Identification of candidate reference chemicals using multidimensional literature and database mining with EPA's PubMed Abstract Sifter

Nancy C. Baker[1], Thomas Knudsen[2]

[1] Leidos, RTP, NC; [2] CCTE, EPA, RTP, NC

Nancy C. Baker I baker.nancy@epa.gov I 0000-0002-8351-9435

www.epa.gov

Introduction

Reference chemicals are essential to calibrate *in vitro* assays and establish confidence in their results for toxicity prediction. The retinoid pathway is critical in embryo development. Chemical perturbation of this pathway has adverse developmental consequences, motivating *in vitro* profiling of bioactivity for targets in the retinoid pathway for predictive toxicity. Using public data sources, we assembled a set of candidate reference chemicals for ten targets of the retinoid pathway[1]:

- Retinol binding protein (serum) – RBP
- Stimulated by retinoic acid 6 (STRA6)
- Cellular retinol binding protein (CRBP)
- Cellular retinoic acid binding protein (CRABP)
- Retinoic acid 4-hydroxylase (CYP26)
- Retinol dehydrogenase (RDH)
- Retinal dehydrogenase (RALDH)
- Retinoic acid receptors (RARA, RARB, RARG)

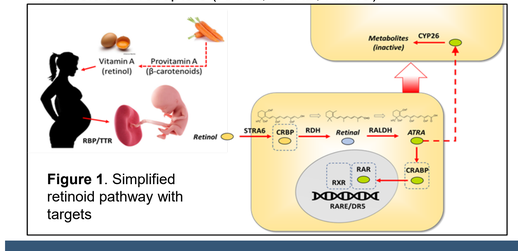


Figure 1. Simplified retinoid pathway with targets

Methods

Structured databases of *in vitro* testing results were searched for candidate chemicals: ToxCast / Tox21; ChEMBL; Protein Data Bank (PDB); MeSH chemical terms and chemicals from publications of *in vitro* test results. MeSH chemicals and pdf extraction were performed with Abstract Sifter[2] v7.3. To establish confidence, we assessed literature connections to adverse outcomes associated with retinoid pathway disruption.

Candidate chemical compilation

Hits in ToxCast RARa assays. The results were filtered to chemicals with AC50 under 2 micromola
Curation steps were used to remove chemical hits with caution flags.

Assay	DSSToxID	Chemical Name	AC50 (uM)
ATG_RARa_trans_up	DTXSID3023556	Retinol	0.0763
	DTXSID2022880	Danazol	0.7338
	DTXSID9020453	Dieldrin	0.7704
	DTXSID8024151	Imazalil	0.9084
	DTXSID8041379	2,6-Di-tert-butyl-4-methoxyphenol	1.0705
	DTXSID9037539	Endosulfan I	1.3836
	DTXSID2032500	Triflumizole	1.4526
	DTXSID4038922	Tetrabromobisphenol A bis(2-hydroxyethyl) ether	1.5032
	DTXSID7047279	CP-532623	1.5610
NVS_NR_hRAR_Antagonist	DTXSID4021559	2,6-Di-tert-butyl-4-nitrophenol	1.8870
	DTXSID2026523	Symclosene	0.2638
	DTXSID8041909	Dysprosium(III) chloride	0.8540
NVS_NR_hRARa_Agonist	DTXSID7021239	all-trans-Retinoic acid	0.1258
	DTXSID9022310	Daidzein	0.9369
	DTXSID0023901	Bentazone	1.8986
	DTXSID8037706	Potassium perfluorooctanesulfonate	1.9118

ToxCast and Tox21 data were culled from the EPA Chemicals Dashboard (comptox.epa.gov) for RARA, RARB, RARG, DR5 and Tox21_RAR. A stringent cutoff of 2.0 micromolar AC50 was applied in order to retrieve only the most potent hits.

Figure 2. Sample data from ToxCast.

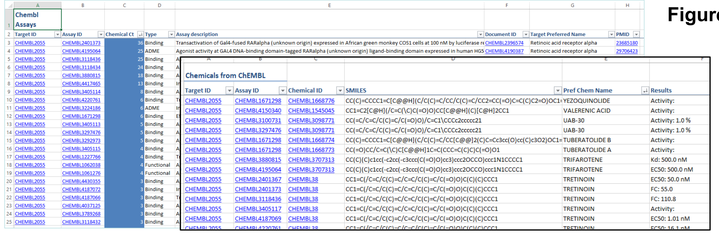


Figure 3. Sample data from ChEMBL.

ChEMBL was searched for each of the targets and records were downloaded to Excel for evaluation and scanning for chemical entities with identified names. For candidate chemicals, the publication was retrieved and reviewed.

Figure 4. Sample data from Abstract Sifter MeSHMine functionality.

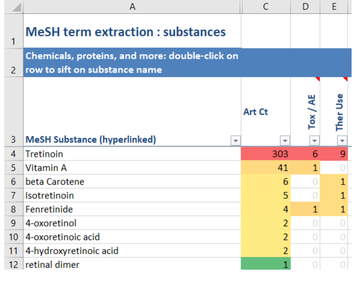


Figure 5. Sample data from Abstract Sifter pdf table extractor and EPA Chemicals Dashboard APIs.

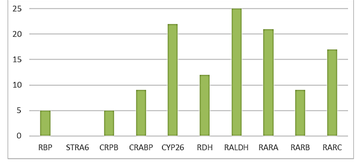
DTXSID	Inhibitor	CYP26A1 IC50 (mM)	CYP26B1 IC50 (mM)	CYP2C8 IC50 (mM)	Substrate	Reference
DTXSID4022652	Benzbromarone	0.63±0.06	7.57±4.93	0.38 (N.R.)	Montelukast (	44
DTXSID0022725	Candesartan	25.6±25.8	58.3±180	36.2±1.7	Amodiaquine	51
DTXSID5020239	Candesartan cileket	0.41±0.32	0.27±0.06	0.496±0.190	Amodiaquine	51
DTXSID7029871	Clotrimazole	0.02±0.01	0.05±0.01	0.725±0.116	Amodiaquine	51
DTXSID0020573	17beta-Estradiol	2.24±1.33	6.73±3.77	6.54±1.22	Amodiaquine	51
DTXSID3020627	Fluconazole	0.70±0.19	19.8±3.21	48.9 (N.R.)	Amodiaquine	50
DTXSID3023180	Itraconazole	0.55±0.09	0.16±0.02	2.16±0.41	Paclitaxel	Unpublished Data
DTXSID10872412	Mometasone	0.90±0.08	6.81±1.19	0.813±0.112	Amodiaquine	51
DTXSID9023334	Montelukast	0.12±0.02	0.61±0.09	0.009±0.001	Amodiaquine	51
DTXSID3037129	Pioglitazone	0.93±0.26	8.48±1.19	11.7±4.0	Amodiaquine	51
DTXSID4021218	Quercetin	1.92±0.39	76.2±148	3.94±0.64	Amodiaquine	51
DTXSID3023550	Raloxifene	1.78±0.77	3.28±1.15	2.15±0.90	Amodiaquine	51
DTXSID3023552	Repaglinide	7.73±2.69	0.61±0.25	11.1 (N.R.)	Montelukast (	44
DTXSID10486627	Ritonavir	3.84±2.82	2.56±0.25	3.03±1.14	Amodiaquine	51
DTXSID7037131	Rosiglitazone	11.9±1.02	8.47±6.64	10.8±3.1	Amodiaquine	51
DTXSID10341187	Tamoxifen	21.4±20.3	14.0±1.11	3.34±1.55	Amodiaquine	51
DTXSID5023746	Zafirlukast	0.06±0.02	0.71±0.23	0.644±0.273	Amodiaquine	51

Results summary

The process produced approximately 220 putative reference chemicals. The results differed among the targets.

- For STRA6, no putative reference chemicals were identified.
- RALDH has many candidates, but this enzyme family has many members that share substrates.
- Targets of interest to drug discovery have the most in vitro testing activity, but fewer environmental chemicals are represented.

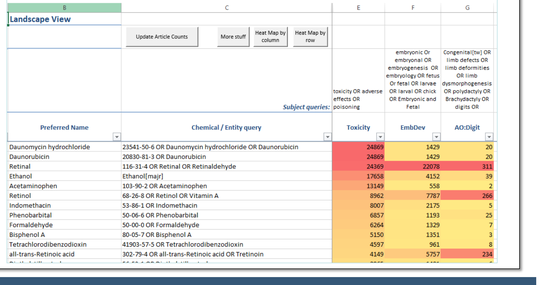
Figure 6. Candidate reference chemical counts by target



PubMed searches with the Abstract Sifter tool were performed and yielded 137 (~ 75%) chemicals showed connections to developmental toxicity.

As an example, 55 chemicals had literature on limb defects.

Figure 7. Abstract Sifter article counts for toxicity, developmental toxicity, and limb defects



Next steps

CYP26 candidate reference chemicals

40 compounds for purchase and evaluation

Tox21 Cross-Partner Project 13 is developing a high-throughput assay for CYP26 bioactivity. The candidate chemical collection was reviewed to select 40 chemicals for a proof-of-concept study (manuscript in preparation). Considerations for selections to the shortlist are:

- Strong supportive data
- Environmental interest
- Known developmental toxicant or known negative
- Chemical diversity

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

This poster does not necessarily represent U.S. EPA policy.

References:

1) Baker N et al. ALTEX. 2022 Jun 23. doi: 10.14573/altex.2202231.2
2) Baker N et al. F1000 Res. 2017 Dec 21;6. pii: Chem Inf Sci-2164.