

www.epa.gov

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

Nancy C. Baker¹, Thomas Knudsen² ¹ Leidos, RTP, NC; ² CCTE, EPA, RTP, NC

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 pathway¹:

- 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)



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 Sifter² v7.3. To establish confidence, we assessed literature connections to adverse outcomes associated with retinoid pathway disruption.

lits	s in <u>ToxC</u> a	<u>ast RARa</u> a	ssays. T	he re	sult	ts were filt	ered to o	che	micals with	AC50 under 2 i	nicromol
Cur	ation step	os were us	ed to re	move	ch	emical hits	with ca	uti	on flags.		
Assay			DSST	DSSToxID			al Name				AC50
				~~~~~							( <u>uM</u> )
			DTXS	ID302	355	6 Retinol					0.0763
			DTXS	ID202	288	0 Danazol					0.7338
			DTXS	ID902	045	3 Dieldrin					0.7704
			DTXS	XSID8024151 Imazalil						0.9084	
	TC DAD-	•		10804	127	2 6-Di-te	2.6 Di tort hutul 4 mothovunhanal				
5	IG RARa	trans_up		10002	752	C Endocul	Z,0-Di-tert-butyi-4-methoxyphenoi				
				10905	133	<u> </u>					1.3030
				10203	250		Iriflumizole				
			DIXS	ID403	892	2 Tetrabro	Tetrabromobisphenol A bis(2-hydroxyethyl) ether				
			DTXS	ID704	727	9 CP-5326	CP-532623				
			DTXS	DTXSID4021559 2,0			2,6-Di-tert-butyl-4-nitrophenol			1.8870	
NVS	NR_hRAR	Antagoni	st <u>DTXS</u>	DTXSID2026523 Symclosene						0.2638	
			DTXS	ID804	190	9 Dyspros	ium(III) cł	nlor	ide		0.8540
			DTXS	ID702	123	9 all-trans	-Retinoic	acio	ł		0.1258
NIV/			DTXS	DTXSID9022310			Daidzein				0.9369
	S INK NKA	Ra Agonisi		10002	200	1 Bentazon	no				1 8086
					770					•-	1.0500
				10803	//0	o Potassiu	m pernuc	proc	octanesultona	te	1.9118
_				_	_			_			
	А	В	с	D					E		
	Chembl										
1	Assays										
2	Target ID	Assay ID 🔹	Chemical Ct 斗	Туре	▼ As	say description					
3	CHEMBL2055	CHEMBL2401373	36	Binding	Tra A σ	insactivation of Gal4-	fused RARalpha DNA-binding do	(unkn) main-	own origin) expressed tagged BABaloba (upkr	n African green monkey COS1	cells at 100 nM by
5	CHEMBL2055	CHEMBL3118436	25	Binding	A	A	B B			own onginy ngana omanig ao	D
6	CHEMBL2055	CHEMBL3118434	24	Binding	Α						
7	CHEMBL2055	CHEMBL3880815	18	Binding	Α	Chemicals from	n ChEMBL				
8	CHEMBL2055	CHEMBL4417465	13	Binding	ln A	Target ID	Assav ID	-	Chemical ID	SMILES	
10	CHEMBL2055	CHEMBL3405114 CHEMBL4220761	8	Binding	A Ti	CHEMBI 2055	CHEMPI 1671	292	CHEMBI 1669776		
11	CHEMBI 2055	CHEMBL3224186		ADME	In	CHEMDE2000	CUENDL10/1	2.70			



2 CHEMBL2055

3 CHEMBL2055

14 CHEMBL2055

15 CHEMBL2055

16 CHEMBL2055

7 CHEMBL2055

18 CHEMBL2055

19 CHEMBL2055

20 CHEMBL2055

1 CHEMBL2055

22 CHEMBL2055

23 CHEMBL2055

24 CHEMBL2055 25 CHEMBL2055

1	MeSH term
2	Chemicals, pro row to sift on
3	MeSH Substan
4	Tretinoin
5	Vitamin A
6	beta Carotene
7	Isotretinoin
8	Fenretinide
9	4-oxoretinol
10	4-oxoretinoic a
11	4-hydroxyretin
12	retinal dimer

# **Candidate chemical compilation**

	DSSToxID	Chemical Name	AC50
			( <u>uM</u> )
	DTXSID3023556	Retinol	0.0763
	DTXSID2022880	Danazol	0.7338
	DTXSID9020453	Dieldrin	0.7704
	DTXSID8024151	Imazalil	0.9084
p	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
	DTXSID4021559	2,6-Di-tert-butyl-4-nitrophenol	1.8870
onist	DTXSID2026523	Symclosene	0.2638
	DTXSID8041909	Dysprosium(III) chloride	0.8540
	DTXSID7021239	all-trans-Retinoic acid	0.1258
nist	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.

вс	D			E		F	G H	Fiau
 Assay ID Chemica 	Ct 斗 Type	 Assay description 			▼ Doct	iment ID 🔄 Target Prefe	rred Name 🔹 PMID	v
CHEMBL2401373	36 Binding	Transactivation of (Gal4-fused RARalpha (unkn	own origin) expressed i	n African green monkey COS1 cells at 100 nM by luciferase re <u>CHE</u>	MBL2396574 Retinoic acid	receptor alpha 23685	180
CHEMBL4195064	25 ADME	Agonist activity at (GAL4 DNA-binding domain-	tagged RARalpha (unkn	own origin) ligand-binding domain expressed in human HG5 <u>CHE</u>	MBL4190387 Retinoic acid	receptor alpha 29706	423
CHEMBL3118436	25 Binding	A A	D		D		C	F
CHEWBL3118434	18 Binding	A Chemicals	from ChEMBI					
CHEMBL4417465	13 Binding	In						
CHEMBL3405114	8 Binding	A Target ID	▼ Assay ID	Chemical ID	SMILES	*	Pref Chem Name	↓ Results
CHEMBL4220761	6 Binding	TI CHEMBL2055	CHEMBL1671298	CHEMBL1668776	CC(C)=CCCC1=C[C@@H](C/C(C)=C/CC/C(C)=C/CC2=C	C(=0)C=C(C)C2=0)OC1	YEZOQUINOLIDE	Activity:
CHEMBL3224186	6 ADME	In CHEMBI 2055	CHEMBI 4150340	CHEMBI 1545045	CC1=C2[C@H](/)=C(\C)=C()=C()=C()=C()=C()=C()=C()=C()=C()=C	2001	VALERENIC ACID	Activity:
CHEMBL1671298	6 Binding	EI CHEMBI 2055	CHEMBI 3100731	CHEMBI 3098771	C(1=C/C=C/C(C)=C/C(1=O)O)/C=C1)/C(CCC2CCC2)		LIAB-30	Activity: 1.0 %
CHEMBL3405113	5 Binding		CHEMPL2207476	CHEMPL2009771	cc(-c/c-c/c(c)-c/c(-0)0)/c-c1/cccc2cccc21			Activity: 1.0 %
CHEMBL3297476	5 Binding		CHEIVIDL3237470			-2(0)(0)-202)001		Activity: 1.0 %
CHEMBL3292973	5 Binding	A CHEIVIBL2055	CHEMBL1071298	CHEIVIBL1008774	CC(C)=CCCC1=C[C@@H](C/C(C)=C/CC[C@@]2(C)C=C	-	TUBERATULIDE B	Activity:
CHEMBL3405115	4 Binding	A CHEMBL2055	CHEMBL1671298	CHEMBL1668773	CC(=0)CC/C=C(\C)C[C@@H]1C=C(CCC=C(C)C)C(=0)C	1	TUBERATOLIDE A	Activity:
CHEMBL1227700	4 Eunctiona	CHEMBL2055	CHEMBL3880815	CHEMBL3707313	CC(C)(C)c1cc(-c2cc(-c3ccc(C(=O)O)cc3)ccc2OCCO)ccc1	N1CCCC1	TRIFAROTENE	Kd: 500.0 nM
CHEMBL1061276	4 Functiona	A CHEMBL2055	CHEMBL4195064	CHEMBL3707313	CC(C)(C)c1cc(-c2cc(-c3ccc(C(=O)O)cc3)ccc2OCCO)ccc1	N1CCCC1	TRIFAROTENE	EC50: 500.0 nM
CHEMBL4430355	3 Binding	A CHEMBL2055	CHEMBL2401367	CHEMBL38	CC1=C(/C=C/C(C)=C/C=C/C(C)=C/C(=O)O)C(C)(C)CCC	1	TRETINOIN	EC50: 50.0 nM
CHEMBL4187072	3 Binding	In CHEMBL2055	CHEMBL2401373	CHEMBL38	CC1=C(/C=C/C(C)=C/C=C/C(C)=C/C(=O)O)C(C)(C)CCC	1	TRETINOIN	FC: 55.0
CHEMBL4187066	3 Binding	TI CHEMBL2055	CHEMBL3118436	CHEMBL38	CC1=C(/C=C/C(C)=C/C=C/C(C)=C/C(=O)O)C(C)(C)CCC	1	TRETINOIN	FC: 110.8
CHEMBL4037125	3 Binding	A CHEMBL2055	CHEMBL3405117	CHEMBL38	CC1=C(/C=C/C(C)=C/C=C/C(C)=C/C(=O)O)C(C)(C)CCC	1	TRETINOIN	Activity:
CHEMBL3789268	3 Binding	A CHEMBI 2055	CHEMBI 4187069	CHEMBL38	C(1=C)/C=C/C(C)=C/C(C)=C/C(C)=C/C(C)/C(C)/C	- 1	TRETINOIN	FC50: 1.01 nM
CHEMBL3118432	3 Binding	A CHEMBL2000	CHEMBL4107005	CUEMBLOO		1	TRETINOIN	EC50, 1,01 mM

Figure 4. Sample data from Abstract Sifter MeSHMine



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

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 cilexet	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	
Ļ	DTXSID3023180	Itraconazole	0.55±0.09	0.16±0.02	2.16±0.41	Paclitaxel	Unpublished [Data
1	DTXSID10872412	Mometasone	0.90±0.08	6.81±1.19	0.813±0.112	Amodiaquine	51	
i	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	
1	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	
	DTXSID1048627	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	
	DTXSID1034187	Tamoxifen	21.4±20.3	14.0±1.11	3.34±1.55	Amodiaquine	51	
Ļ	DTXSID5023746	Zafirlukast	0.06±0.02	0.71±0.23	0.644±0.273	Amodiaquine	51	

References:

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.

Nancy C. Baker I <u>baker.nancy@epa.gov</u> I 0000-0002-8351-9435

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



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.

Next steps

В	
Landscape View	
	Update Article Counts
Preferred Name	Chemical /
Daunomycin hydrochloride	23541-50-6 OR Daunomycin hydroc
Daunorubicin	20830-81-3 OR Daunorubicin
Retinal	116-31-4 OR Retinal OR Retinaldeh
Ethanol	Ethanol[majr]
Acetaminophen	103-90-2 OR Acetaminophen
Retinol	68-26-8 OR Retinol OR Vitamin A
Indomethacin	53-86-1 OR Indomethacin
Phenobarbital	50-06-6 OR Phenobarbital
Formaldehyde	50-00-0 OR Formaldehyde
Bisphenol A	80-05-7 OR Bisphenol A
Tetrachlorodibenzodioxin	41903-57-5 OR Tetrachlorodibenzo
all-trans-Retinoic acid	302-79-4 OR all-trans-Retinoic acid
er al 1 au 1 a 1	

Considerations for selections to the shortlist are:

- Strong supportive data
- Environmental interest
- Chemical diversity

le data from ChEMBL

was searched for ne targets and records nloaded to Excel for and scanning for entities with identified or candidate s, the publication was and reviewed.

CYP26 candidate reference chemicals 40 compounds for

purchase and evaluation



Tox21 Cross-Partner Project 13 is developing a highthroughput assay for CYP26 bioactivity. The candidate chemical collection was reviewed to select 40 chemicals for a proof-of-concept study (manuscript in preparation).

Known developmental toxicant or known negative