Identification of Chemical Features Linked to Thyroperoxidase Inhibition

Steven O. Simmons[1], Eric D. Watt[1,2], Katie Paul Friedman[4], Keith A. Houck[1], Richard S. Judson[1], Michael W. Hornung[3], Kevin M. Crofton[1]

[1]National Center for Computational Toxicology; [2]Oak Ridge Institute for Science Education Postdoctoral Fellow;
[3]Mid-Continent Ecology Division, NHEERL; ORD, U.S. EPA, RTP, NC and Duluth, MN and; [4]Bayer CropScience LP, Research Triangle Park, NC

simmons.steve@epa.gov | 919-541-1475 | orcid.org/0000-0001-9079-1069

Abstract # 1886

Introduction

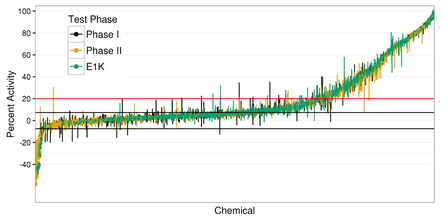
- The US Environmental Protection Agency (EPA) developed the Endocrine Disrupter Screening Program (EDSP) to identify environmentally-relevant chemicals that interfere with endocrine pathways, including thyroid hormone signaling
- A key target for chemically-induced thyroid disruption is inhibition of thyroperoxidase (TPO), which catalyzes thyroid hormone synthesis
- The AUR-TPO assay was recently developed and used to screen nearly 2,000 ToxCast chemicals for potential TPO inhibition activity and selectivity

Objective

Use ToxCast TPO inhibition assay data to identify chemical structures correlated with *in vitro* TPO inhibition

ToxCast Screen for TPO Inhibitors

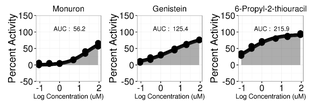
Figure 1: Median % TPO inhibition by Sample and ToxCast Phase Single-concentration Screen

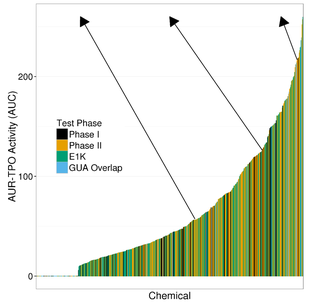


- 489 ToxCast chemicals were active in initial single-concentration screening
- These 489 AUR-TPO actives were tested in concentration-response (0.1-100 µM final concentration) (N=3)
- Multi- and single-concentration data were analyzed using the ToxCast Analysis Pipeline (tcpl v2.0)
- Chemicals that failed to inhibit TPO activity by ≥ 20% were defined as 'inactive' and AUC = 0
- Area under the fitted curve (AUC) scores incorporate both potency and efficacy for ranking
- High AUC example: 6-propyl-2-thiouracil (highly potent and efficacious)
- Low AUC example: monuron (less potent and efficacious)

- 1,903 ToxCast chemicals were tested in the AUR-TPO assay at 87.5 µM final concentration, solubility permitting (N=3) Chemicals eliciting a median TPO inhibition ≥ 20% (above red line) were defined as 'active' and tested in 6-8 pt concentration-response (red line, Fig. 1) Black lines represent ± 3*baseline median absolute deviation (3*bmad)

Figure 2: AUC Ranking and ToxCast Phase Multiple-Concentration Screen





Enriched Feature Identification Using Hypergeometric Distribution

Figure 3: Distribution of 429 Represented Chemotypes (p = 0.005 threshold for over- and under-enrichment)

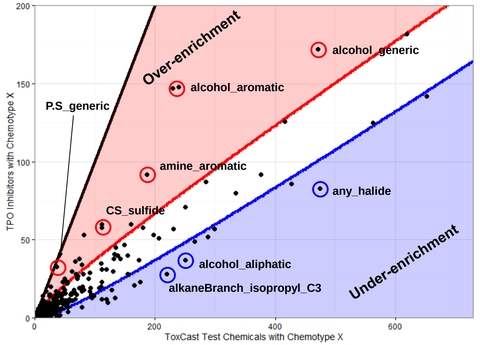
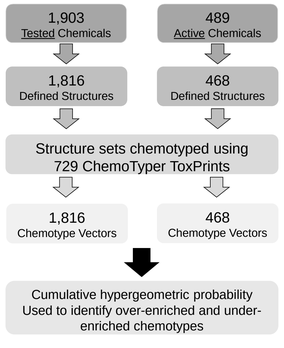


Figure 4: Workflow for identifying enriched chemotypes



- 1,816 of 1,903 tested chemicals have defined structures; others were largely mixtures
- 429 of 729 ToxPrint chemotypes were represented among the 1,816 tested structures
- Cumulative hypergeometric probability: what is the probability that there will be ≥ X active chemicals with chemotype Y if there are 1,816 tested chemicals, 486 actives, and Z tested chemicals with chemotype Y?

25 Most Over-enriched Chemotypes

Chemotype	Tested	Active	p-value
bond.COH_alcohol_aromatic_phenol	230	147	1.70E-40
bond.COH_alcohol_aromatic	239	148	2.70E-38
ring.aromatic_benzene	1012	341	3.61E-19
bond.P.S_generic	37	33	1.31E-17
bond.CN_amine_pri.NH2_aromatic	82	53	6.07E-15
bond.CN_amine_aromatic_generic	186	92	4.98E-14
bond.CS_sulfide	111	60	1.27E-11
bond.O.O_O.S._sulfhydride	19	17	2.85E-10
bond.CN_amine_pri.NH2_generic	112	58	2.94E-10
bond.COH_alcohol_generic	471	172	4.70E-10
bond.C.S_carbonyl_thio_generic	27	20	1.78E-08
chain.aromaticAlkane_Ph.C1.Ph	68	36	2.74E-07
chain.aromaticAlkane_Ar.C.Ar	74	38	3.95E-07
bond.P.O_phosphate_dithio	13	11	7.69E-07
bond.C.O_carbonyl_1_2.di	14	11	4.13E-06
bond.C..O.N_carbamate_dithio	9	8	4.73E-06
bond.NC.O_urea_thio	15	11	1.58E-05
chain.aromaticAlkane_Ph.C6	8	7	1.86E-05
bond.CC..O.C_ketone_alkene_cyclic_2.en.1.one_generic	62	30	2.68E-05
ring.hetero_.6..6._O_benzopyrone_.1_4..	12	9	4.70E-05
chain.aromaticAlkane_Ph.C4	14	10	5.03E-05
ring.fused_.6..6._naphthalene	50	25	5.18E-05
bond.N.N_azo_aromatic	25	15	5.67E-05
chain.alkeneLinear_diene_1_3.butene	7	6	7.30E-05
bond.CN_amine_ter.N_aromatic_aliphatic	46	23	9.48E-05

Alcohols (aromatic)
Amines (aromatic)
Thiocarbonyls (C=S)
Phosphothioates (P-S)
Aromatic (ring/chain)

25 Most Under-enriched Chemotypes

Chemotype	Tested	Active	p-value
bond.C..O.O_carboxylicEster_acyclic	122	10	3.74E-07
chain.alkaneBranch_isopropyl_C3	220	28	4.26E-07
bond.X.any._halide	474	83	6.22E-07
bond.COC_ether_aliphatic	134	13	1.33E-06
bond.COH_alcohol_aliphatic_generic	250	37	5.63E-06
bond.COH_alcohol_pri.alkyl	119	12	9.76E-06
bond.C..O.O_carboxylicEster_alkyl	166	21	1.26E-05
bond.C.O_carbonyl_generic	913	196	1.53E-05
bond.CX_halide_alkyl.X_generic	175	23	1.58E-05
bond.C..O.O_carboxylicEster_aliphatic	111	12	5.09E-05
bond.CX_halide_alkyl.F_trifluoro_.1_1_1..	78	7	1.36E-04
chain.oxy.alkaneLinear_ethyleneOxide_EO1	37	1	2.03E-04
bond.CX_halide_alkyl.X_benzyl_alkane	42	2	3.91E-04
bond.CC..O.C_ketone_aliphatic_acyclic	55	4	4.09E-04
chain.alkaneLinear_butyl_C4	288	52	5.20E-04
bond.CX_halide_alkyl.X_dihalo_.1_1..	128	18	6.98E-04
chain.alkaneLinear_propyl_C3	427	86	1.23E-03
chain.alkaneCyclic_ethyl_C2_connect_noZ.	266	49	1.50E-03
bond.CX_halide_aromatic.X_generic	299	57	1.90E-03
chain.alkaneLinear_ethyl_C2.H_gt_1.	652	142	2.01E-03
bond.CX_halide_alkyl.X_trihalo_.1_1_1..	95	13	2.66E-03
chain.oxy.alkaneLinear_ethylenOxide_EO1.O.	27	1	3.16E-03
bond.C..O.N_carboxamide_.NR2.	98	14	3.70E-03
bond.CX_halide_alkyl.X_secondary	18	0	4.54E-03
bond.CC..O.C_ketone_methyl_aliphatic	17	0	6.14E-03

Halides
Carboxilic Esters
Aliphatic Alcohols
Alkanes
Aliphatic Ketones

Predictive Feature Identification Using a Random Forest Model

Figure 5: Workflow for Identifying the Most Predictive Chemotypes

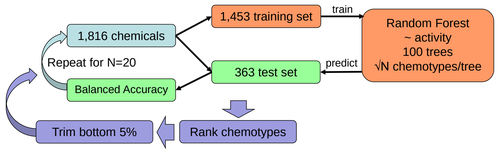
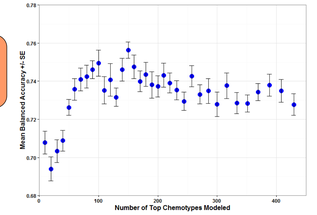


Figure 6: Model Performance by Top Chemotypes



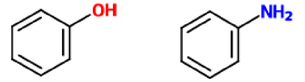
25 Most Predictive Chemotypes

Chemotype	Mean ↓ Gini Index
bond.COH_alcohol_aromatic_phenol	21.71
bond.COH_alcohol_aromatic	19.51
bond.P.S_generic	12.19
ring.aromatic_benzene	9.57
bond.CN_amine_aromatic_generic	9.13
bond.CS_sulfide	8.74
bond.CN_amine_pri.NH2_aromatic	8.23
bond.COH_alcohol_generic	6.13
bond.CN_amine_pri.NH2_generic	5.95
bond.OQ.O.O_S._sulfhydride	5.87
bond.X.any._halide	5.30
bond.C.O_carbonyl_generic	5.11
bond.C.S_carbonyl_thio_generic	5.05
chain.alkaneLinear_ethyl_C2._connect_noZ_CN.4.	5.05
chain.alkaneLinear_ethyl_C2.H_gt_1.	4.98
ring.aromatic_phenyl	4.59
chain.aromaticAlkane_Ph.C1_acyclic_generic	4.59
bond.P.O_phosphate_thioate	4.37
bond.CX_halide_aromatic.X_generic	4.24
bond.COC_ether_aliphatic_.aromatic	4.22
chain.aromaticAlkane_Ph.C1_cyclic	4.14
bond.OZ_oxide_hyroxy	4.02
bond.C.O_carbonyl_1_2.di	3.98
chain.aromaticAlkane_Ph.C1_acyclic_connect_noDblBd	3.98
bond.N..O._nitro_C	3.76

Over-enriched
Under-enriched

- Chemotypes were trimmed from 429 to 10 using a looping random forest (R randomForest 4.6-12) that evaluated balanced accuracy and ranked the chemotypes most predictive of TPO inhibition activity (Fig. 5)
- Fig. 6 shows optimal model performance peaks at a trimmed chemotype set of 150
- The 25 most predictive chemotypes were identified at the model optima (top 150) using the mean decrease in Gini index, a measure of node impurity in tree structures
- 6 of the 25 most predictive chemotypes were not identified as over- or under-enriched by probability analysis
- Using chemotypes alone, a predictive TPO model can be built with a balanced accuracy of 76%

Conclusions and Future Directions



Aromatic Alcohols | Aromatic Amines | Thiocarbonyls | Phosphorothioates

- Four general chemotypes predict *in vitro* TPO inhibition: aromatic alcohols, aromatic amines, thiocarbonyls and phosphorothioates
- Both probabilistic and predictive models readily identify chemotypes that impact TPO activity; these methods are complementary and taken together increase our understanding of what structural features impact activity
- Some chemotypes identified as enriched using probability analysis were not among the most predictive chemotypes; conversely, some of the most predictive chemotypes were not among the most statistically enriched
- Both over-and under-enriched chemotypes help improve predictive modelling
- Supplementing a predictive model with additional data, such as physical-chemical properties are expected to improve model performance over using chemotypes alone