	11585	17	50.0	99.9	99.7	0.474
Histamine Receptors									
H1	Alert HH1R 1 - Doxepine-like	24	134	11106	366	15.2	96.8	95.7	0.077
	Alert HH1R 2 - 4-Phenoxypiperidine	21	137	11412	60	13.3	99.5	98.3	0.178
H2	Alert HH2R 1 - Imidazole	15	33	10635	947	31.3	91.8	91.6	0.054
	Alert HH2R 2 - Guanidine	14	34	11408	174	29.2	98.5	98.2	0.141
	Alert HH2R 4 - Indole	20	28	10609	973	41.7	91.6	91.4	0.076
	Alert HH2R 1, 2 OR 4	38	10	9558	2024	79.2	82.5	82.5	0.104

Receptor	MIE	TP	FN	TN	FP	SE	SP	Q	MCC
Muscarinic Acetylcholine Receptors									
M1 and M2	Alert MAR 1 - Formanilide	102	270	9782	1476	27.4	86.9	85.0	0.074
M2 and M3	Alert MAR 2 - N-Ethyl-N,N-dimethylpropanaminum	129	282	11175	44	31.4	99.6	97.2	0.473
M3	Alert MAR 3 - Tetramethylamonium	118	177	11245	90	40.0	99.2	97.7	0.465
All	Alert MAR 4 - MAR Pharmacophore	198	394	10541	497	33.4	95.5	92.3	0.268
Opioid Receptors									
	Alert OR 1 - Morphine-like (4 or more)	164	1221	10238	7	11.8	99.9	89.4	0.317
	Alert OR 1 - Morphine-like (3 or more)	459	926	9867	378	33.1	96.3	88.8	0.369
	Alert OR 2 - 1-Methyl-4-phenylpiperidine	668	717	9957	288	48.2	97.2	91.4	0.536
Serotonin Receptors									
All	Alert SR 1 - 3-Ethyl Indole	125	1041	9946	518	10.7	95.0	86.6	0.076
1A, 1B, 2A	Alert SR 1a1b2a 1 - 1-Methyl-4-Phenylpiperazine-like	397	621	10011	601	39.0	94.3	89.5	0.336
1A	Alert SR 1a 1 - SR 1a Frag 2	33	471	11118	8	6.5	99.9	95.9	0.222
1B	Alert SR 1b 1 - 2-Methyl-1,2-dihydroquinoline	46	68	11323	193	40.4	98.3	97.8	0.269
2B	Alert SR 2b 1 - CID 15206310	15	180	11425	10	7.7	99.9	98.4	0.211
	Alert SR 2b 2 - SR 2b Frag 2	8	187	11435	0	4.1	100.0	98.4	0.201

Receptor	MIE	TP	FN	TN	FP	SE	SP	Q	MCC
Vasopressin V1a									
	Alert VV1a 1 - VV1a Frag 1	70	110	10811	639	38.9	94.4	93.6	0.172
	Alert VV1a 2 - n-benzyl-n-ethylmethylamine	90	90	9928	1522	50.0	86.7	86.1	0.131
	Alert VV1a 3 - VV1a Frag 3	66	114	11144	306	36.7	97.3	96.4	0.238
	Alert VV1a 4 - VV1a Frag 4	41	139	11372	78	22.8	99.3	98.1	0.271
Enzymes									
Acetylcholinesterase									
	Alert AC 1a - Tacrine	74	292	11261	3	20.2	100.0	97.5	0.435
	Alert AC 1b - 4-Quinolinamine	77	289	11233	31	21.0	99.7	97.2	0.378
	Alert AC 1c - AC Frag 1	97	269	11261	3	26.5	100.0	97.7	0.501
	Alert AC 1d - Dimethyl benzylamine	127	239	9596	1668	34.7	85.2	83.6	0.096
	Alert AC 1e - Dimethyl benzylamine-like SMARTS	142	224	9710	1554	38.8	86.2	84.7	0.124
Cyclooxygenases									
	Alert COX 1 - Sulfonated aromatic rings	187	168	10567	708	52.7	93.7	92.5	0.299
	Alert COX 2 - Cinnamaldehyde-like	58	297	10912	363	16.3	96.8	94.3	0.121
	Alert COX 3 - 5-Phenyl-1H-pyrazole-like	34	321	11057	218	9.6	98.1	95.4	0.090
	Alerts COX 1 OR 2	226	129	10218	1057	63.7	90.6	89.8	0.298
Dihydrofolate Reductase									
	Alert DHFR 1 (SMILES) - Diaminopyrimidine	63	37	11397	133	63.0	98.8	98.5	0.444
	Alert DHFR 1 (SMARTS) - Diaminopyrimidine-like	65	35	10535	995	65.0	91.4	91.1	0.181

Receptor	MIE	TP	FN	TN	FP	SE	SP	Q	MCC
GAR Transformylase									
	Alert GART 1 - N-Acetyl-DL-glutamic acid	10	3	11592	25	76.9	99.8	99.8	0.468
Histone Deacetylase 1									
	Alert HDAC 1 1 - Hydroxamic acids	147	173	11296	14	45.9	99.9	98.4	0.641
	Alert HDAC 1 2a - Benzamide	83	237	10066	1244	25.9	89.0	87.3	0.077
	Alert HDAC 1 2b - Benzamide-like SMARTS	100	220	9846	1464	31.3	87.1	85.5	0.088
	Alert HDAC 1 2c - Acetanilide	127	193	10380	930	39.7	91.8	90.3	0.179
	Alert HDAC 1 2d - Acetanilide-like SMARTS	135	185	9654	1656	42.2	85.4	84.2	0.125
	Alert HDAC 1 2e - Benanilide	61	259	11173	137	19.1	98.8	96.6	0.226
	Alert HDAC 1 2f - Benanilide-like SMARTS	73	247	11135	175	22.8	98.5	96.4	0.241
Monoamine Oxidase									
	Alert MAO A 1 - 3-Phenyl-2-oxazolidone	6	145	11477	2	4.0	100.0	98.7	0.171
	Alert MAO A 2a - 4-Methylphenol	37	114	9431	2048	24.5	82.2	81.4	0.020
	Alert MAO A 2b - 3-Methylphenol	35	116	9698	1781	23.2	84.5	83.7	0.024
	Alert MAO A 2c - Methyl-3,4-diphenol	3	148	10901	578	2.0	95.0	93.8	-0.016
	Alert MAO A 3a - Aminomethaziale	26	125	11415	64	17.2	99.4	98.4	0.215
	Alert MAO A 3b - 4-Phenyl-1,3-thiazole-2-amine	24	127	11452	27	15.9	99.8	98.7	0.268
	Alert MAO A 3c - CID 7958070	24	127	11479	0	15.9	100.0	98.9	0.396
Phosphodiesterases									
3A	Alert PDE 3A 1 - PDE 3A Frag 2	9	29	11592	0	23.7	100.0	99.8	0.486
	Alert PDE 3A 2a - PDE Frag 3	6	32	11592	0	15.8	100.0	99.7	0.397
	Alert PDE 3A 2b - PDE Frag 3-like SMARTS	7	31	11592	0	18.4	100.0	99.7	0.429
	Alert PDE 3A 3 - PDE 3A Frag 4	14	24	11574	18	36.8	99.8	99.6	0.400
	Alert PDE 3A 4 - Veratrol-like	13	25	11139	453	34.2	96.1	95.9	0.088

Receptor	MIE	TP	FN	TN	FP	SE	SP	Q	MCC
4D	Alert PDE 4D 1 - Homoveratrol-like	47	71	11311	201	39.8	98.3	97.7	0.264
	Alert PDE 4D 2 - 8-Isopropyl-1H-purine Frag	19	99	11512	0	16.1	100.0	99.1	0.400
	Alert PDE 4D 3 - PDE 4D Frag 3	50	68	10932	580	42.4	95.0	94.4	0.165
Thymidylate Synthase									
	Alert TS 1 - N-Acetyl-DL-glutamic acid	13	41	11554	22	24.1	99.8	99.5	0.296
	Alert TS 2 - Pyrimidine	2	52	11423	153	3.7	98.7	98.2	0.014
	Alert TS 1 AND 2	1	53	11576	0	1.9	100.0	99.5	0.136
	Alert TS 3 - Uracil	3	51	11537	39	5.6	99.7	99.2	0.059
Tyrosine Protein Kinase									
	Alert LCK 1 - Pyrimidine-like	78	71	10494	987	52.3	91.4	90.9	0.171
Ion Channels									
Glutamate (NMDA) Receptor									
	Alerts GRSZ1 1 - GRSZ1 Frag 1	7	0	11623	0	100.0	100.0	100.0	1.000
Potassium Voltage Gated Channel Subunit H Member 2									
	Alert HERG 1a - Flex Aromatic Amine 1-5	60	20	5493	6057	75.0	47.6	47.7	0.037
	Alert HERG 1b - Flex Aromatic Amine 2-5	53	27	6046	5504	66.3	52.3	52.4	0.031
Serotonin 3A Receptor									
	Alert S3AR 1 - S3AR Frag 1	12	74	11544	0	14.0	100.0	99.4	0.372
	Alert S3AR 2 - 1-[2-(Phenylsulfanyl)phenyl]piperazine	7	79	11533	11	8.1	99.9	99.2	0.175
	Alert S3AR 1 or 2	19	67	11533	11	22.1	99.9	99.3	0.372

Receptor	MIE	TP	FN	TN	FP	SE	SP	Q	MCC
Sodium Channel V Subunit alpha									
	Alert SCA-A 1a - Pyrimidine-like aromatic	24	18	10466	1122	57.1	90.3	90.2	0.096
	Alert SCA-A 1b - Pyrimidine-like wildcard	24	18	10466	1122	57.1	90.3	90.2	0.096
	Alert SCA-A 2 - 4-(4-Piperidineoxy)pyrimidine	9	33	11588	0	21.4	100.0	99.7	0.462
	Alert SCA-A 3 - SCV Frag 1	8	34	11588	0	19.0	100.0	99.7	0.436
	Alert SCA-A 4 - N-(2-Phenylethyl)acetamide	11	31	11133	455	26.2	96.1	95.8	0.068
Voltage Gated K Channel Subunit Kv7.1									
	Alert VGKC Kv7.1 1 - VGKV Kv7.1 Frag 1	7	0	11618	5	100.0	100.0	100.0	0.764
Nuclear Receptors									
Androgen Receptor									
	Alert AR 1a - Benzonitrile	132	205	11075	218	39.2	98.1	96.4	0.366
	Alert AR 1b - Nitrobenzene	39	298	11191	102	11.6	99.1	96.6	0.164
	Alert AR 1c - Quinolone	59	278	11247	46	17.5	99.6	97.2	0.303
	Alert AR 1a, 1b OR 1c	230	107	10927	366	68.2	96.8	95.9	0.495
Glucocorticoid Receptor									
	Alert GR 1a - GR Frag 1	74	378	11159	19	16.4	99.8	96.6	0.352
	Alert GR 1b - tert-Butylcyclohexane	55	397	11145	33	12.2	99.7	96.3	0.265
	Alert GR 1c - GR Frag 4	68	384	11168	10	15.0	99.9	96.6	0.354
	Alert GR 1d - GR Frag 5	68	384	11023	155	15.0	98.6	95.4	0.192
	Alerts GR 1a OR 1c OR 1d	76	376	11005	173	16.8	98.5	95.3	0.204

Receptor	MIE	TP	FN	TN	FP	SE	SP	Q	MCC
Transporters									
Dopamine Transporter									
	Alert DT 1 - 3-Phenyl-8-azabicyclo[3.2.1]octane	63	372	11136	59	14.5	99.5	96.3	0.260
	Alert DT 2 - Diphenylmethane-like	101	334	9902	1293	23.2	88.5	86.0	0.068
Norepinephrine Transporter									
	Alert NT 1a - Cocaine-like 1	15	623	10961	31	2.4	99.7	94.4	0.075
	Alert NT 1b - Cocaine-like 2	16	622	10960	32	2.5	99.7	94.4	0.079
	Alert NT 1c - Cocaine-like 3	33	605	10903	89	5.2	99.2	94.0	0.098
	Alert NT 1d - Cocaine-like 4	37	601	10789	203	5.8	98.2	93.1	0.063
	Alert NT 2 - Amphetamines-like	486	152	4158	6834	76.2	37.8	39.9	0.066
Serotonin Transporter									
	Alert ST 1 - 3-Ethyl-indole	78	749	10238	565	9.4	94.8	88.7	0.047
	Alert ST 2 - DMEA	68	759	4256	6547	8.2	39.4	37.2	-0.272
	Alert ST 3 - Benzyloxybenzene	81	746	10635	168	9.8	98.4	92.1	0.146
	Alert ST 4 - Diphenylmethane	227	600	10258	545	27.4	95.0	90.2	0.231

Table 3. Results for test sets of each receptor with the remaining test sets acting as negatives. TP=True Positive, FN=False Negative, TN=True Negative, FP=False Positive, SE=Sensitivity, SP=Specificity, Q=Overall Quality, MCC=Matthews Correlation Coefficient.

Target		TP	FN	SE
GPCRs				
Adenosine A2a Receptor	Alert 1, 2 3 OR 4	505	305	62.35
Alpha-1a Adrenergic Receptor	Alert 1, 2 OR 3	57	138	29.23
Alpha-2a Adrenergic Receptor	Alert 1, 2, 3, 4, 5 OR 6	43	67	39.09
Beta-1 Adrenergic Receptor	Alert 1, 2 OR 3 (beta alerts)	182	7	96.30
Beta-2 Adrenergic Receptor	Alert 1, 2 OR 3 (beta alerts)	215	15	93.48
Cannabinoid CB1 Receptor	Alert 1, 2 OR 3	179	871	17.05
Cannabinoid CB2 Receptor	Alert 1 OR 2	207	763	21.34
Cholecystokinin Receptor A	Alert 1, 2 OR 3	76	2	97.44
Dopamine D1 Receptor	Alert 1, 2, 3 OR 4	45	86	34.35
Dopamine D2 Receptor	Alert 1, 2 OR 3	417	341	55.01
Endothelin Receptor A	Alert 1, 2, 3, 4 OR 5	25	3	89.29
Histamine H1 Receptor	Alert 1 OR 2	45	113	28.48
Histamine H2 Receptor	Alert 1, 2 OR 4	38	10	79.17
Muscarinic Acetylcholine Receptor M1	Alert 1 OR 4	109	139	43.95
Muscarinic Acetylcholine Receptor M2	Alert 1, 2 OR 4	74	90	45.12
Muscarinic Acetylcholine Receptor M3	Alert 2, 3 OR 4	128	167	43.39
Delta Opioid Receptor	Alert 1 (3 or more) OR 2	268	241	52.65
Kappa Opioid Receptor	Alert 1 (3 or more) OR 2	286	271	51.35
Mu Opioid Receptor	Alert 1 (3 or more) OR 2	353	306	53.57
Serotonin 1A Receptor	Alert SR1, SR1a1b2a1 OR SR1a1	313	191	62.10
Serotonin 1B Receptor	Alert SR1, SR1a1b2a1 OR SR1b1	98	16	85.96
Serotonin 2A Receptor	Alert SR1 OR SR1a1b2a1	140	310	31.11
Serotonin 2B Receptor	Alert SR1, SR2b1 OR SR2b2	37	158	18.97
Vasopressin V1A Receptor	Alert 1, 2, 3 OR 4	133	47	73.89

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Target		TP	FN	SE
Enzymes				
Acetylcholinesterase	Alert 1a, 1b, 1c, 1d, OR 1e	244	122	66.67
Cyclooxygenase 1	Alert 1, 2 OR 3	45	60	42.86
Cyclooxygenase 2	Alert 1, 2 OR 3	197	81	70.86
Dihydrofolate Reductase	Alert 1 (SMARTS)	65	35	65.00
GAR Transformylase	Alert 1	10	3	76.92
Histone Deacetylase 1	Alert 1, 2b, 2d OR 2f	244	76	76.25
Monoamine Oxidase A	Alert 1, 2a, 2b, 2c, 3a, 3b OR 3c	77	74	50.99
Phosphodiesterase 3A	Alert 1, 2b, 3 OR 4	27	11	71.05
Phosphodiesterase 4D	Alert 1, 2 OR 3	78	40	66.10
Thymidylate Synthase	Alert 1, 2 OR 3	17	37	31.48
Tyrosine-Protein Kinase	Alert 1	78	71	52.35
Ion Channels				
Glutamate (NMDA) Receptor	Alert 1	7	0	100.00
Potassium Voltage Gated Channel KQT 1	Alert 1a	60	20	75.00
Serotonin 3A Receptor	Alert 1 OR 2	19	67	22.09
Sodium Channel V Subunit Alpha	Alert 1b, 2, 3 OR 4	28	14	66.67
Voltage Gated K Channel Subunit Kv7.1	Alert 1	7	0	100.00

Target		TP	FN	SE
Nuclear Receptors				
Androgen Receptor	Alert 1a, 1b OR 1c	230	107	68.25
Glucocorticoid Receptor	Alert 1a, 1b, 1c OR 1d	76	376	16.81
Transporters				
Dopamine Transporter	Alert 1 OR 2	164	271	37.70
Norepinephrine Transporter	Alert 1a, 1b, 1c, 1d OR 2	490	148	76.80
Serotonin Transporter	Alert 1, 2, 3 OR 4	637	190	77.03

Table 4. Results for models of combined structural alerts for each receptor. TP=True Positive, FN=False Negative, SE=Sensitivity.