Atomic contribution mapping and exploration with reverse fingerprinting (ACME-RF): Assigning toxicological endpoints to chemical structure at atomic resolution

Michael R Goldsmith (1), Daniel T Chang (1), Chris Williams (2)

1) U.S. Environmental Protection Agency, Office of Research and Development, Durham, NC 27713 (2) Chemical Computing Group Inc., 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7

Rocky Goldsmith I goldsmith.michael@epa.gov I 919-541-7770

Abstract

Purpose

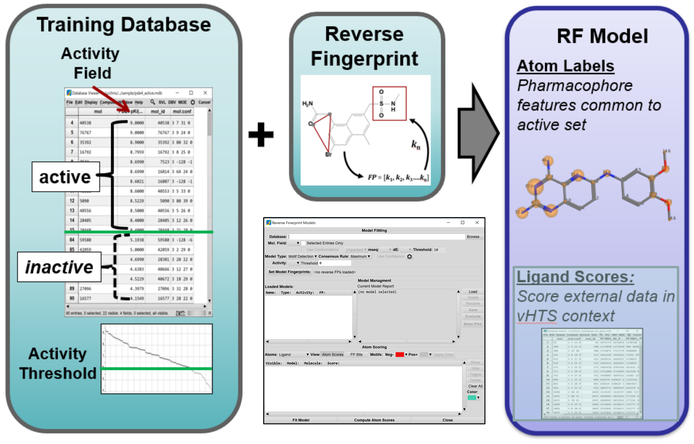
In silico tools and models for assessing activity are usually defined by endpoints and quantitative structural metrics. Although it is useful to obtain categorical/continuous estimates of activity, traditional SAR provide limited guidance as to the molecular moieties giving rise to the endpoint. Reverse-fingerprinting (RF), provides a useful marriage between discretized endpoints and feature-based molecular fingerprint. RF produces both a quantitative and visual representation of atomic contribution to an endpoint, mapped on to structure (Williams C, 2009 PMID: 19442069). Here we introduce the concept of atomic contribution mapping and exploration (ACME) using the RF framework, as implemented in the Molecular Operating Environment (MOE)..

Reverse Fingerprint Modeling Workflow

Method

Using public datasets, we explore three different ACME-RF examples. First, we demonstrate the rapid identification of a class of pyrethroid acaricide that is not-toxic to honeybees while still being toxic to the varroa mite using very basic insecticide-class information of 80 pyrethoids as inputs. Second, we used the ToxCast NVS_NR_hER dataset (165/2645) to build a RF model that was used to identify the toxicophore of hER-a that directly map to known crystal structures. Finally, we explore photostability half-lives (Blum, Kristin M. 2013) and identify critical photolabile moieties.

The generalized scheme involve in RF modeling workflow consists of identifying an active / inactive dataset, computing a fingerprint metric of choice, calculating the mutual information between a bit state K and the active state as a function of its probability in an active structure set relative to all fingerprint bits in an inactive structure set. This metric is subsequently normalized and expressed as contributing positively or negatively to the endpoint being mapped.

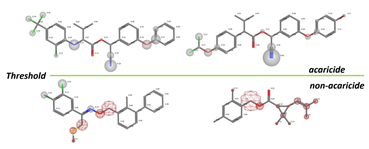


The generalized workflow consists of a training database, a reverse fingprinting algorithm, a panel to control fingperint types and comparison metrics (average versus maximum similarity) and a scored dataset along with the atomic contributyion map for exploration (ACME) and riverse fingerprint contribution.

Identifying Bee-Friendlier Chemistries

Acaricide vs Insecticide

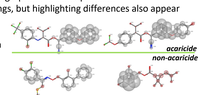
- **Arachnid Selective Pesticide**
  - Identify features for arachnid selectivity (kill mites not bees)
    - Set Activity as 'acaricide> 0' 14 active, 66 inactive **MACCS** fingerprint



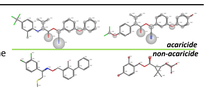
  - Cyano group important
  - Corresponds to Typell pyrethroid – acaricides (ticks and mites)

Acaricide vs Insecticide

- **Arachnid Selective Pesticide**
  - Apply FCFP and 2D_PCH fingerprints to same problem
  - FPs highlight common things, but highlighting differences also appear
- **FCFP6**
  - Heavily highlights Ph-O-Ph moiety
  - Highlights –CN slightly
  - Down scores cyclopropyl



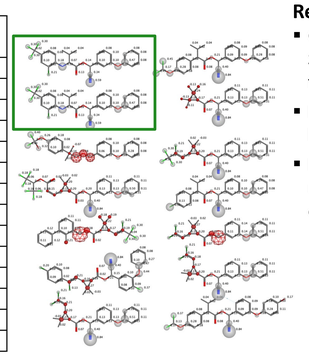
- **2D_PCH**
  - Heavily highlights –CN and C=O moieties
  - Slightly down scores alkene group



Relative Selectivity of Acaricides in Database

- **Goal:**
  - Identify compounds most selective for the varroa mite (which feeds on the honeybee) but be a poor insecticide – not kill the bee.
  - Use MACCS RF model to compute all molecule scores (Sm) and atomic scores for all 14 acaricides.

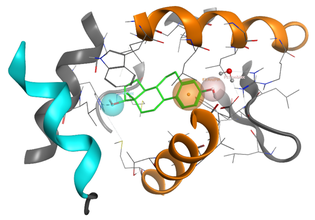
Molecule:	Score (Sm):
Fluvalinate	5.9
tau-fluvalinate	5.9
Flucythrinate	5.6
Brofluthrinate	5.6
Fenvalerate	5.0
Acrinathrin	4.4
Cyhalothrin	4.3
Flumethrin	3.8
α-cypermethrin	3.6
Cypermethrin	3.6
Fenpropathrin	3.0
Halfenprox	1.9
Permethrin	0.8
bifenthrin	-0.5



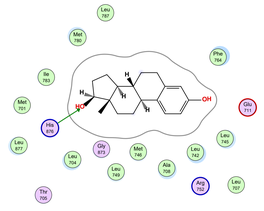
Result

- Compounds with highest S_M are tau-fluvalinate and fluvalinate (in green box)
- High S_M scoring compounds have few red spheres
- Fluvalinate is compound used by beekeepers to control mites

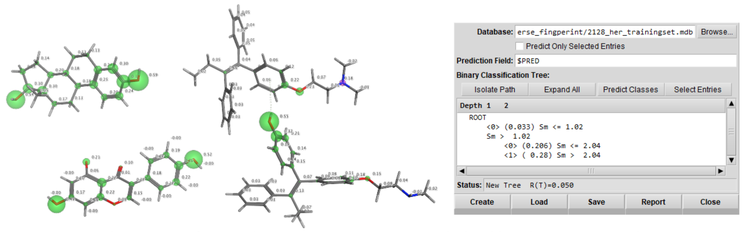
Finding the Estrogen Receptor Pharmacophore



Crystal structure of hER alpha (PDB 6CBZ) With pharmacophore elucidated using autoPH4



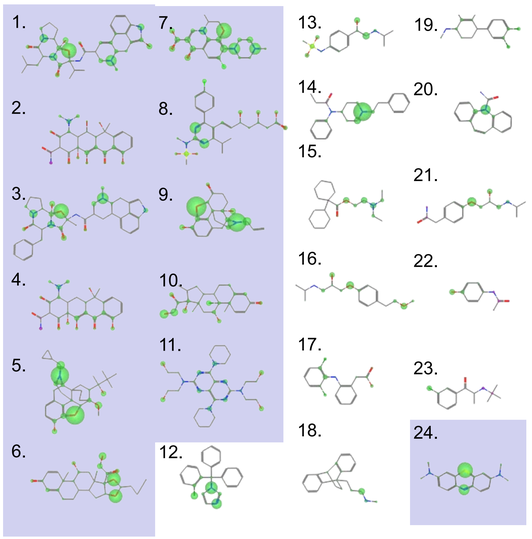
Estradiol binding site, 2D representation from PDB 6CBZ



Atomic contribution maps of activity (green = +ve, magenta = -ve) calculated based on training set of 2128 chemistries (126 actives), an 80% subset of the full NVS_NR_hER dataset. (A) estradiol showing the classic motif (S = 4.9) (B) Genistein showing similar activity motif (S = 3.9), and then (C) Tamoxifen (S=1.4) versus its active metabolite (D) endoxifen (S= 3.0). Test set of 503 chemistries (33 actives) provided a total of 21 chemistries with S > 2.04 of which 13 were active (nearly a 10 X enrichment from 6.5% active to 62% active). The model score can differentiate active from inactive with a misclassification rate of 5% (i.e. 95% accuracy).

Anticipating Photolabile Moieties

1	Bromocriptine
2	Oxytetracycline
3	Dihydroergotamine
4	Tetracycline
5	Buprenorphine
6	Budesonide
7	Ofloxacin
8	Rosuvastatin
9	Naloxone
10	Beclometasone
11	Dipyridamole
12	Clotrimazole
13	Sotalol
14	Fentanyl
15	Dicycloverine
16	Metoprolol
17	Diclofenac
18	Maprotiline
19	Sertraline
20	Carbamazepine
21	Atenolol
22	Paracetamol
23	Bupropion
24	Methylene Blue



- Light sensitive moieties are highlighted in green spheres (violet background are photo-labile < 24 hours half-life in filtered river water (Blum K, 2013)
- Photostable moieties are highlighted in magenta.
- Decision tree classifier indicates key groups (see below)
- Testing this approach on methylene blue provides well known liabilities

BINARY CLASSIFICATION TREE:
Misclassification rate = 0.117

```
ROOT
  MACCS(146) <= 0.5
    MACCS(-85) <= 1.5
      <0> (0.064) MACCS(-36) == 0
      <1> ( 0.25) MACCS(-36) <> 0
    MACCS(-85) >  1.5
      MACCS(154) <= 0.5
        <1> (    0) MACCS(-72) <= 1
        <0> (    0) MACCS(-72) >  1
      <0> (  0.2) MACCS(154) >  0.5
  MACCS(146) >  0.5
    MACCS(125) == 0
      <1> (0.286) MACCS(117) <= 0.5
      <0> (    0) MACCS(117) >  0.5
    <1> (0.179) MACCS(125) <> 0
```

LEGEND

MACCS(146)	Key(164)-2 if key(164)>2; else 0
MACCS(-36)	#S atoms in rings
MACCS(-72)	#O separated by 3 bonds
MACCS(-85)	#N bonded to >= 3 C
MACCS(117)	#N 2 bonds from an O
MACCS(125)	Is # AROMATIC RING > 1?
MACCS(146)	Key(164)-2 if key(164)>2; else 0
MACCS(154)	#O in C=O
MACCS(164)	#oxygens

Conclusion and Future DIrection

Results/Conclusion

Using ACME-RF we identified and visualized moieties of molecules that resulted in (I) apical endpoints across species (II) chemical-biological interactions and (III) photodegradation liabilities. The method can be used to identify toxic chemicals and critical toxicophore fragments or sub-structures essential for molecular discovery and de-risking. Future efforts will include adapting the ToxPrint fingerprints into the same framework. [This abstract solely represents the views of the authors and not the view of the Agency.]

Williams C, Schreyer SK. Reverse fingerprinting and mutual information-based activity labeling and scoring (MIBALS). Comb Chem High Throughput Screen. 2009 May;12(4):424-39. doi: 10.2174/138620709788167953. PMID: 19442069.

Blum, Kristin. "Phototransformation of pharmaceuticals in the environment: Multivariate modeling and experimental determination of photolysis half-lives." (2013).

U.S. Environmental Protection Agency
Office of Research and Development