



Guide to Nuclear Receptor-mediated Endocrine Activity
(NRMEA) Model
version 1.1.1

Table of Contents

1. Model explanation.....	2
1.1 Introduction.....	2
1.2 Model details.....	2
1.3 Applicability Domain.....	5
1.4 Hierarchy featured fragments for 12 nuclear receptors-mediated compounds.....	5
1.5 Hierarchy screening workflow for 12 nuclear receptors-mediated compounds.....	5
1.6 Model statistics.	6
2. Model usage.....	6
2.1 Input.....	6
2.2 Select models and prediction	6
2.3 Output	6
3. Differences from other softwares.....	7
Figures	8
Tables.....	24

1. Model explanation

1.1 Introduction

The model provides a qualitative prediction of 12 classical nuclear receptor-mediated endocrine disruptions, including Androgen Receptor (AR), Estrogen Receptor α/β (ER α/β), Glucocorticoid Receptor (GR), Mineralocorticoid Receptor (MR), Progesterone Receptor (PR), Retinoic Acid Receptor $\alpha/\beta/\gamma$ (RAR $\alpha/\beta/\gamma$), Thyroid Hormone Receptor α/β (TR α/β), and Vitamin D Receptor (VDR). It is implemented inside the VEGA online platform, accessible at: <https://www.vegahub.eu>.

Developers:

Wei Shi (njushiwei@nju.edu.cn), Haoyue Tan, Qinchang Chen, Hongxia Yu

State Key Laboratory of Pollution Control and Resources Reuse, School of the Environment,
Nanjing University, Nanjing, Jiangsu 210023, China

Jiangsu Environmental Monitoring Center, Nanjing University, Nanjing, Jiangsu 210023, China

Jiangsu Key Laboratory of Chemical Pollution Control and Resources Reuse, Nanjing University,
Nanjing, Jiangsu 210023, China

1.2 Model details

With N RMEA, you can predict chemicals without disrupting data about 12 NRs-mediated disrupting activities aforementioned. Chemical data were collected from two open free libraries, including ToxCast/Tox21 (<https://www.epa.gov/chemical-research/toxcast-dashboard>) and

ChEMBL (<https://www.ebi.ac.uk/chembl/>). It's worth noting that three predictive models (AR, ER α , GR) are built based on three in vitro assays: i) competitive binding assay, ii) reporter gene assay and iii) cytotoxicity (Table 1). Whereas other predictive models (ER β , PR, MR, RAR $\alpha/\beta/\gamma$, TR α/β , VDR) are built based on two in vitro assays: i) reporter gene assay and ii) cytotoxicity. Therefore, the predicted results of first three models are under the classical nuclear genomic mechanism, in which ligand-NR binding directly and leading abnormal transcription, and the later couldn't predict the disrupting mechanism result from ligand-bound NRs (The detailed information is in Table 2).

Hierarchy fragment features built NRMEA, including primary fragment, secondary fragment, and tertiary fragment. Primary fragments are the essential features of active compounds; specific complex fragments sufficient for active compounds were further extracted based on the primary fragments and were named as secondary fragments; and then the critical featured fragments about three types of disrupting activities (agonist/a-anta/antagonist) were extracted and we named them as tertiary fragment. Firstly, we took advantage of primary fragments to select compounds have the possibility of being active. Secondly, compounds were screened by secondary fragments according to their different types of primary fragments. Each compound contains primary and secondary fragments simultaneously, would be predicted as active compounds. Next, the types of disrupting activities of identified active compounds would be distinguished based on their tertiary fragments (Figure 1). This hierarchy fragments allow the exploration of a wide range of structurally diverse active compounds which could bind to receptors, and build up a narray of rules focusing on Agonist/A-Anta/Antagonist disruption.

All of the chemical data were prepared by removing all false SMILES strings, and deleting all duplicate compounds. Then, the data set was split randomly into training set and test set in the ratio of 4:1 (KNIME Analytics Platform, <https://www.knime.com/>). The training set was used to provide feature fragments and local physicochemical properties present across the completely active set. The test set was used in the external validation. By using substructure frequency analysis (eq 1) and substructure percentage analysis (eq 2), and PubChem fingerprint database was used to generate primary fragments to describe active disruptors as much as possible. According to the structural features, chemicals were divided into several subtypes based on their primary fragments. Then, specific secondary fragments were generated by using SARpy software (<https://www.vegahub.eu/portfolio-item/sarpy/>). Active disruptors, containing primary and related secondary fragments, were used to extract disrupting tertiary fragments. Disruptors were clustered according to their primary and secondary fragments, then, tertiary fragments were extracted independently by using SARpy software.

$$\text{Frequency of a fragment} = \frac{(N_{\text{fragment}}^A \times N_{\text{total}})}{(N_{\text{fragment_total}} \times N_A)} \quad (1)$$

$$\text{Percentage of a fragment} = \frac{N_{\text{fragment}}^A}{N_A} \text{ or } \frac{N_{\text{fragment}}^I}{N_I} \quad (2)$$

Where N_{fragment}^A is the number of active compounds that contain this fragment, N_{total} is the total number of compounds in the dataset, $N_{\text{fragment_total}}$ is the total number of compounds containing this fragment, and N_A is the total number of active compounds in the dataset. N_{fragment}^I is the number of inactive compounds that contain this fragment and N_I is the total number of inactive compounds in the dataset.

1.3 Applicability Domain

NRMEA is built on the base of hierarchy featured fragments, thus diverse compounds can be predicted except for mixtures, and we pose model into two steps.

- Small molecules with known experimental value.

If the predicted compound is the same with compounds in the dataset used to build the model, whose experimental values and predicted values will be written.

- Small molecules with unknown experimental value.

If the predicted compound isn't the same with compounds in the dataset used to build the model, the prediction will be extrapolated by model.

1.4 Hierarchy featured fragments for 12 nuclear receptors-mediated compounds.

The model goes through three types of featured fragments. The hierarchy featured fragments are the following Table 3-14.

1.5 Hierarchy screening workflow for 12 nuclear receptors-mediated compounds.

The model goes through three types of featured fragments. The hierarchical featured fragments-based screening workflow are the following Figure 2-13.

1.6 Model statistics.

Following, statistics obtained applying the model to its original dataset (Table 15, Figure 14-15).

2. Model usage

2.1 Input

Select a single structural file or a directory containing the molecules' structural files. Most common file formats (e.g. CSV) are supported, or you can input smile directly into the table. In order to obtain a standardized representation of compound, canonical smiles should be used firstly.

2.2 Select models and prediction

We have two kinds of different prediction modules. First, if you input one chemical, select one or several NR models that you wanna know, and click the bottom “Prediction”, predicted results will be got finally.

2.3 Output

The predicted results will be saved to software, automatically. Secondly, click the bottom “Save Predictions”, the results will be saved as data file of “results.xlsx” (Figure 19). The results includes

6 parts: 1) the model you use; 2) active/inactive prediction; 3) primary fragments that compound has; 4) secondary fragment that compound has; 5) agonist/a-anta/antagonist prediction; 6) tertiary fragment that compound has. Additionally, if you input a list of chemicals, the predicted results will be saved to data file of “results.xlsx” in the same way. However, subsequent rows will contain the predicted results for one molecule per row. The first column is the molecule's name, which is either obtained from the structural file. Subsequent columns are the predicted results for the molecules.

3. Differences from other softwares.

NRMEA is the first software built by hierarchy featured fragment, and established the direct relations

1 **Figures**

2 Figure 1. Hierarchy featured fragment-based screening workflow of NRMEA.

3 Figure 2. The detailed screening workflow of Estrogen Receptor Alpha (ERa).

4 Figure 3. The detailed screening workflow of Androgen Receptor (AR).

5 Figure 4. The detailed screening workflow of Estrogen Receptor Beta (ERb).

6 Figure 5. The detailed screening workflow of Glucocorticoid Receptor (GR).

7 Figure 6. The detailed screening workflow of Mineralocorticoid Receptor (MR).

8 Figure 7. The detailed screening workflow of Progesterone Receptor (PR).

9 Figure 8. The detailed screening workflow of Retinoic Acid Receptor Alpha (RARa).

10 Figure 9. The detailed screening workflow of Retinoic Acid Receptor Beta (RARb).

11 Figure 10. The detailed screening workflow of Retinoic Acid Receptor Gamma (RARg).

12 Figure 11. The detailed screening workflow of Thyroid Hormone Receptor Alpha (TRa).

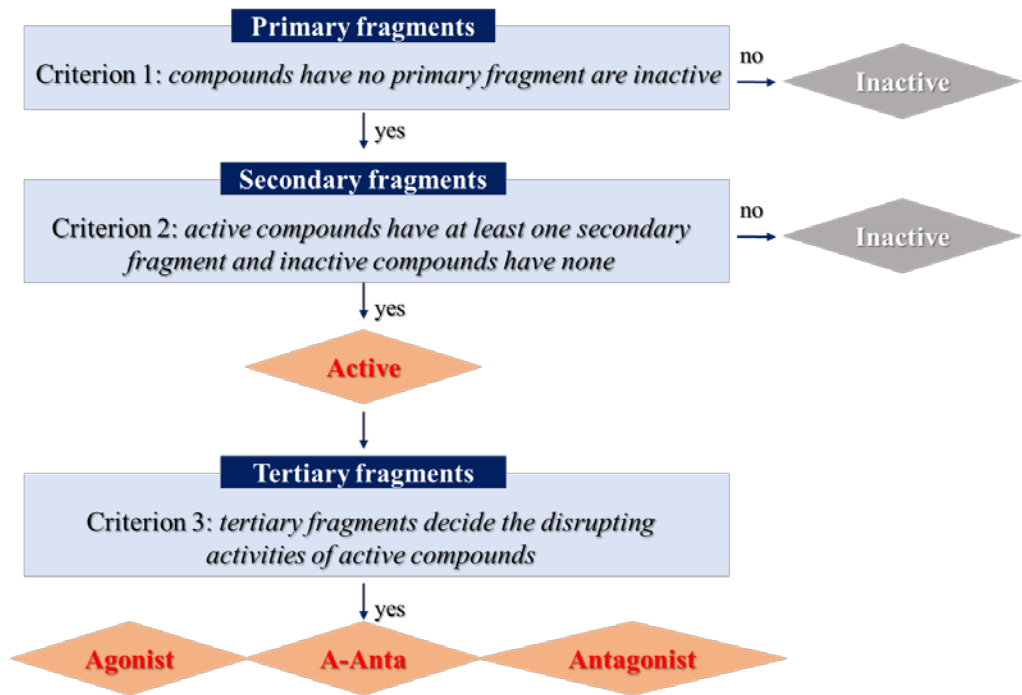
13 Figure 12. The detailed screening workflow of Thyroid Hormone Receptor Beta (TRb).

14 Figure 13. The detailed screening workflow of Vitamin D Receptor (VDR).

15 Figure 14. Performance of active/inactive prediction of external and internal validations for training
16 set and test set of our 12 nuclear receptor featured fragments-based screening models.

17 Figure 15. Performance of agonist/a-anta/antagonist prediction of external and internal validations
18 for training set and test set of our 12 nuclear receptor fragments- based screening models.

Hierarchy featured fragment-based screening workflow



Prediction category

Active: chemicals have potential disrupting activities.

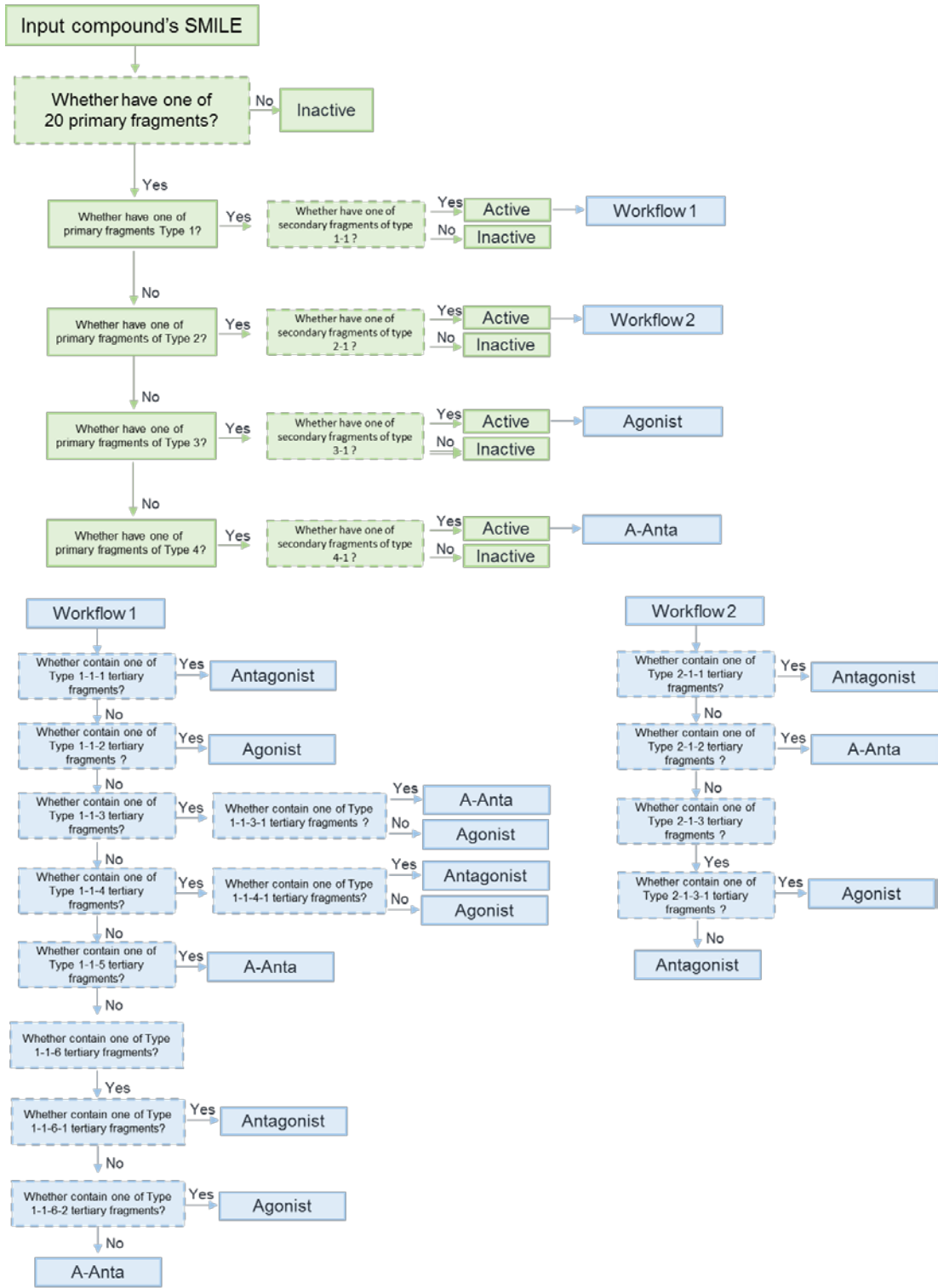
Inactive: chemicals are non-toxic.

Agonist: Active chemicals can only mimic natural hormone's action.

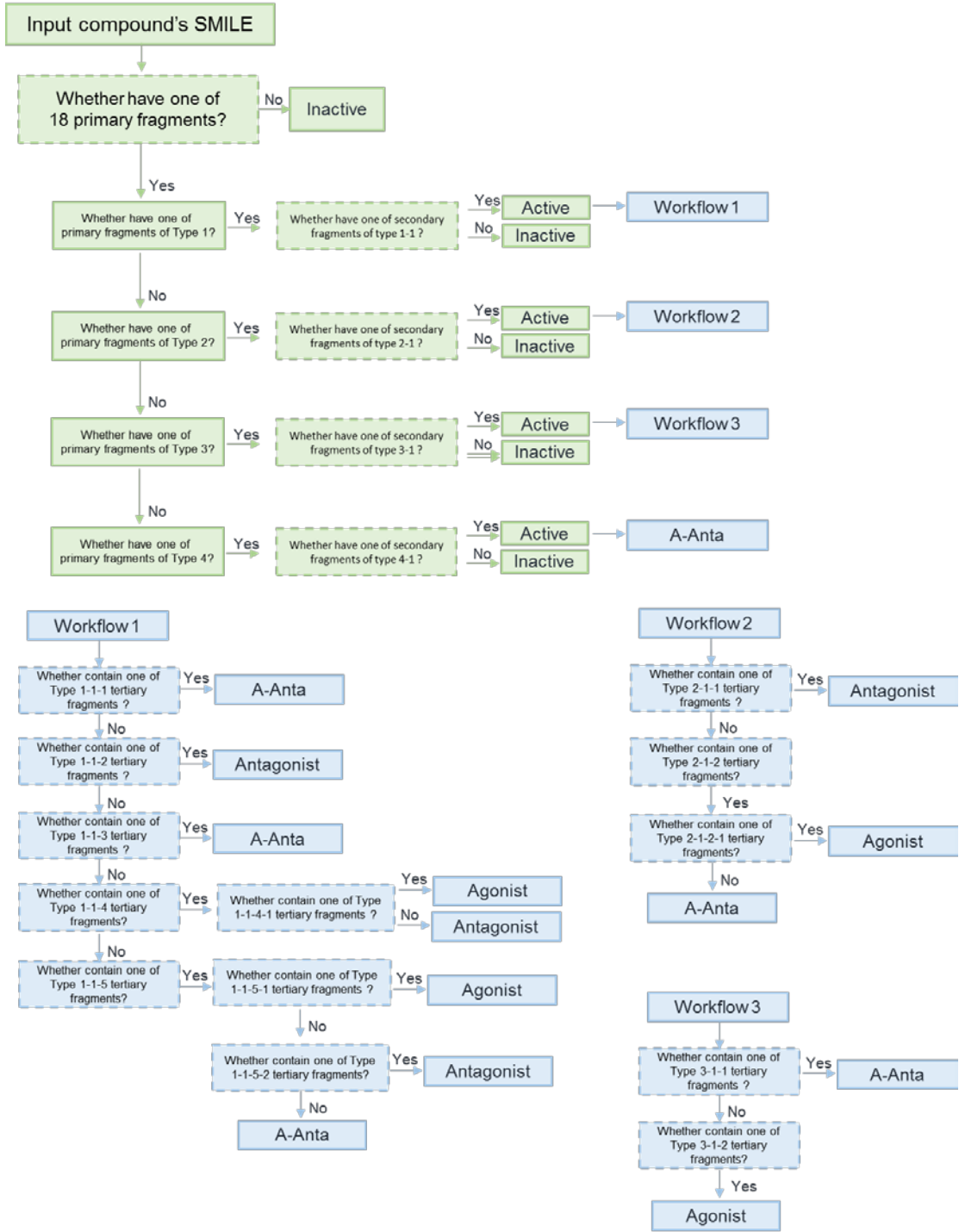
A-anta: Active chemicals not only can mimic but also can antagonize natural hormone's action.

Antagonist: Active chemicals can only antagonize natural hormone's action.

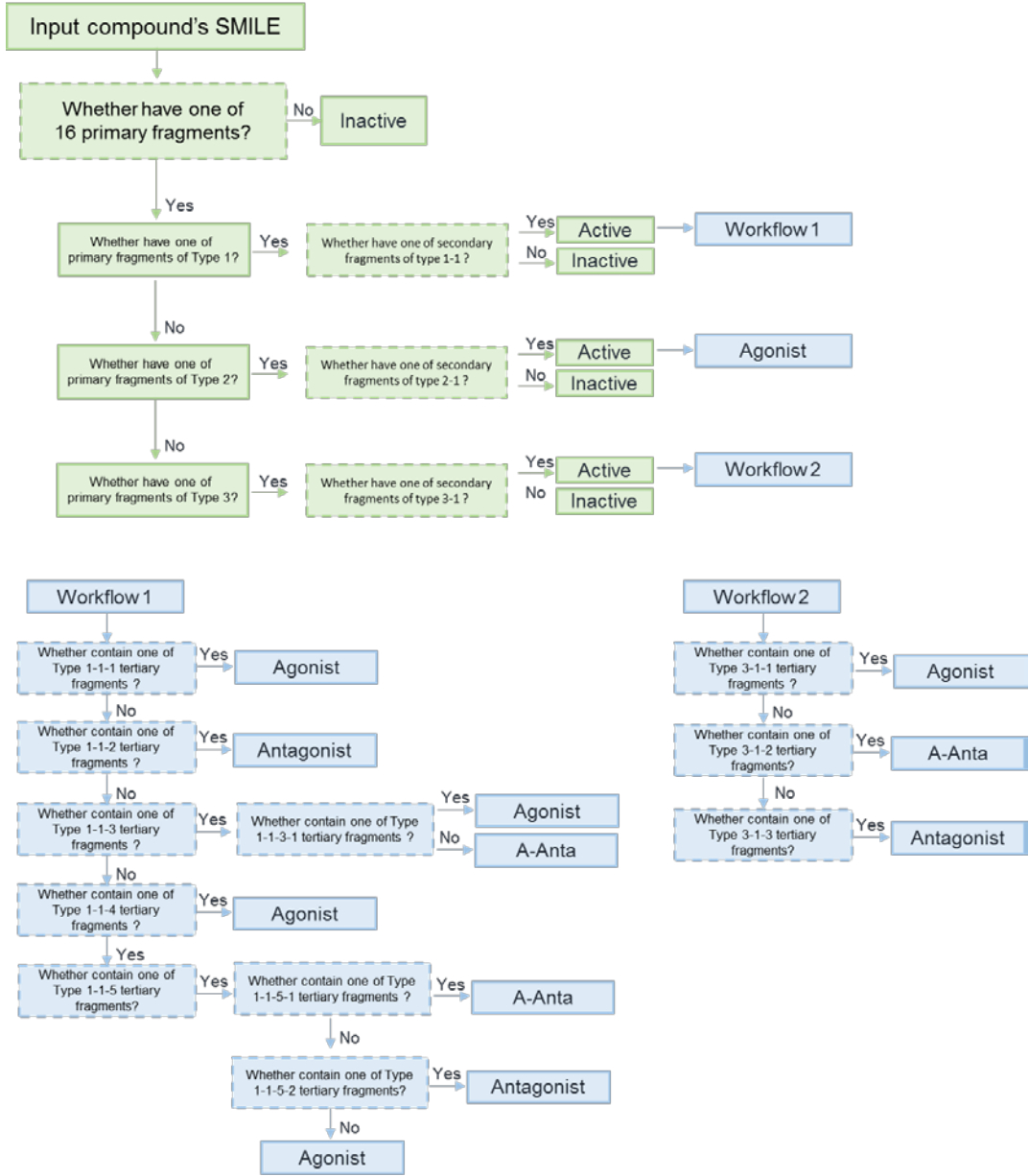
Estrogen Receptor Alpha Screening Workflow



Androgen Receptor Screening Workflow



Estrogen Receptor Beta Screening Workflow

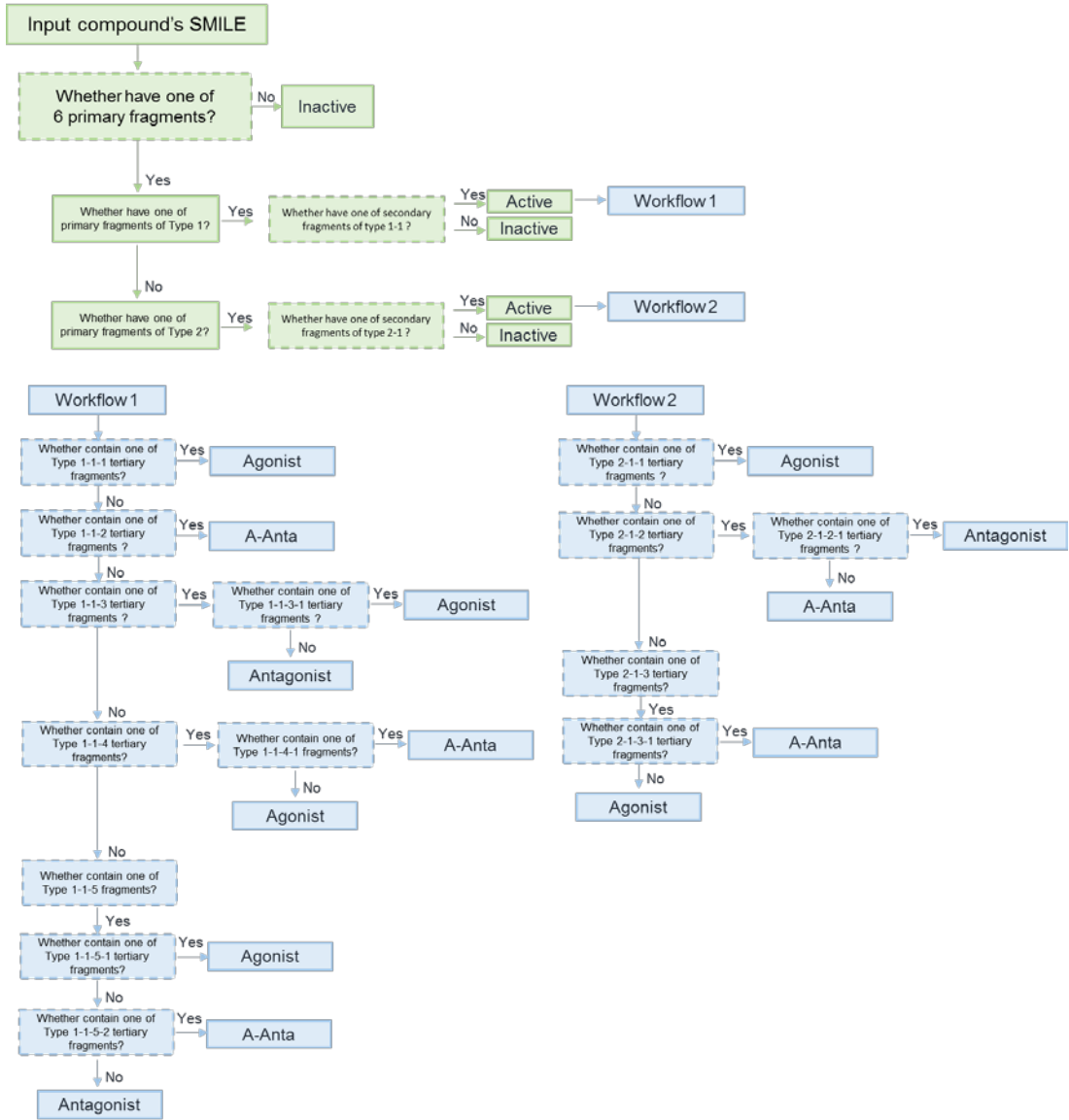


29

30

31 Figure 5.

Glucocorticoid Receptor Screening Workflow

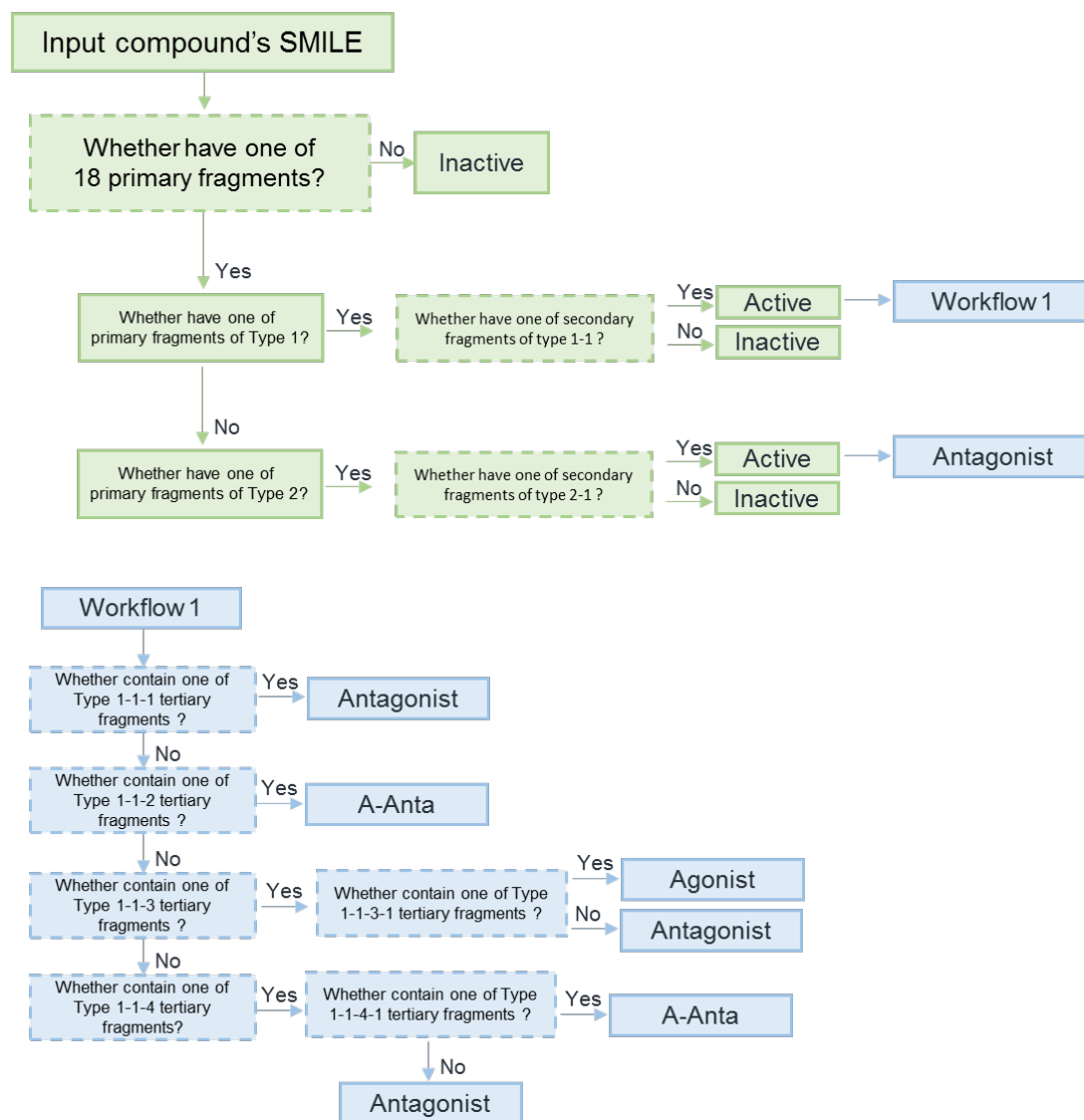


32

33

34 Figure 6.

Mineralocorticoid Receptor Screening Workflow

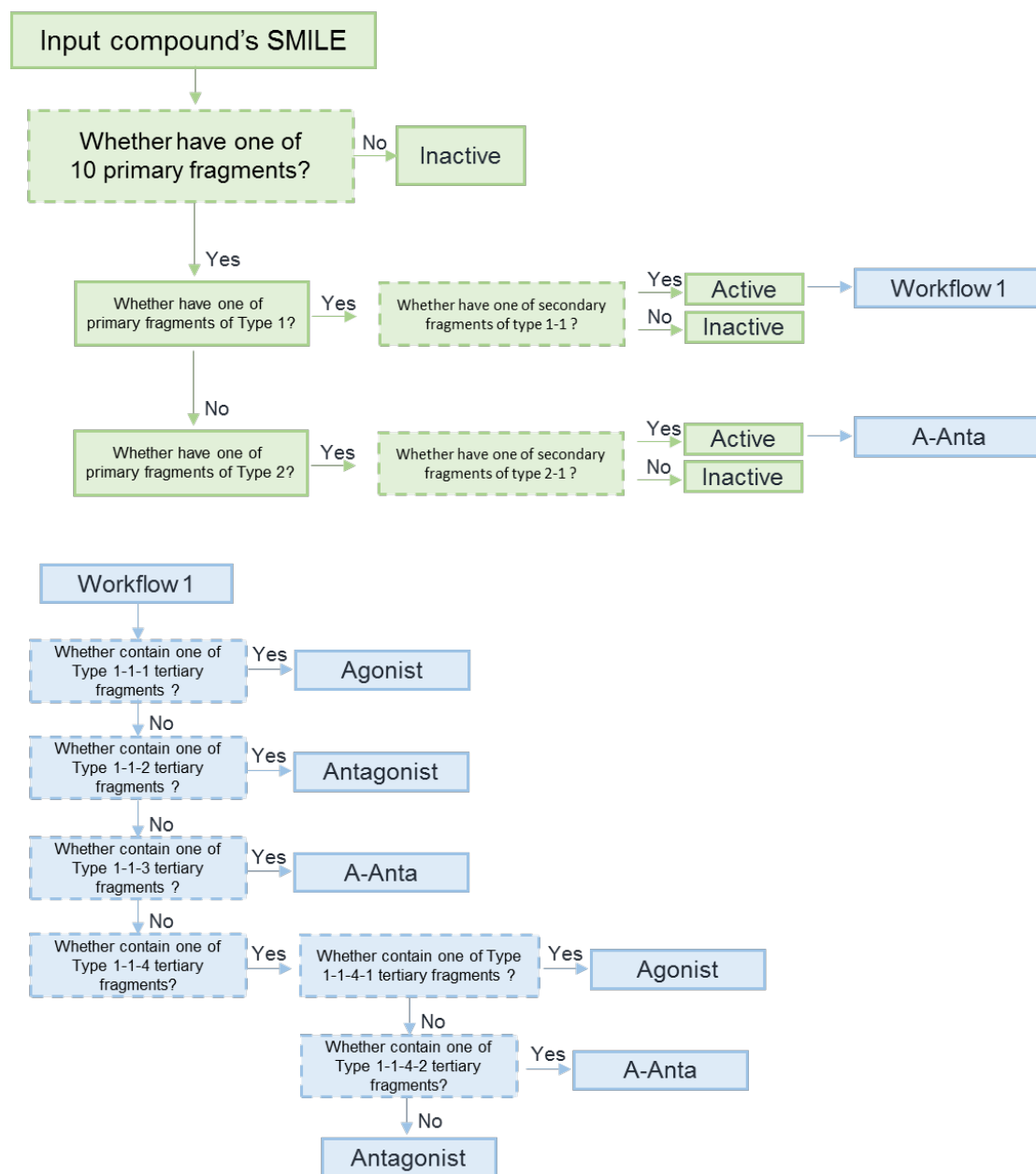


35

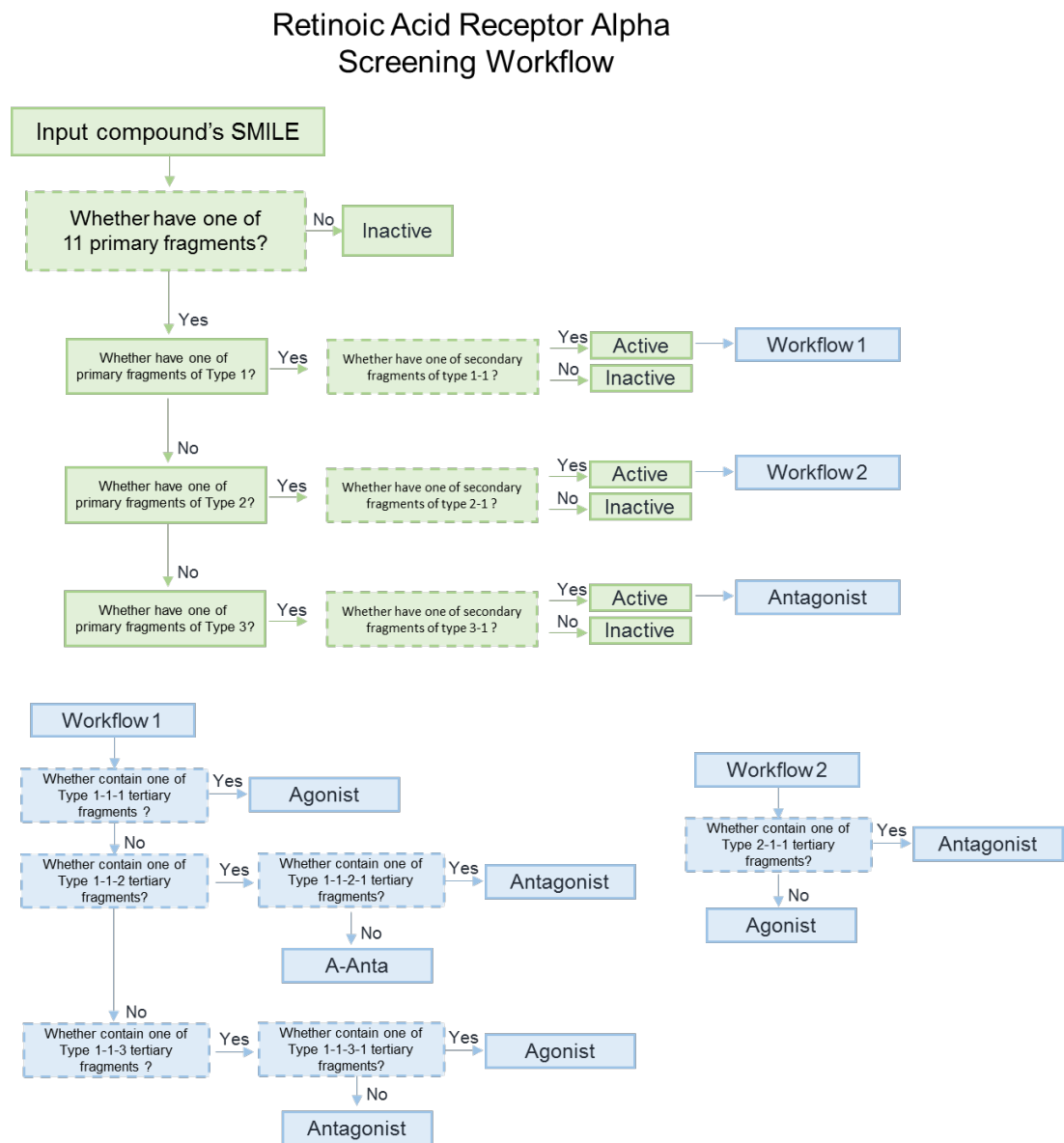
36

37

Progesterone Receptor Screening Workflow



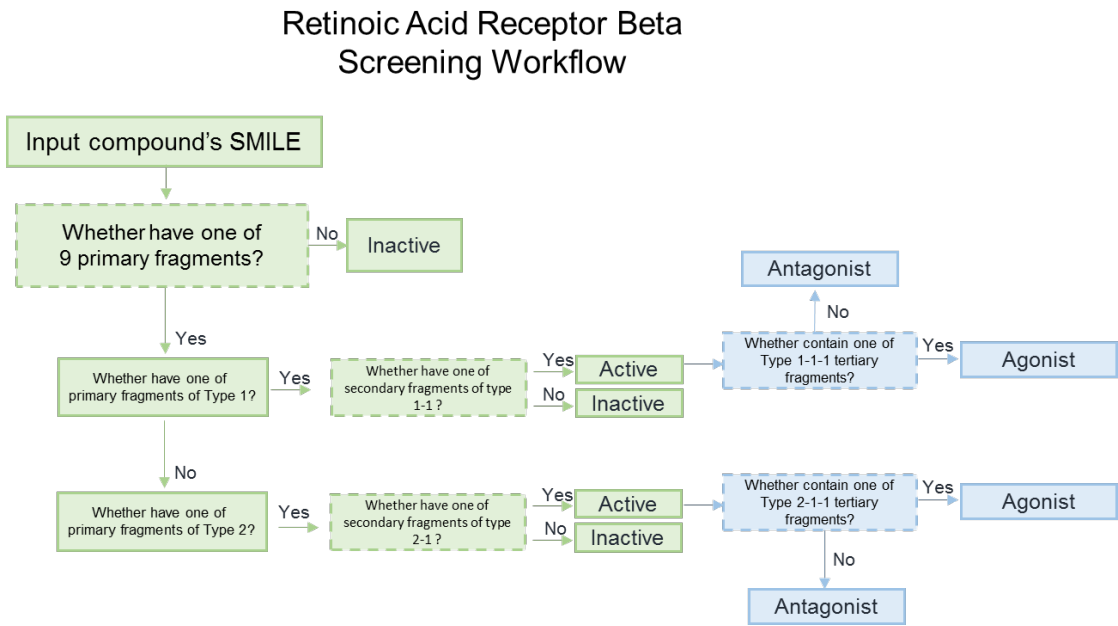
41 Figure 8.



42

43

44 Figure 9.

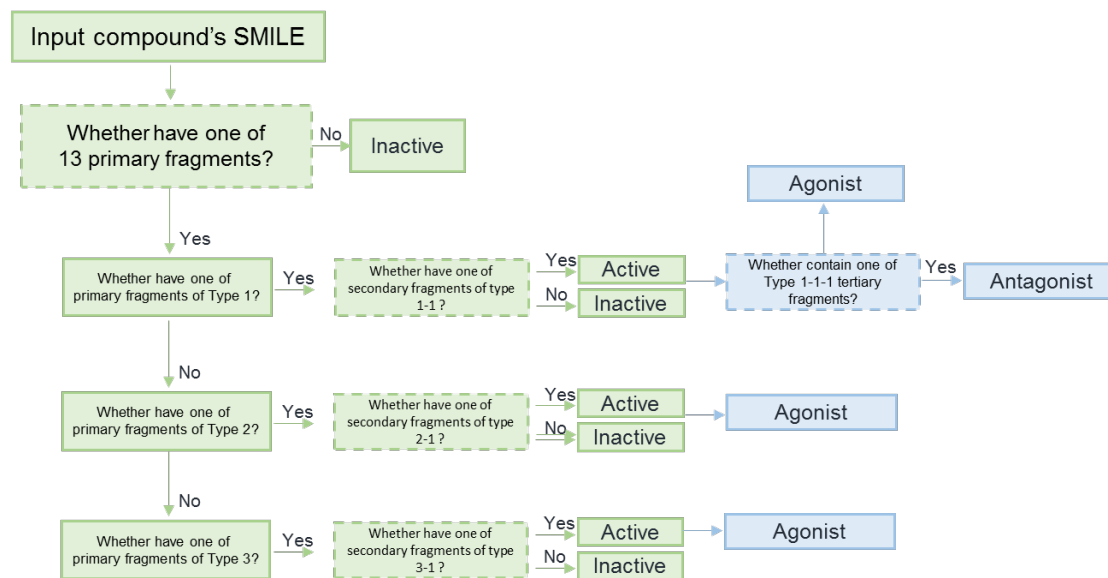


45

46

47 Figure 10.

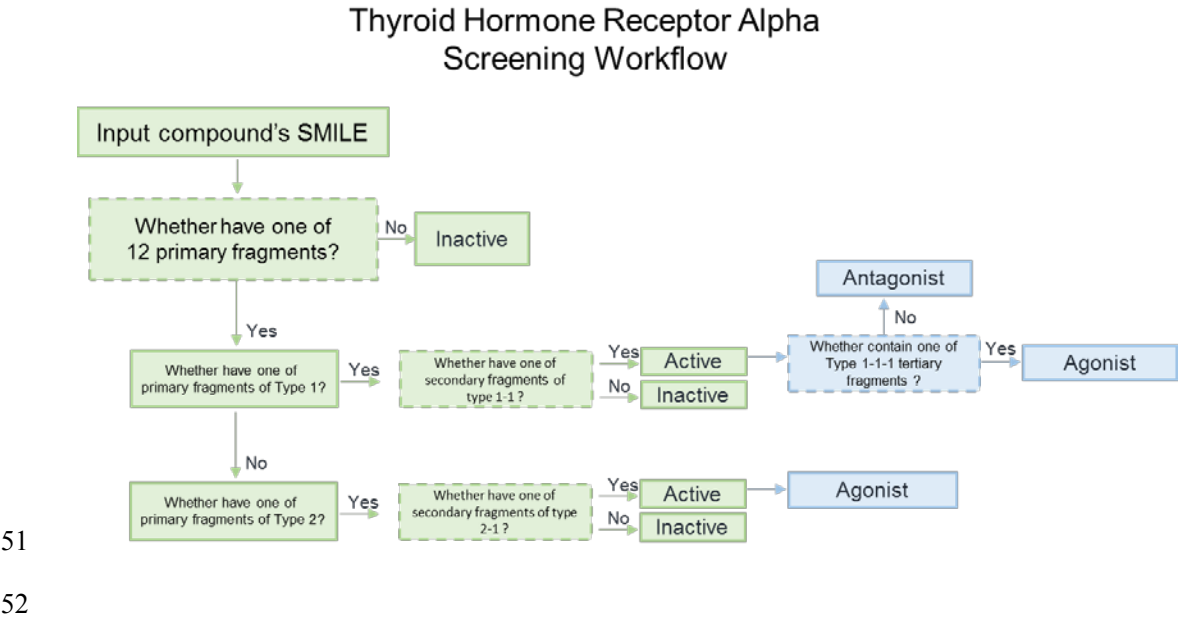
Retinoic Acid Receptor Gamma Screening Workflow



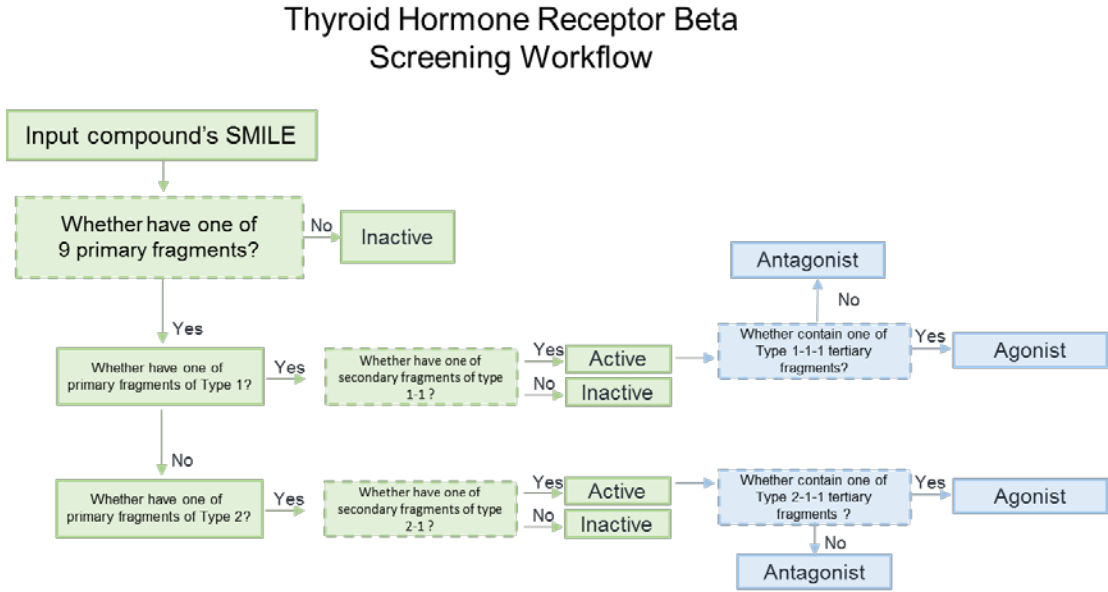
48

49

50 Figure 11.



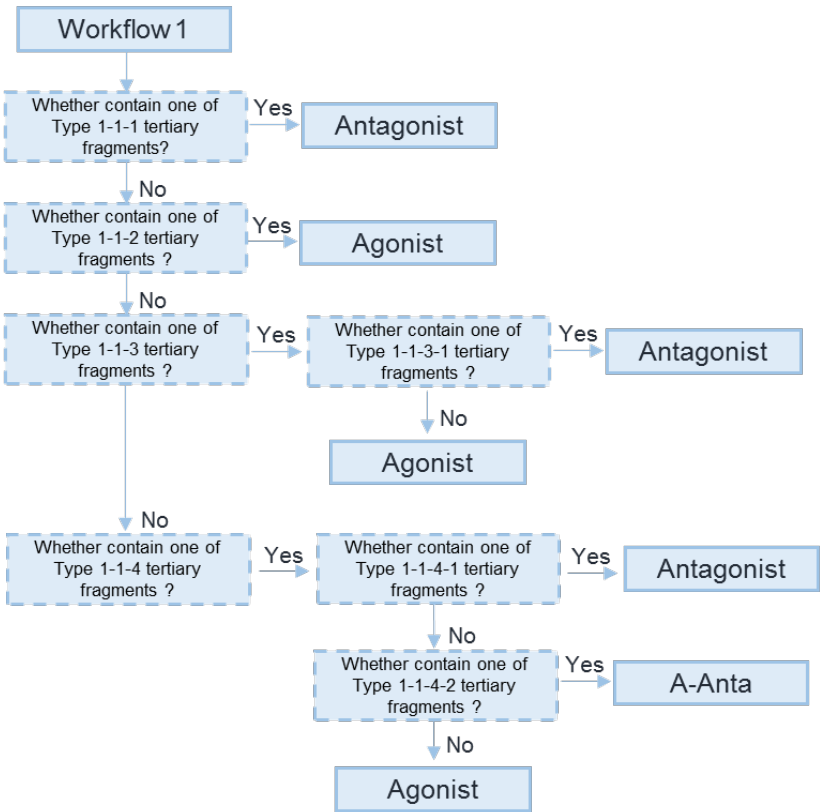
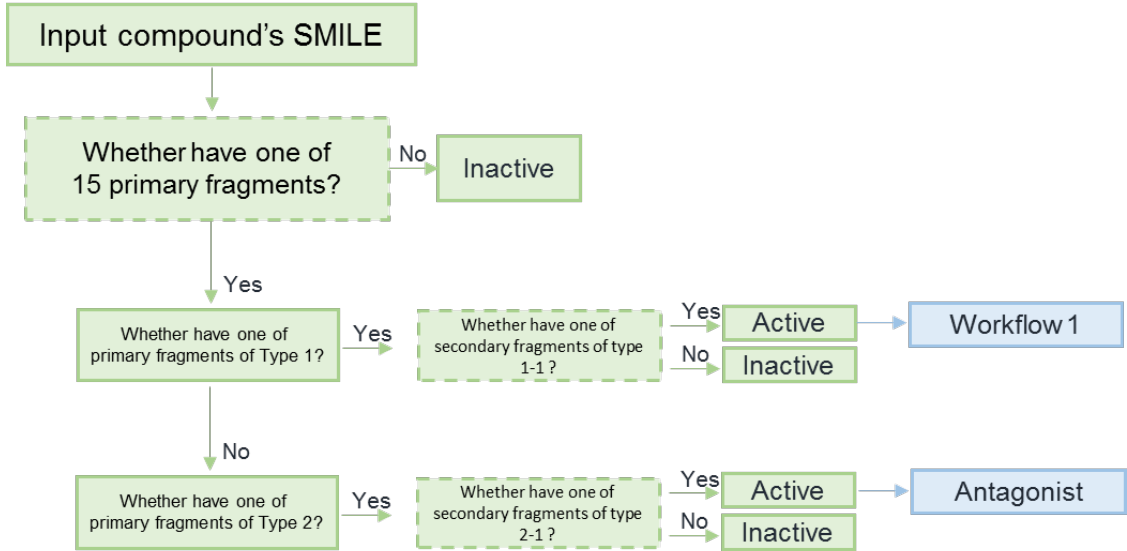
53 Figure 12.



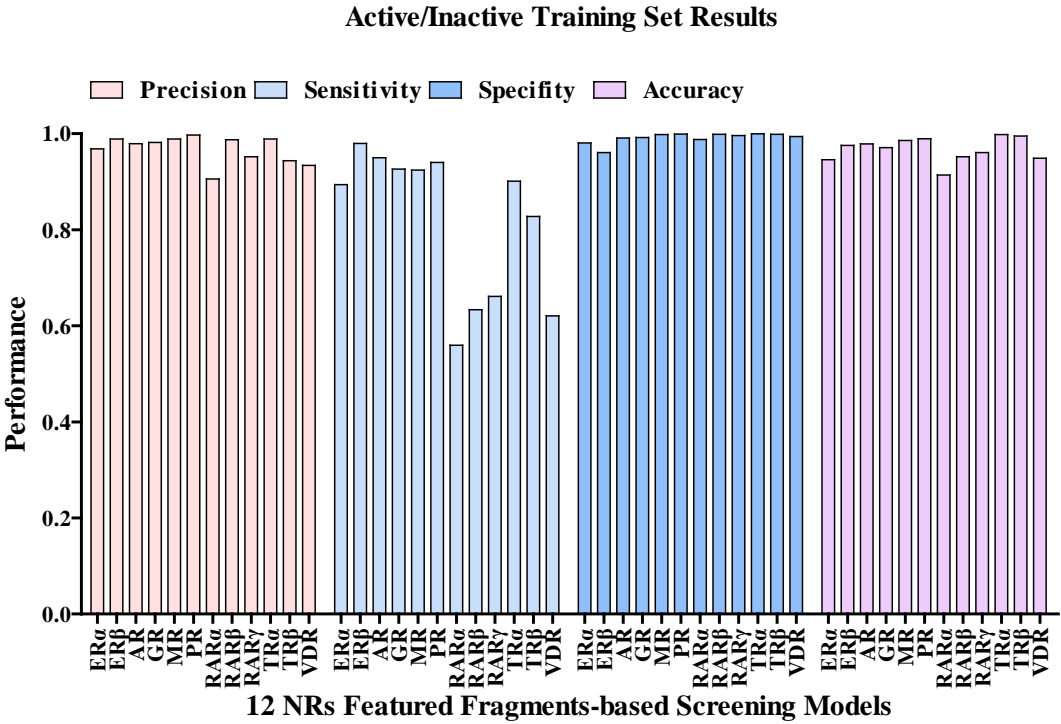
54

55

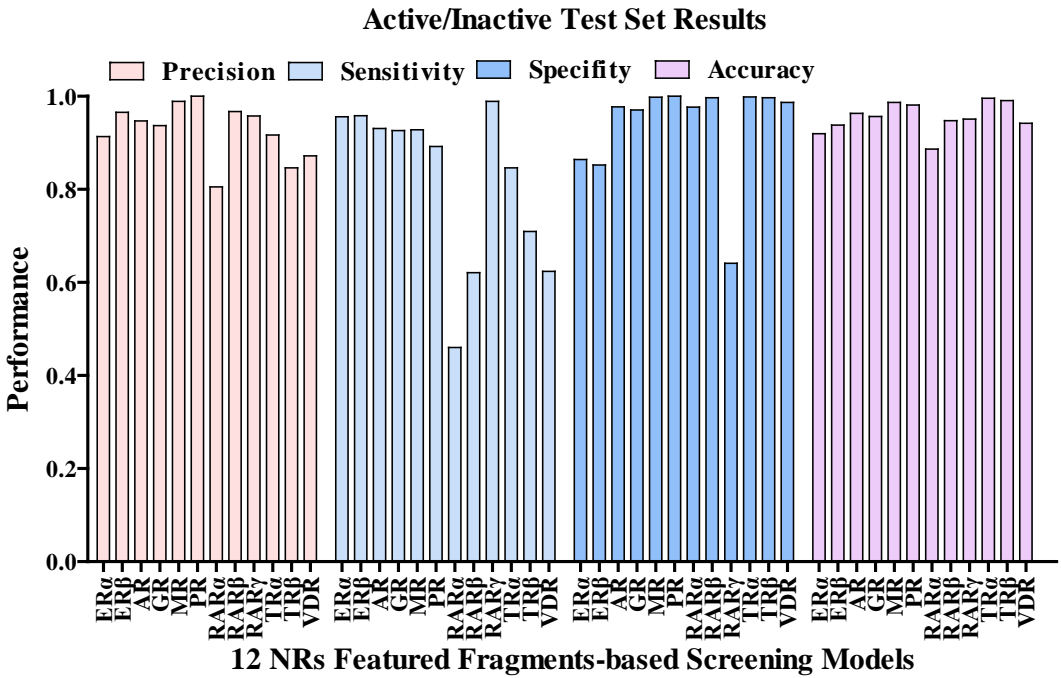
Vitamin D Receptor Screening Workflow



60 Figure 14.



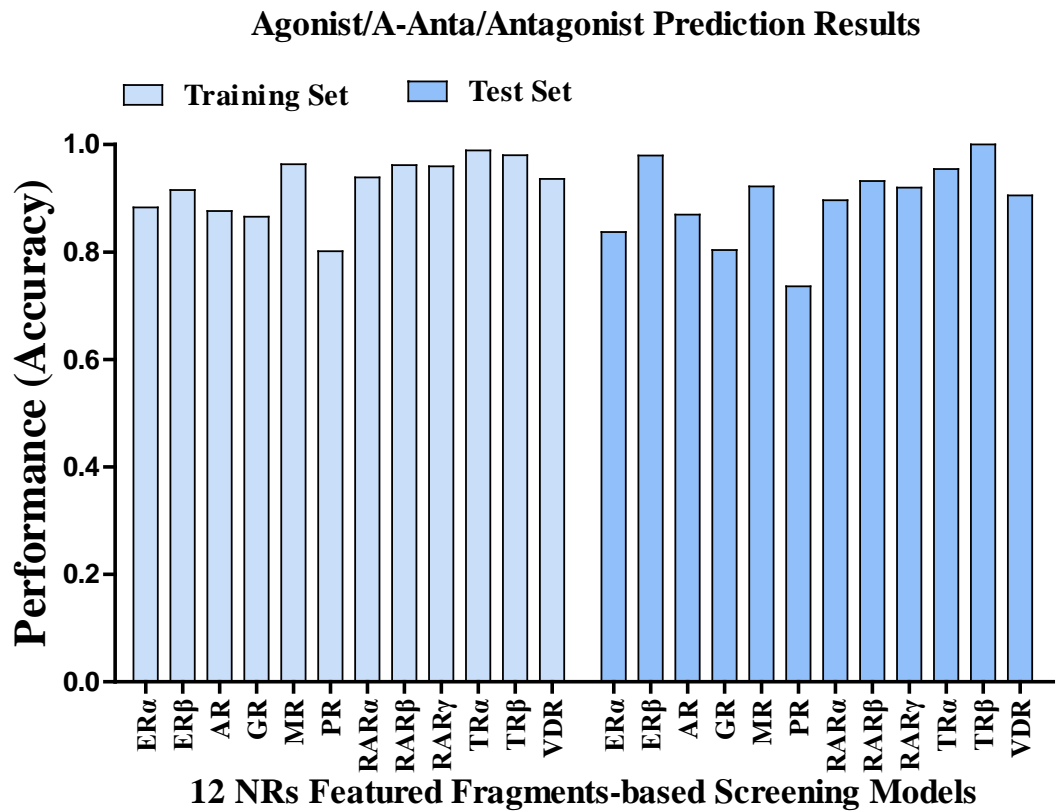
61



62

63

64 Figure 15.



65

66

Tables

Table 1. ToxCast/Tox21 database and ChEMBL database *In Vitro* Assays Used in 12 NRs Activity

Data Collection

ER α			
ToxCast/Tox21	Assay Name	Species	Type
	NVS_NR_hER	Homo sapiens	Competitive binding
	TOX21_ERa_BLA_Agonist_ratio		receptor gene
	TOX21_ERa_LUC_BG1_Agonist		receptor gene
	TOX21_ERa_BLA_Antagonist_ratio		receptor gene
	TOX21_ERa_LUC_BG1_Antagonist		receptor gene
	NCCT_HEK293T_CellTiterGLO		cytotoxicity
ChEMBL	Activity comment	Species	Assay type
	Binding affinity	Homo sapiens	B
	Agonist/Antagonist	sapiens	F
ER β			
ToxCast/Tox21	Assay Name	Species	Type
	NVS_NR_hER	Homo sapiens	Competitive binding
	NCCT_HEK293T_CellTiterGLO	sapiens	cytotoxicity
ChEMBL	Activity comment	Species	Assay type
	Binding affinity	Homo sapiens	B
	Agonist/Antagonist	sapiens	F
AR			
ToxCast/Tox21	Assay Name	Species	Type
	NVS_NR_hAR	Homo sapiens	Competitive binding
	Tox21_AR_BLA_Agonist_ratio		receptor gene
	Tox21_AR_BLA_Antagonist_ratio		receptor gene
	TOX21_AR_LUC_MDAKB2_Agonist		receptor gene
	TOX21_AR_LUC_MDAKB2_Antagonist		receptor gene
	NCCT_HEK293T_CellTiterGLO		cytotoxicity
ChEMBL	Activity comment	Species	Assay type
	Binding affinity	Homo sapiens	B
	Agonist/Antagonist	sapiens	F
GR			
ToxCast/Tox21	Assay Name	Species	Type
	NVS_NR_hGR	Homo sapiens	Competitive binding
	TOX21_GR_BLA_Agonist_ratio		receptor gene
	TOX21_GR_BLA_Antagonist_ratio		receptor gene

	NCCT_HEK293T_CellTiterGLO		cytotoxicity
ChEMBL	Activity comment	Species	Assay type
	Binding affinity	Homo	B
	Agonist/Antagonist	sapiens	F
MR			
ToxCast/Tox21	Assay Name	Species	Type
	NVS_NR_rMR	Homo	Competitive binding
	NCCT_HEK293T_CellTiterGLO	sapiens	cytotoxicity
ChEMBL	Activity comment	Species	Assay type
	Binding affinity	Homo	B
	Agonist/Antagonist	sapiens	F
PR			
ToxCast/Tox21	Assay Name	Species	Type
	NVS_NR_hPR	Homo	Competitive binding
	NCCT_HEK293T_CellTiterGLO	sapiens	cytotoxicity
ChEMBL	Activity comment	Species	Assay type
	Binding affinity	Homo	B
	Agonist/Antagonist	sapiens	F
RAR α			
ToxCast/Tox21	Assay Name	Species	Type
	ATG_RAR α _TRANS_up	Homo sapiens	receptor gene
	ATG_RAR α _TRANS_dn		receptor gene
	NCCT_HEK293T_CellTiterGLO		cytotoxicity
ChEMBL	Activity comment	Species	Assay type
	Binding affinity	Homo	B
	Agonist/Antagonist	sapiens	F
RAR β			
ToxCast/Tox21	Assay Name	Species	Type
	ATG_RAR β _TRANS_dn	Homo sapiens	receptor gene
	ATG_RAR β _TRANS_up		receptor gene
	NCCT_HEK293T_CellTiterGLO		cytotoxicity
ChEMBL	Activity comment	Species	Assay type
	Binding affinity	Homo	B
	Agonist/Antagonist	sapiens	F
RAR γ			
ToxCast/Tox21	Assay Name	Species	Type
	ATG_RAR γ _TRANS_up	Homo sapiens	receptor gene
	ATG_RAR γ _TRANS_dn		receptor gene
	NCCT_HEK293T_CellTiterGLO		cytotoxicity
ChEMBL	Activity comment	Species	Assay type
	Binding affinity	Homo	B
	Agonist/Antagonist	sapiens	F

TR α			
ToxCast/Tox21	Assay Name	Species	Type
	TOX21_TR_LUC_GH3_Agonist	Homo sapiens	receptor gene
	TOX21_TR_LUC_GH3_Antagonist		receptor gene
	NCCT_HEK293T_CellTiterGLO		cytotoxicity
ChEMBL	Activity comment	Species	Assay type
	Binding affinity	Homo sapiens	B
	Agonist/Antagonist	sapiens	F
TR β			
ToxCast/Tox21	Assay Name	Species	Type
	TOX21_TR_LUC_GH3_Agonist	Homo sapiens	receptor gene
	TOX21_TR_LUC_GH3_Antagonist		receptor gene
	NCCT_HEK293T_CellTiterGLO		cytotoxicity
ChEMBL	Activity comment	Species	Assay type
	Binding affinity	Homo sapiens	B
	Agonist/Antagonist	sapiens	F
VDR			
ToxCast/Tox21	Assay Name	Species	Type
	TOX21_VDR_BLA_agonist_ratio	Homo sapiens	receptor gene
	TOX21_VDR_BLA_antagonist_ratio		receptor gene
	NCCT_HEK293T_CellTiterGLO		cytotoxicity
ChEMBL	Activity comment	Species	Assay type
	Binding affinity	Homo sapiens	B
	Agonist/Antagonist	sapiens	F

1 Table 2. Statistical Data of Chemicals Used in the Training Sets and the Test Sets of 12 NRs

2 Active/Inactive

ER α					
		Active			Inactive
		Agonist	A-Anta	Antagonist	
ToxCast/Tox21 ChEMBL	Training Set	338	155	290	1186
	Test Set	86	39	74	290
	Total	424	194	364	1476
ER β					
		Active			Inactive
		Agonist	A-Anta	Antagonist	
ToxCast/Tox21 ChEMBL	Training Set	295	60	102	1996
	Test Set	74	14	26	500
	Total	369	74	128	2496
AR					
		Active			Inactive
		Agonist	A-Anta	Antagonist	
ToxCast/Tox21 ChEMBL	Training Set	106	242	349	1592
	Test Set	21	60	78	395
	Total	127	302	427	1987
GR					
		Active			Inactive
		Agonist	A-Anta	Antagonist	
ToxCast/Tox21 ChEMBL	Training Set	236	326	40	1495
	Test Set	59	81	10	373
	Total	295	407	50	1868
MR					
		Active			Inactive
		Agonist	A-Anta	Antagonist	
ToxCast/Tox21 ChEMBL	Training Set	21	23	349	1994
	Test Set	6	6	75	499
	Total	27	29	424	2493
PR					
		Active			Inactive
		Agonist	A-Anta	Antagonist	
ToxCast/Tox21 ChEMBL	Training Set	164	116	123	1942
	Test Set	39	29	33	485
	Total	203	145	156	2427
RAR α					
		Active			Inactive

		Agonist	A-Anta	Antagonist	
ToxCast/Tox21 ChEMBL	Training Set	392	20	79	2387
	Test Set	99	5	23	596
	Total	491	25	102	2983
RAR β					
		Active			Inactive
		Agonist	A-Anta	Antagonist	
ToxCast/Tox21 ChEMBL	Training Set	244	-	127	2515
	Test Set	62	-	33	630
	Total	306	-	160	3145
RAR γ					
		Active			Inactive
		Agonist	A-Anta	Antagonist	
ToxCast/Tox21 ChEMBL	Training Set	226	-	75	2548
	Test Set	56	-	20	637
	Total	282	-	95	3185
TR α					
		Active			Inactive
		Agonist	A-Anta	Antagonist	
ToxCast/Tox21 ChEMBL	Training Set	79	-	21	4760
	Test Set	20	-	5	1190
	Total	99	-	26	5950
TR β					
		Active			Inactive
		Agonist	A-Anta	Antagonist	
ToxCast/Tox21 ChEMBL	Training Set	91	-	30	4760
	Test Set	23	-	7	1190
	Total	114	-	37	5950
VDR					
		Active			Inactive
		Agonist	A-Anta	Antagonist	
ToxCast/Tox21 ChEMBL	Training Set	262	31	427	5272
	Test Set	62	9	110	1319
	Total	324	40	537	6591

4 Table 3. The hierarchy featured fragments of Estrogen Receptor Alpha (ERa).

Primary fragments (n=20)			
Type 1 (n=6)	Type 2 (n=4)	Type 3 (n=5)	Type 4 (n=5)
O-C:C-[#1] C-C:C-O-[#1] O-C:C-C O-C:C:C-C O-C-C:C O-C-C-C:C	C:C-C-C C:C-C:C C:C-C=C C-C:C-C-C	O-C=C-C O-C-C=C O-C-C-C=C C-C-C-O-[#1] O-C-C-C-C	C=C-C=C [#1]-C=C-[#1] C=C-C-C-C C-C(C)-C-C C-C-C-C-C-C-C
Secondary fragments (n=38)			
Type 1-1 (follow type 1 primary fragments) (n=27)	Type 2-1 (follow type 2 primary fragments) (n=7)	Type 3-1 (follow type 3 primary fragments) (n=3)	
C(c2ccc(O)cc2)c2ccc(O)cc2 Cn1cccc1 CCCC(C)c1cccc1 NCCc2ccc(O)cc2 Oc1ccc(Cc2ccccc2)cc1 c1cccc(Cl)c1C Oc1ccc(cc1)C(=Cc1cccc1) Oc1ccc(CC(c2ccccc2))cc1 Oc1ccc2C(N(CCc2c1)) CCCCCc1ccc(O)cc1 c1ccc(cc1)c1ccc2c(cccc2)c1 Oc1ccc(cc1)c1ccc1c1cccc1 Oc1ccc(cc1)Sc1ccc(O)cc1 c1cc(=O)c2c(O)cc(O)cc2o1	c1ccc(cc1)c1cccc1 Oc1ccc(cc1)C(=Nc1cccc1) n1ccc2cc(O)ccc2c1 c1ccc1c1cccc1 Oc1ccc(cc1)c1cccc1 CCC(=C(c1cccc1)) n1cc(cc2ccccc12)c1cccc1 CC(c1ccc(O)cc1)c1ccc(O)cc1 Oc1ccc(cc1)n1cc2ccccc2n1 C(C)OC(=O)c1ccc(O)cc1 COc1ccc(OC)n1 Oc1ccc(CCCc2ccc(O)cc2)c(O)c1 Oc1ccc(cc1)c1ccc(cc1)c1cccc1	CCC(=C)c1cccc1 c1cc(C)nc(NC)n1 CCN1CCc2c([nH]c3ccccc23) C1 C(=O)CCCCCCCCCO CS(=O)(=O) Cc1cccc(c1)c1cccc1 CNC(=N)NCCC	C=CCC=CCCC CC1CCC(=O)C(C1) CC1CCCC(C1)C(O)C Type 4-1 (follow type 4 primary fragments) (n=1) CCCCCCCCCCCCC
Tertiary fragments (n=65)			
Type 1-1-1 (follow type 1-1 secondary fragments) (n=4, Target: Antagonist)	Type 1-1-2 (n=4, Target: Agonist)	Type 1-1-3 (n=5)	Type 1-1-3-1 (n=4, Target: A-Anta)
Oc1ccc2C(N(CCc2c1)) Oc1ccc(cc1)n1cc2ccccc2n1 n1ccc2cc(O)ccc2c1 COc1ccc(OC)n1	Oc1ccc(CCCc2ccc(O)cc2)c(O)c1 Oc1ccc(cc1)c1ccc(cc1)c1cccc1 Oc1ccc(cc1)c1ccc1c1cccc1 n1cc(cc2ccccc12)c1cccc1	Oc1ccc(cc1)C(=Nc1cccc1) c1ccc1c1cccc1 Oc1ccc(cc1)c1cccc1 C(C)OC(=O)c1ccc(O)cc1 c1cc(=O)c2c(O)cc(O)cc2o1	Oc1ccc(cc1)C(=N)c1ccc(O)cc1 c1ccc(OCC(=O)O)cc1 C(CC)COC(=O)c1ccc(O)cc1 c1ccc2c(c(O)cc(O)c2)c1=O
Type 1-1-3-2 (Target: Agonist)	Type 1-1-4 (n=3)	Type 1-1-4-1 (n=2, Target: Antagonist)	Type 1-1-4-2 (Target: Agonist)
If compounds at least contain one of the type 1-1-3 fragments while don't contain any type 1-1-3-1 fragments,	Cn1cccc1 c1ccc(Cl)c1C c1ccc(cc1)c1cccc1	CCCCCNCC Clc1cc(Cl)c(cc1)	If compounds at least contain one of the type 1-1-4 fragments while don't contain any type 1-1-4-1

they will be defined as Agonists.			fragments, they will be defined as Agonists.
Type 1-1-5 (n=1, Target: A-Anta) CC(c1ccc(O)cc1)c1ccc(O)cc1	Type 1-1-6-1 (n=9, Target: Antagonist)	Type 1-1-6-2 (n=13, Target: Agonist)	Type 1-1-6-3 (Target: A-Anta)
Type 1-1-6 (n=10) CCCC(C)c1ccccc1 NCCc2ccc(O)cc2 Oc1ccc(Cc2ccccc2)cc1 Oc1ccc(cc1)C(=Cc1ccccc1) Oc1ccc(CC(c2ccccc2))cc1 CCCCCc1ccc(O)cc1 c1ccc(cc1)c1ccc2c(c1)ccc2 CCC(=C(c1ccccc1)) Oc1ccc(cc1)Sc1ccc(O)cc1 C(c2ccc(O)cc2)c2ccc(O)cc2	OC(=O)C=Cc1ccc(C)cc1 CCN(C)CCCCC c1cc(CC)c(O)cc1 CCCCCSCC C(c1ccc(O)cc1)c1ccc(OC)cc1 C(c1ccc(O)cc1)c1ccc(C)cc1 ONC(=O)CCCCCCC(=O)N OC(=O)C=Cc1ccccc1 OCCO	CC(=O)CC CC#N Oc1ccc2OC(C(Sc2c1)c1ccccc1) Oc1ccc(cc1)C1Sc2ccc(O)cc2 OC1 OC(CCCCC)c1ccccc1 CNC(=O)C C(C#C)C CC(Cc1ccccc1)c1ccc(O)cc1 Oc1ccc(C(=O)c2ccccc2)c(O)c 1 Oc1ccccc1Cl Cc1cc(O)ccc1CC CCN2CCCCC2 Oc1ccc(C=C(C)c2ccc(O)cc2)c c1	If compounds at least contain one of the type 1-1-6 fragments while don't contain any type 1-1-6-1 and type 1-1-6-2 fragments, they will be defined as A-Anta
Type 2-1-1 (follow type 2-1 secondary fragments) (n=2, Target: Antagonist)	Type 2-1-2 (n=1, Target: A-Anta) Cc1ccccc1)c1ccccc1	Type 2-1-3-1 (n=2, Target: Agonist)	Type 2-1-3-2 (Target: Antagonist)
c1cc(C)nc(NC)n1 CCN1CCc2c([nH]c3ccccc23)C1	Type 2-1-3 (n=4) CCC(=C)c1ccccc1 C(=O)CCCCCCCCCO CS(=O)(=O) CNC(=N)NCCC	CCC(=C)c1ccccc1 C(=O)CCCCCCCCCO	If compounds at least contain one of the type 2-1-3 fragments while don't contain any type 1-1-3-1 fragments, they will be defined as Agonists.
Type 3-1-1 (follow type 3-1 secondary fragments)	Type 4-1-1 (follow type 4-1 secondary fragments)		
All compounds which meet one of the type 4-1 secondary fragments will be defined as Agonist	All compounds which meet one of the type 4-1 secondary fragments will be defined as A-Anta		

6 Table 4. The hierarchy featured fragments of Androgen Receptor (AR).

Primary fragments (n=18)			
Type 1 (n=6)	Type 2 (n=3)	Type 3 (n=6)	Type 4 (n=3)
N-C-C-C N-C-C-[#1] C-N-C-C N-C-C-C C-N-C-C-C N-C-C-C-C	C-C-C-C C-C-C=C C-C-C-C	N-C-C-[#1] N-C-C=C C-N-C-[#1] C-C-N-C-C C-N-C-C-C N-C-C-C-C	C-C(C)-C-C C=C-C=C C-C-C-C-C-C
Secondary fragments (n=29)			
Type 1-1 (follow type 1 primary fragments) (n=14)	Type 2-1 (follow type 2 primary fragments) (n=8)	Type 3-1 (follow type 3 primary fragments) (n=3)	Type 4-1 (follow type 3 primary fragments) (n=4)
c1ccc2ccc(c2c1)C c1ccc(cc1)C#N CN(c1ccc(cc1)[N+](=O)[O-]) CC1CNC(=O)C1 CC1=CC(C)(C)Ne2ccc(cc12) Cc1cc(c(C)n1)c1ccc(cc1) CN(CCO)C(=O)	FC(F)c1cc(=O)oc2cc3NCCCc3cc12 Cn1cc(Ne2ccc(cc2))cc1C(=O) c1cc(c(en1)C#N)C CN(CC(F)(F)F)c1ccc2[nH]c(=O)cc(c2c1) Ne1ccc(C)c(c1)C(F)(F)F C(CC1CCCC1)C c1ccc(C=N)cc1	N#Cc1ccc(cc1) c2ccc(C)c(c2)C(F)(F)F CC(C)(C)C(CC)C c1ccccc1C1=NOC2CCCCC12 COc1ccccc1C#N CC(c1ccccc1)c1ccc(O)cc1 c1ccc(OCCN(C)C)cc1 Fe1ccc(cc1C(F)(F)F)	CCCNC(=O)CCCCC NC(=O)CCCCCCCCC CC(C)N(C)C(=O)C CCCCCCCCCCCCCC(=O)C=C C CCCCC(CCCC(F)C)C(=O) CCCCCCCCCCCCCC=CCC CCCCCCCCCCCCCC(=O)C
Tertiary fragments (n=47)			
Type 1-1-1 (follow type 1-1 secondary fragments) (n=2, Target: A-Anta)	Type 1-1-2 (n=4, Target: Antagonist)	Type 1-1-3 (n=1, Target: A-Anta)	Type 1-1-4 (n=2)
Cc1cc(c(C)n1)c1ccc(cc1) FC(F)c1cc(=O)oc2cc3NCCCc3cc12 2	c1ccc(C=N)cc1 CC1=CC(C)(C)Ne2ccc(cc12) Cn1cc(Ne2ccc(cc2))cc1C(=O) c1cc(c(en1)C#N)C	CN(CC(F)(F)F)c1ccc2[nH]c(=O)cc(c2c1)	CC1CNC(=O)C1 CN(CCO)C(=O)
Type 1-1-4-1 (n=6, Target: Agonist)	Type 1-1-4-2 (Target: Antagonist)	Type 1-1-5 (n=5)	Type 1-1-5-1 (n=3, Target: Agonist)
Cc1cc(ccc1)C#N CN1C(=O)N(C(=O)C1(CO)c1ccccc1) COc1cc(ccc1)C Cc1ccc(cc1)Br C1(CO)N(C)C(=O)N(C1) N2C(=O)C3C4CCC(C=C4)C3C2=	If compounds at least contain one of the type 1-1-4 fragments while don't contain any type 1-1-4-1 fragments, they will be defined as Antagonists.	C(CC1CCCC1)C c1ccc2nccc(c2c1)C c1ccc(cc1)C#N CN(c1ccc(cc1)[N+](=O)[O-]) Ne1ccc(C)c(c1)C(F)(F)F	CC1COc2cc3[nH]c(=O)cc(c3cc2N1) Ce3enene3 Oe1cc(c2cc3C4CCCC4Nc3cc2n1)

O			
Type 1-1-5-2 (n=12, Target: Antagonist)	Type 1-1-5-3 (Target: A-Anta)		Type 3-1-1 (follow type 3-1 secondary fragments) (n=1, Target: A-Anta)
CCCNC(=O)N CCCNc1ccc2c(cc(=O)[nH]c2c1)C(F)(F) CC1CC(C)(C)Nc2cc3nccc(c3cc12) Cn1c(=O)cc(c2cc3CCC(C)(C)Nc3cc12) c1ccccc1 FC(F)(F)c1cc(=O)[nH]c2cc(NC)cc1c12 NC3CCCC3 CCC(C)(C)N CC1(C)N(CCCC)C(=S)N(C1=O) [O-][N+](=O)c1ccc(cc1)N(C) n1c(=O)cc(c2cc3CCNc3cc12)C(F)(F) CS(=O)(=O)c1ccccc1	If compounds at least contain one of the type 1-1-5 fragments while don't contain any type 1-1-5-1 and type 1-1-5-2 fragments, they will be defined as A-Anta Type 2-1-1 (follow type 2-1 secondary fragments) (n=7, Target: Antagonist) N#Cc1ccc(cc1) c2ccc(C)c(c2)C(F)(F)F c1ccccc1C1=NOC2CCCCC12 COc1ccccc1C#N CC(c1ccccc1)c1ccc(O)cc1 c1ccc(OCCN(C)C)cc1 Fe1ccc(cc1C(F)(F)F)	Type 2-1-2 (n=1) CC(C)(C)C(CC)C Type 2-1-2-1 (n=1, Target: Agonist) CC1(C)CCC(=Cc2ccccc2)C1(O) Type 2-1-2-2 (Target: A-Anta) If compounds at least contain one of the type 2-1-2 fragments while don't contain any type 2-1-2-1 fragments, they will be defined as A-Anta	NC(=O)CCCCCCCCC Type 3-1-2 (n=2, Target: Agonist) CCCN(C(=O)CCCCC CC(C)N(C)C(=O)C Type 4-1-1 (follow type 4-1 secondary fragments) (Target: A-Anta) All compounds which meet one of the type 4-1 secondary fragments will be defined as A-Anta

8 Table 5. The hierarchy featured fragments of Estrogen Receptor Beta (ERb).

Primary fragments (n=16)			
Type 1 (n=7)	Type 2 (n=6)	Type 3 (n=3)	
<chem>O-C-C-[#1]</chem> <chem>O-C-C-C</chem> <chem>C-C-C-O-[#1]</chem> <chem>Cc1ccc(O)cc1</chem> <chem>O-C-C-C-C</chem> <chem>O-C-C:C</chem> <chem>O-C-C-C:C</chem>	<chem>O-C=C-C</chem> <chem>O-C-C=C</chem> <chem>C=C-C-O-[#1]</chem> <chem>O-C-C-C=C</chem> <chem>C-C-C-O-[#1]</chem> <chem>O-C-C-C-C</chem>	<chem>[#1]-C=C-[#1]</chem> <chem>C-C(C)-C-C</chem> <chem>C-C-C-C-C-C-C</chem>	
Secondary fragments (n=29)			
Type 1-1 (follow type 1 primary fragments) (n=23)	Type 2-1 (follow type 2 primary fragments) (n=2)	Type 3-1 (follow type 3 primary fragments) (n=4)	
<chem>CC(=C)c1ccc(O)cc1</chem> <chem>Oc1ccc(cc1)C(=Nc1ccccc1)</chem> <chem>n1cnc2cc(O)ccc2c1</chem> <chem>Cc1ccc(OCCN2CCCC2)cc1</chem> <chem>c1ccc(s1)c1ccccc1</chem> <chem>CC12CCC(CC1CCC2)</chem> <chem>n1cc(c(=O)c2ccccc12)c1ccc(O)c1</chem> <chem>Oc1ccc2C(N(Cc2c1)c1ccccc1)</chem> <chem>Oc1ccc(cc1)c1oc(cc1)</chem> <chem>Oc1ccc(cc1)c1oc2c(cc(O)cc2n1)</chem> <chem>OC(CC(=O))c1ccccc1</chem>	<chem>Oc1ccc(cc1)c1ccc2cc(O)ccc2c1</chem> <chem>Oc1ccc(cc1)n1cc2ccccc2n1</chem> <chem>CCC(=Cc1ccc(O)cc1)</chem> <chem>Oc1ccc(CC(C#N)c2ccccc2)cc1</chem> <chem>CC(Cc1ccc(O)cc1)c1ccc(O)cn1</chem> <chem>ON=Cc1cccc(c1)c1ccccc1</chem> <chem>Oc1ccc(CCc2ccccc2)cc1</chem> <chem>Oc1ccc(cc1)c1ccc1c1ccccc1</chem> <chem>CCC2=NOC(C2)</chem> <chem>c1ccc(O)cc1F</chem> <chem>C=Cc2cc(O)ccc2</chem> <chem>C(CCc1ccccc1)CC(=O)</chem>	<chem>CC#C</chem> <chem>CCC=C</chem>	<chem>CCC12CCCC=C1c1ccc3[nH]nnc3c1</chem> <chem>C2</chem> <chem>CCC12CCCC=C1c1ccc3[nH]ncc3c1</chem> <chem>C2</chem> <chem>ON=Cc1cccc(c1)c1ccccc1</chem> <chem>c1cc(NCC(C)C)nc(NCC(C)C)n1</chem>
Tertiary fragments (n=32)			
Type 1-1-1 (follow type 1-1 secondary fragments) (n=13, Target: Agonist)	Type 1-1-2 (n=3, Target: Antagonist)	Type 1-1-3 (n=1) <chem>Oc1ccc(cc1)C(=Nc1ccccc1)</chem> Type 1-1-3-1 (n=1, Target: Agonist) <chem>C1Oc2cc(O)ccc2N=C1c1ccc(O)c1</chem> Type 1-1-3-2 (Target: A-Anta) If compounds at least contain one of the type 1-1-3 fragments while don't contain any type 1-1-3-1 fragments, they will be defined as A-Anta	Type 1-1-4 (n=2, Target: Agonist)
<chem>C(CCc1ccccc1)CC(=O)</chem> <chem>c1ccc(O)cc1F</chem> <chem>c1ccc(s1)c1ccccc1</chem> <chem>CC12CCC(CC1CCC2)</chem> <chem>n1cc(c(=O)c2ccccc12)c1ccc(O)c1</chem> <chem>Oc1ccc(cc1)c1oc(cc1)</chem> <chem>Oc1ccc(cc1)c1oc2c(cc(O)cc2n1)</chem> <chem>Oc1ccc(cc1)c1ccc2cc(O)ccc2c1</chem> <chem>Oc1ccc(CC(C#N)c2ccccc2)cc1</chem> <chem>CC(Cc1ccc(O)cc1)c1ccc(O)cn1</chem>	<chem>n1cnc2cc(O)ccc2c1</chem> <chem>Oc1ccc2C(N(Cc2c1)c1ccccc1)</chem> <chem>Oc1ccc(cc1)n1cc2ccccc2n1</chem>		<chem>Cc1ccc(OCCN2CCCC2)cc1</chem> <chem>OC(CC(=O))c1ccccc1</chem>

<chem>ON=Cc1cccc(c1)c1ccccc1</chem> <chem>Oc1ccc(cc1)c1ccccc1c1ccccc1</chem> <chem>CCC2=NOC(C2)</chem>			
Type 1-1-5 (n=4)	Type 1-1-5-1 (n=3, Target: A-Anta)	Type 1-1-5-2 (n=1, Target: Antagonist)	Type 1-1-5-3 (Target: Agonist)
<chem>Oc1ccc(CCc2ccccc2)cc1</chem> <chem>C=Cc2cc(O)ccc2</chem> <chem>CC(=C)c1ccc(O)cc1</chem> <chem>CCC(=Cc1ccc(O)cc1)</chem>	<chem>Oc1ccc(cc1)C(=C1CCCCC1)</chem> <chem>ONC(=O)CCCCCCC(=O)N</chem> <chem>C(=O)CCCCCCC(=O)Nc1ccc(cc1)</chem>	<chem>C1=C(c2ccccc2)c2ccc(O)cc2C1</chem>	If compounds at least contain one of the type 1-1-5 fragments while don't contain any type 1-1-5-1 and type 1-1-5-2 fragments, they will be defined as Agonists
Type 2-1-1 (follow type 2-1 secondary fragments)	Type 3-1-1 (follow type 3-1 secondary fragments) (n=2, Target: Agonist)	Type 3-1-2 (n=1, Target: A-Anta)	Type 3-1-3 (n=1, Target: Antagonist)
All compounds which meet one of the type 2-1 secondary fragments will be defined as Agonist	<chem>CCC12CCCC=C1c1ccc3[nH]nnc3c1</chem> <chem>C2</chem> <chem>CCC12CCCC=C1c1ccc3[nH]nnc3c1</chem> <chem>C2</chem>	<chem>ON=Cc1cccc(c1)c1ccccc1</chem>	<chem>c1cc(NCC(C)C)nc(NCC(C)C)n1</chem>

10 Table 6. The hierarchy featured fragments of Glucocorticoid Receptor (GR).

Primary fragments (n=6)			
Type 1 (n=4)	Type 2 (n=2)		
O-C-C-C:C O-C-C-C=C O-C-C-C-C-C-C-C O-C-C-C-C-C(C)-C	C-C(C)-C-C C-C-C-C-C-C-C-C		
Secondary fragments (n=23)			
Type 1-1 (follow type 1 primary fragments) (n=17)	Type 2-1 (follow type 2 primary fragments) (n=6)		
c1ccccc1C(C)CC(C)(O)C CCc1cc(ccc1)c1ccccc1 CC(C)(C(c1ccccc1))C c2cnn(c2)c2ccccc2 CC1CCCC2=CCCCC12 CCCCC1CCCc2ccccc12 C(C)(C)CC(O)C(F) S(=O)(=O)NC(COc1ccccc1)C	c1cc(Cl)c(cc1)C(C) CCC(O)(CC)C(=O) C(Cn1ccnc1) CCC(=C(c1ccccc1))c1ccccc1 c1cc2c(NC(C)(C)C(=O)C2(C)C)cc1 CCCCCOC(=O)c1ccc(O)cc1 CC(c1ccccc1)c1ccc(O)cc1 C(O)(c1ccccc1)c1ccccc1 Cc1cc(l)c(O)c(l)c1		
Tertiary fragments (n=29)			
Type 1-1-1 (follow type 1-1 secondary fragments) (n=1, Target: Agonist)	Type 1-1-2 (n=1, Target: A-Anta)	Type 1-1-3 (n=2)	Type 1-1-3-1 (n=1, Target: Agonist)
S(=O)(=O)NC(COc1ccccc1)C	c1cc2c(NC(C)(C)C(=O)C2(C)C)cc1	CCC(=C(c1ccccc1))c1ccccc1 Cc1cc(l)c(O)c(l)c1	c1ccc(OC)cc1
Type 1-1-3-2 (Target: Antagonist)	Type 1-1-4 (n=7)	Type 1-1-4-1 (n=14, Target: A-Anta)	Type 1-1-4-2 (Target: Agonist)
If compounds at least contain one of the type 1-1-3 fragments while don't contain any type 1-1-3-1 fragments, they will be defined as Antagonists.	C(Cn1ccnc1) CCCCCOC(=O)c1ccc(O)cc1 c1ccccc1C(C)CC(C)(O)C CCc1cc(ccc1)c1ccccc1 CC(C)(C(c1ccccc1))C c2cnn(c2)c2ccccc2 C(C)(C)CC(O)C(F)	C(O)(CC=C) CC(C)(Cc1ccc(O)cc1) OC(=O)N c1cncnc1 c1ccc(C(=O))c(F)c1 c2cc(C)c3NC(C)(C)CC(C)c3c2Cl C(c1ccccc1)C(C)(C)C(=O)Nc1nncs1 Cc1cnn(c1)c1ccc(F)cc1 c1cc(O)c(O)cc1 C=C(C)C CCc1ccc(cc1F) CC(C)C1(C)CCCc2ccccc12 c1ccc(cc1)C(c1ccccc1) CC1(CC(O)C)CCCc2ccccc12	If compounds at least contain one of the type 1-1-4 fragments while don't contain any type 1-1-4-1 fragments, they will be defined as Agonists.
Type 1-1-5 (n=6)	Type 1-1-5-1 (n=3, Target: Agonist)	Type 1-1-5-2 (n=5, Target: A-Anta)	Type 1-1-5-3 (Target: Antagonist)
CC1CCCC2=CCCCC12 CCCCC1CCCc2ccccc12 c1cc(Cl)c(cc1)C(C) CCC(O)(CC)C(=O) CC(c1ccccc1)c1ccc(O)cc1 C(O)(c1ccccc1)c1ccccc1	c1ccc(F)nc1 OCCC(O)(CC)C CC(c1ccccc1)c1ccccc1	CC1(O)CCC2(C)C(CCc3cc(O)ccc23)C1 CCC12CC(O)C(CC1CCc1cc(O)ccc21) C(O)C1CCCC2=CC(=O)CCC12C CC1CCCC2=CC(=O)CCC12C c1cc(ccc1O)C	If compounds at least contain one of the type 1-1-5 fragments while don't contain any type 1-1-5-1 and type 1-1-5-2 fragments, they will be defined as Antagonists
Type 2-1-1 (follow type 2-1 secondary	Type 2-1-2 (n=2)	Type 2-1-2-1 (n=2, Target: Antagonist)	Type 2-1-2-2 (Target: A-Anta)

fragments) (n=1, Target: Agonist)			
<chem>CC(C)NS(=O)(=O)c1ccccc1</chem>	<chem>C1CC(C)(C)Nc2ccc(cc12)CCCCCCCCCCCC</chem>	<chem>CCCCCCCCCCCCCCCCC1(C)Nc2c3CCCCc3c(cc2C(C)(C)C1=O)</chem>	If compounds at least contain one of the type 2-1-2 fragments while don't contain any type 2-1-2-1 fragments, they will be defined as A-Anta
Type 2-1-3 (n=3)	Type 2-1-3-1 (n=3, Target: A-Anta)	Type 2-1-3-2 (Target: Agonist)	
<chem>c1enn(c1)c1ccccc1</chem> <chem>CC(C)(C(c1ccccc1))C(=O)N</chem> <chem>[N+](CC)CC</chem>	<chem>CC(CNC(=O))C(c1ccccc1)</chem> <chem>c1ccc2c(nen2c1)</chem> <chem>CC(CNS(=O)(=O)C)C(c1ccccc1)</chem>	If compounds at least contain one of the type 2-1-3 fragments while don't contain any type 2-1-3-1 fragments, they will be defined as Agonist	

11

12 Table 7. The hierarchy featured fragments of Mineralocorticoid Receptor (MR).

Primary fragments (n=18)			
Type 1 (n=9)		Type 2 (n=9)	
O-C-C:C	N-C:C-C	O-C-C-C-C	N-C=C-[#1]
O-C-C:C-C	C-N-C:C	O-C-C=C	N-C-C=C
O-C-C-C:C	N-C:C:C-C	O-C-C-C=C	N-C-C-C-C
N-C-C-[#1]	N-C-C:C	C-N-C-[#1]	C-N-C-C-C
	N-C-C-C:C		C-C-N-C-C
Secondary fragments (n=22)			
Type 1-1 (follow type 1 primary fragments) (n=21)			Type 2-1 (follow type 2 primary fragments) (n=1)
CC2OC(=O)N(C)C2 C1=NN(C(C1))c1ccc(cc1)C CC1=CC(C)(C)Nc2ccc(cc12) Cn1ccc(c2ccc3OCC(=O)Nc3c2)c 1 CC(C)(C(c1ccccc1))CN COC(=O)C1=CNC(=C(C)C1)C CC(C(c1ccccc1))C(=O)N	c1cc(CCC2CCCC2(C))ccc1 Cn1cc(c2ccc(cc2)C#N)c2ccccc12 CC=CCC(F)C(F)(F) CC1Cc2cc3c(ccnc3cc2N1)C(F)(F)F C1OCC(=O)Nc2ccc(cc12) C1=NN(C(C1)C1CCCC1) C1=NN(C(C1)c1ccc(F)cc1)c1ccccc1 1	C(c1ccccc1)c1c[nH]c2ccccc12 C1CCNc2cc3oc(=O)cc(c3cc12) CC1(C)OCNc2ccc(cc12) Oc1ccc2c(CCC3=CC(=O)CCC23C) c1 OC(C)c1cccs1 CC2C(=C)CCCC2(C) c2ccc3OCC(=O)Nc3c2	CC1=C(C(C(=C(C)N1)C)c1ccccc1) 1)
Tertiary fragments (n=22)			
Type 1-1-1 (follow type 1-1 secondary fragments) (n=15, Target: Antagonist)	Type 1-1-2 (n=1, Target: A-Anta)	Type 1-1-4 (n=2)	
CC2C(=C)CCCC2(C) CC2OC(=O)N(C)C2 C1=NN(C(C1))c1ccc(cc1)C COC(=O)C1=CNC(=C(C)C1)C c1cc(CCC2CCCC2(C))ccc1 Cn1cc(c2ccc(cc2)C#N)c2ccccc12 CC1Cc2cc3c(ccnc3cc2N1)C(F)(F)F C1OCC(=O)Nc2ccc(cc12)	C1=NN(C(C1)C1CCCC1) C(c1ccccc1)c1c[nH]c2ccccc12 C1CCNc2cc3oc(=O)cc(c3cc12) CC1(C)OCNc2ccc(cc12) Oc1ccc2c(CCC3=CC(=O)CCC23C) c1 OC(C)c1cccs1 C1=NN(C(C1)c1ccc(F)cc1)c1ccccc1 1	CC=CCC(F)C(F)(F)	c2ccc3OCC(=O)Nc3c2 Cn1ccc(c2ccc3OCC(=O)Nc3c2)c 1
		Type 1-1-3 (n=3)	Type 1-1-4-1 (n=2, Target: A-Anta)
		CC(C(c1ccccc1))C(=O)N CC1=CC(C)(C)Nc2ccc(cc12) CC(C)(C(c1ccccc1))CN	CCCO Cn1ccc(c2ccc3OCC(=O)Nc3c2)c 1
		Type 1-1-3-1 (n=2, Target: Agonist)	Type 1-1-4-2 (Target: Antagonist)
		CC(C)(C(c1ccccc1)c1ccc(O)cc1) c4ccncc4C	If compounds at least contain one of the type 1-1-4 fragments while don't contain any type 1-1-4-1 fragments, they will be defined as Antagonists.
		Type 1-1-3-2 (Target: Antagonist)	Type 2-1-1 (follow type 2-1 secondary fragments) (Target: Antagonist)

		If compounds at least contain one of the type 1-1-3 fragments while don't contain any type 1-1-3-1 fragments, they will be defined as Antagonists.	All compounds which meet one of the type 4-1 secondary fragments will be defined as Antagonists
--	--	--	---

13

14 Table 8. The hierarchy featured fragments of Progesterone Receptor (PR).

Primary fragments (n=10)			
Type 1 (n=6)		Type 2 (n=4)	
N-C:C-[#1] N-C:C-C N-C:C:C-C C-N-C:C N-C-C:C N-C-C-C:C		C:C-C-C C:C-C:C C:C-C=C C-C:C:C-C	
Secondary fragments (n=17)			
Type 1-1 (follow type 1 primary fragments) (n=13)			Type 2-1 (follow type 2 primary fragments) (n=4)
CN(C)c1ccc(C#N)c(Cl)c1 N(Cc1ccccc1)c1cccc(Cl)c1 Cc1c(C#N)cnc2c3C(CC(=O)Nc3sc12) Cn1cc(c2ccc(cc2)C#N)c2ccccc12	n1ccc(c2cc3CCC(C)(C)Nc3cc12) Oe1nc2cc(ccc2n1)c1ccccc1 C12CC3CC(CC(C3)(C1)c1ccc(cc1)C#N)C2 NC(CC(=O)c2ccccc2)	CC1=CC(=C)CCC1 CCc1cc(ccc1N)c1ccccc1 Cn1c(ccc1)C#N c1csc(c1)C#N CC1=CC(C)(C)Nc2ccc(cc12)	COc1cc(C)c(O)cc1Br S(=O)(=O)N1CCCCC(=N1)c1ccc(cc1) Oe1ccc2c(CCC3=CC(=O)CCC23C)c1 c1cc(CC2CCCC2)ccc1
Tertiary fragments (n=17)			
Type 1-1-1 (follow type 1-1 secondary fragments) (n=3, Target: Agonist)	Type 1-1-2 (n=4, Target: Antagonist)	Type 1-1-3 (n=2, Target: A-Anta)	Type 1-1-4 (n=4)
CN(C)c1ccc(C#N)c(Cl)c1 N(Cc1ccccc1)c1cccc(Cl)c1 Cc1c(C#N)cnc2c3C(CC(=O)Nc3sc12)	Cn1cc(c2ccc(cc2)C#N)c2ccccc12 n1ccc(c2cc3CCC(C)(C)Nc3cc12) Oe1nc2cc(ccc2n1)c1ccccc1 C12CC3CC(CC(C3)(C1)c1ccc(cc1)C#N)C2	NC(CC(=O)c2ccccc2) CC1=CC(=C)CCC1	CCc1cc(ccc1N)c1ccccc1 Cn1c(ccc1)C#N c1csc(c1)C#N CC1=CC(C)(C)Nc2ccc(cc12)
Type 1-1-4-1 (n=4, Target: Agonist)	Type 1-1-4-2 (n=6, Target: A-Anta)	Type 1-1-4-3 (Target: Antagonist)	Type 2-1-1 (follow type 2-1 secondary fragments) (Target: A-Anta)
CC1Nc2ccc(cc2C(C)(C)O1)c1csc(c1) CC1Nc2ccc(cc2C(C)O1) CC(C)c1ccccc1C CC(C)CO	CCCO COc1ccc(cc1) c1sc(cc1C)C c1csc(c1)C#N CC(O)c1cc(ccc1) C3CCCC3	If compounds at least contain one of the type 1-1-4 fragments while don't contain any type 1-1-4-1 and type 1-1-4-2 fragments, they will be defined as Antagonists	All compounds which meet one of the type 4-1 secondary fragments will be defined as A-Anta

16 Table 9. The hierarchy featured fragments of Retinoic Acid Receptor Alpha (RARα).

Primary fragments (n=11)			
Type 1 (n=4)	Type 2 (n=5)	Type 3 (n=2)	
<chem>C:C-C-C</chem> <chem>C:C-C=C</chem> <chem>C:C-C:C</chem> <chem>C-C:C:C-C</chem>	<chem>[#1]-C-C=C-[#1]</chem> <chem>[#1]-C=C-[#1]</chem> <chem>C-C=C-C</chem> <chem>C-C-C=C</chem> <chem>C=C-C=C</chem>	<chem>C-C-C-C-C-C-C</chem> <chem>C-C(C)-C-C</chem>	
Secondary fragments (n=21)			
Type 1-1 (follow type 1 primary fragments) (n=14)	Type 2-1 (follow type 2 primary fragments) (n=6)	Type 3-1 (follow type 3 primary fragments) (n=1)	
<chem>C#Cc2ccccc2</chem> <chem>CCc1c(Cl)cccc1</chem> <chem>c1ccc(F)cc1C</chem> <chem>CC1(C)CCCc2cc(ccc12)</chem> <chem>CC=CC(=CC)C</chem> <chem>c1ccc(O)c(c1)C12CC3CC(CC(C3)C1)C2</chem>	<chem>n1c(ccc1)c1ccc(cc1)C(=O)O</chem> <chem>NCc1cc(cc(c1)C(C)C(C)C)</chem> <chem>C(=C)c1ccc(cc1)C(=O)O</chem> <chem>CC1(C)CCC(C)(C)c2nc(cnc12)</chem> <chem>Cc1cc(N)cc(C)c1</chem> <chem>c1cccc2OCC(=Cc12)</chem> <chem>CN1CCCc2cc(ccc12)C</chem> <chem>CCCC(C)S</chem>	<chem>CC(=CC=CC(=CC(=O)O)C)C</chem> <chem>=C</chem> <chem>CC=CC(=CC(=O)O)C</chem> <chem>C1=CCCCC1C=C</chem> <chem>C=CCC=CCCCCCCCC</chem> <chem>CCCC(=O)OCCCCC</chem> <chem>CC1=C(CC)CCCC1</chem>	<chem>CCCCCCCCCCCCCCCC[N+](C)(C)C</chem>
Tertiary fragments (n=11)			
Type 1-1-1 (follow type 1-1 secondary fragments) (n=4, Target: Agonist)	Type 1-1-2 (n=1)	Type 1-1-2-1 (n=1, Target: Antagonist)	Type 1-1-2-2 (Target: A-Anta)
<chem>CC1(C)CCC(C)(C)c2nc(cnc12)</chem> <chem>n1c(ccc1)c1ccc(cc1)C(=O)O</chem> <chem>NCc1cc(cc(c1)C(C)C(C)C)</chem> <chem>CCCC(C)S</chem>	<chem>c1ccc(O)c(c1)C12CC3CC(CC(C3)C1)C2</chem>	<chem>C(=O)C=Cc1ccc(cc1)C(=O)O</chem>	If compounds at least contain one of the type 1-1-2 fragments while don't contain any type 1-1-2-1 fragments, they will be defined as A-Anta
Type 1-1-3 (n=9)	Type 1-1-3-1 (n=14, Target: Agonist)		Type 1-1-3-2 (Target: Antagonist)
<chem>C#Cc2ccccc2</chem> <chem>CCc1c(Cl)cccc1</chem> <chem>c1ccc(F)cc1C</chem> <chem>CC1(C)CCCc2cc(ccc12)</chem> <chem>CC=CC(=CC)C</chem> <chem>C(=C)c1ccc(cc1)C(=O)O</chem> <chem>Cc1cc(N)cc(C)c1</chem> <chem>c1cccc2OCC(=Cc12)</chem> <chem>CN1CCCc2cc(ccc12)C</chem>	<chem>CC(=CCO)C</chem> <chem>C=C(C)c1cc(cc(c1)C(C)C)</chem> <chem>c3ccc4cc(ccc4c3)C(=O)O</chem> <chem>c1ccccc1</chem> <chem>c1ccc(cc1)c1noc(=O)[nH]1</chem> <chem>OCc1ccc(cc1)C(=O)</chem> <chem>c2c(O)cc(O)cc2</chem>		<chem>CC1(C)CCNc2ccc(cc12)C</chem> <chem>CCc1ccc(cc1)CO</chem> <chem>C=CC(=O)O</chem> <chem>CCc1ccc(cc1)C(=O)</chem> <chem>c1ccc(F)cc1C</chem> <chem>C1CCCc2ccc(C)cc12</chem> <chem>C=Cc1ccc(cc1)C(=O)</chem>
Type 2-1-1 (follow type	Type 2-1-2 (Target:	Type 3-1-1 (follow	

2-1 secondary fragments) (n=1, Target: Antagonist)	Agonist)	type 3-1 secondary fragments) (Target: Antagonist)	
<chem>C2CCC(=C2C)CC</chem>	If compounds don't contain any type 2-1-1 fragments, they will be defined as Agonists	All compounds which meet one of the type 4-1 secondary fragments will be defined as Antagonist	

17

18

19 Table 10. The hierarchy featured fragments of Retinoic Acid Receptor Beta (RARβ).

Primary fragments (n=9)			
Type 1 (n=4)		Type 2 (n=5)	
C:C-C:C C:C-C=C C:C-C-C C-C:C:C-C		C=C-C-C C-C=C-C [#1]-C-C=C-[#1] [#1]-C=C-[#1] C=C-C=C	
Secondary fragments (n=19)			
Type 1-1 (follow type 1 primary fragments) (n=15)			Type 2-1 (follow type 2 primary fragments) (n=4)
CC=CC(=CC)C CC(=Cc1ccc(cc1)C(=O)) CNc1ccc(cc1)C(=O)O n1c(ccc1)c1ccc(cc1)C(=O)O C#Cc1ccc(cc1)	CCCCc1ccc(cc1)c1ccc(cc1) CC(=C)c1ccc2c(c1)C(C)(C)CCC2(C)C Cc1ccc(cc1)c1ccc(O)c(c1) CC1(C)CCC(C)(C)c2nc(cnc12) CC1CCCc2cc(N)ccc12	CCOc1nc(sc1C)c1ccc(cc1)C(=O)) O CC1(C)CCC(C)(C)c2cc(S)ccc12 CC=Cc1ccc2cccc2c1 c1ccc2c(c1)C(C)(C)CCC2 NC(=O)c3ccc(cc3)C(=O)O	C1CCC(C)C=C1C(=CC=CC(=C C)) C=C(C)C(=CC=C(C))C CCC(=CCCC=CCCC) CC(=CC=CC(=CC(=O))C)
Tertiary fragments (n=11)			
Type 1-1-1 (follow type 1-1 secondary fragments) (n=16, Target: Agonist)	Type 1-1-2 (Target: Antagonist)	Type 2-1-1 (follow type 2-1 secondary fragments) (n=1, Target: Agonist)	Type 2-1-2 (Target: Antagonist)
CC(=O)O Cc1cc(ccc1)C(=O) C=C(C)c1cc(cc(c1))C(C)(C) n1c(ccc1) c1ccc3cc(ccc3c1)C(=O)O C(=O)Nc1ccc(cc1) CCOc1ccc(cc1)c1ccc(cc1)C(=O)O CC1(C)CCC(C)(C)c2nc(cnc12) c1nc(sc1)c1ccc(cc1)C CC1=C(C)C(C)(C)CCC1 c1ccc(cc1)c1ccc(C)c(C)c1 c1ccc(cc1)C#Cc1ccc(cc1)C CCCCc1ccc(cc1)c1ccc(cc1) Cc1cc2c(cc1C(=NO))C(C)(C)CCC2(C)C Sc3ccc(cc3)C Cc1cc(cc(c1)C(C)(C)C)	If compounds don't contain any 1-1-1 fragments, they will be defined as Antagonists.	CC(=CC(=O)O)C	If compounds don't contain any type 2-1-1 fragments, they will be defined as Antagonists.

21 Table 11. The hierarchy featured fragments of Retinoic Acid Receptor Gamma (RARγ).

Primary fragments (n=13)			
Type 1 (n=6)	Type 2 (n=5)	Type 3 (n=2)	
<chem>C:C-C=C</chem> <chem>C:C-C:C</chem> <chem>C-C:C:C-C</chem> <chem>C:C-C-C:C</chem> <chem>C-C-C-C:C</chem> <chem>C-C:C-C-C</chem>	<chem>[#1]-C-C=C-[#1]</chem> <chem>[#1]-C=C-[#1]</chem> <chem>C-C=C-C</chem> <chem>C-C-C=C</chem> <chem>C=C-C=C</chem>	<chem>C-C(C)-C-C</chem> <chem>C-C-C-C-C</chem>	
Secondary fragments (n=17)			
Type 1-1 (follow type 1 primary fragments) (n=11)		Type 2-1 (follow type 2 primary fragments) (n=4)	Type 3-1 (follow type 3 primary fragments) (n=2)
<chem>C(=C)c1ccc(cc1)C(=O)</chem> <chem>CC=CC(=CC(=O)O)C</chem> <chem>c2cc3CCCC4CCCc(c2)c34</chem> <chem>C#Cc2ccc(cc2)C</chem> <chem>c1ccc([nH]1)c1ccc(cc1)C(=O)O</chem>	<chem>C(=O)c3ccc4cc(c3c4)CO</chem> <chem>CC1(C)CCC(C)(C)c2nc(cnc12)</chem> <chem>C=C(C)c1ccc2c(c1)C(C)(C)CCC2(C)C</chem> <chem>c1cc2CCCC3CCCCc(c1)c23</chem> <chem>CC1(C)CCC(C)(C)c2c1ccc(c2)CC(C)CCOC(=O)</chem>	<chem>CC(=CC=C)C=CC1=CCCC(C1)C</chem> <chem>CC(=CC=C(C#N))C=CC1=C(C)CCCC1(C)C</chem> <chem>CC(=CC=CC(=CC(=O)O)C)C=C</chem> <chem>CC(=CCCC(=CCCC))</chem>	<chem>CCCCCNCCCCC</chem> <chem>C1C1C(C1)C(C1)C(C1)C(C1)C1</chem>
Tertiary fragments (n=2)			
Type 1-1-1 (follow type 1-1 secondary fragments) (n=2, Target: Antagonist)	Type 1-1-2 (Target: Agonist)	Type 2-1-1 (follow type 2-1 secondary fragments) (Target: Agonist)	Type 3-1-1 (follow type 3-1 secondary fragments) (Target: Agonist)
<chem>CC(C)C=C(c1ccc(cc1)C)COCC</chem>	If compounds don't contain any type 1-1-1 fragments, they will be defined as Agonists.	All compounds which meet one of the type 4-1 secondary fragments will be defined as Agonists	All compounds which meet one of the type 4-1 secondary fragments will be defined as Agonists

22

23 Table 12. The hierarchy featured fragments of Thyroid Hormone Receptor Alpha (TRa).

Primary fragments (n=12)			
Type 1 (n=5)		Type 2 (n=7)	
N-C:C-C N-C:C-[#1] C-N:C:C	N-C-C:C N-C-C-C:C	O=C-N-C-C O=C-C-C N-C=C-[#1] N-C-C=C	N-C-C-C-C C-N-C-C-C C-C-N-C-C
Secondary fragments (n=9)			
Type 1-1 (follow type 1 primary fragments) (n=7)			Type 2-1 (follow type 2 primary fragments) (n=2)
c1csc(n1)c1ccc(c(c1)[N+](=O)[O-])S(=O)(=O)C c1c(C)cc(CC2SC(=O)NC2=O)cc1C	c1ccc2CCCc2c1Oc1ccc(O)cc1 c2cc(Cl)c(Oc3ccc(O)cc3)c(Cl)c2	Oc1ccc(Oc2ccc(C)cc2)cc1 CC(C)c1cc(nnc1=O) CC(C)c1cc(O)ccc1O	CC(C)c1cc(nnc1=O) CC(C)c1cc(O)ccc1O
Tertiary fragments (n=5)			
Type 1-1-1 (follow type 1-1 secondary fragments) (n=5, Target: Agonist)	Type 1-1-2 (Target: Antagonist)	Type 2-1-1 (follow type 2-1 secondary fragments) (Target: Agonist)	
c1ccc2CCCc2c1Oc1ccc(O)cc1 c2cc(Cl)c(Oc3ccc(O)cc3)c(Cl)c2 Oc1ccc(Oc2ccc(C)cc2)cc1 CC(C)c1cc(nnc1=O) CC(C)c1cc(O)ccc1O	If compounds don't contain any type 1-1-1 fragments, they will be defined as Antagonists.	All compounds which meet one of the type 4-1 secondary fragments will be defined as Agonists	

24

25 Table 13. The hierarchy featured fragments of Thyroid Hormone Receptor Beta (TRb).

Primary fragments (n=9)			
Type 1 (n=5)		Type 2 (n=4)	
N-C:C-C C-N-C:C N-C:C-[#1]	N-C-C:C N-C-C-C:C	O=C-C-C N-C-C=C	O=C-N-C-C N-C=C-[#1]
Secondary fragments (n=11)			
Type 1-1 (follow type 1 primary fragments) (n=9)			Type 2-1 (follow type 2 primary fragments) (n=2)
c1esc(n1)c1ccc(c(c1)[N+](=O)[O-])S(=O)(=O)C N(C)c1c(C)cc(CC2SC(=O)NC2=O)cc1C CC(C)c1cc(C)ccc1O	Oc2ccc(NC(=O)CC(=O)O)cc2 Oc1c(Cl)cc(NCC(=O))cc1Cl Oc1ccc(Oc2ccc(C)cc2)cc1	c1c(Cl)cc(cc1Cl)n1ncc(=O)[nH]c1=O Cn2ccc3c(NC(=O)C)cccc23 CC(C)c1cc(O)ccc1O	CC(C)c1cc(nnc1=O) Oc1c(Br)cc(CC)cc1Br
Tertiary fragments (n=8)			
Type 1-1-1 (follow type 1-1 secondary fragments) (n=7, Target: Agonist)	Type 1-1-2 (Target: Antagonist)	Type 2-1-1 (follow type 2-1 secondary fragments) (n=1, Target: Agonist)	Type 2-1-2 (Target: Antagonist)
CC(C)c1cc(C)ccc1O Oc2ccc(NC(=O)CC(=O)O)cc2 Oc1c(Cl)cc(NCC(=O))cc1Cl Oc1ccc(Oc2ccc(C)cc2)cc1 c1c(Cl)cc(cc1Cl)n1ncc(=O)[nH]c1=O Cn2ccc3c(NC(=O)C)cccc23 CC(C)c1cc(O)ccc1O	If compounds don't contain any type 1-1-1 fragments, they will be defined as Antagonists.	CC(C)c1cc(nnc1=O)	If compounds don't contain any type 2-1-1 fragments, they will be defined as Antagonists.

26

27 Table 14. The hierarchy featured fragments of Vitamin D Receptor (VDR).

Primary fragments (n=15)			
Type 1 (n=8)		Type 2 (n=7)	
C-C-C-O-[#1]	O-C-C=C	C-C-C-C-C	C=C-C-C
O-C-C-C-C	C=C-C-O-[#1]	C=C-C=C	C-C=C-C
O-C-C-C-O	O-C-C-C=C	C=C-C-C-C	C-C(C)-C-C
[#1]-C-O-[#1]	OC1CC(O)CCC1	[#1]-C-C=C-[#1]	
Secondary fragments (n=25)			
Type 1-1 (follow type 1 primary fragments) (n=16)		Type 2-1 (follow type 2 primary fragments) (n=9)	
CC=C3CC(O)CCC3=C C(CCO)CC1CCC2CCCC12C Cc1cc(ccc1)C(O)(CC) CCC(=C(c1ccc(O)cc1)) CC(COC)(COCC)COCC c1ccc(cc1)c1cccc(c1)C CC=C3CC(O)CC(O)C3 CCC(CC)(c1ccc(OCCC)c(C)c1)	Cc1[nH]e2cccc2c1C(Nc1cccc1) CCCCC(C)C1CCC2C(O)CCCC12C CC[Si](CC)(c1ccc(OCCC)c(C)c1) Cc1ccc(C=CC(C))cc1 Oc1cc(C=CC(=CCC))ccc1 CCS(=O)(=O)c1oc(nc1S(=O)(=O))c1cc cs1 CN(C)c1ccc2c(C)cc(=O)oc2c1 CCC(=O)NCCS(=O)(=O)	C(N)c1c[nH]e2cccc12 CC(C)CC(C)N(c1cccc1) n1cccc1=O CCCCCCCCCCC[N+](C)(C) C CCCCCCCCCCCn1cc[n+](C) c1	FC(F)(F)c1c(Cl)ccc(N)c1 c1ccc(cc1)C(=O)CC1 Sc1cc2c(Sc3cccc3N2CC)c c1 c1cc2nc3cccc3cc2cc1
Tertiary fragments (n=8)			
Type 1-1-1 (follow type 1-1 secondary fragments) (n=4, Target: Antagonist)	Type 1-1-2 (n=6, Target: Agonist)	Type 1-1-3 (n=2)	Type 1-1-3-1 (n=3, Target: Antagonist)
Cc1[nH]e2cccc2c1C(Nc1cccc1) CCS(=O)(=O)c1oc(nc1S(=O)(=O))c1cc cs1 CCC(=O)NCCS(=O)(=O) CC(COC)(COCC)COCC	CCC(CC)(c1ccc(OCCC)c(C)c1) CCCCC(C)C1CCC2C(O)CCCC12C CC[Si](CC)(c1ccc(OCCC)c(C)c1) CN(C)c1ccc2c(C)cc(=O)oc2c1 Cc1ccc(C=CC(C))cc1 Oc1cc(C=CC(=CCC))ccc1	CC=C3CC(O)CCC3=C Cc1cc(ccc1)C(O)(CC)	CC(CC1CC(=C)C(=O)O1) CC(CC1CC(C)(O)C(=O)N1 C) OC5CC(N)C(O)C(C)O5
Type 1-1-3-2 (Target: Agonist)	Type 1-1-4 (n=4)	Type 1-1-4-1 (n=3, Target: Antagonist)	Type 1-1-4-2 (n=2, Target: A-Anta)
If compounds at least contain one of the type 1-1-3 fragments while don't contain any type 1-1-3-1 fragments, they will be defined as Agonists.	C(CCO)CC1CCC2CCCC12C CCC(=C(c1ccc(O)cc1)) c1ccc(cc1)c1cccc(c1)C CC=C3CC(O)CC(O)C3	c1ccc(cc1)C(=O)O CC=C3CC(O)C(=CCF)C(O)C3 Cc3cccc(O)c3	C(CCO)C(=O) CCC(CC(CC)CC)C(C)
Type 1-1-4-3 (Target: Agonist)			
If compounds at least contain one of the type 1-1-4 fragments while don't contain any type 1-1-4-1 and type 1-1-4-2 fragments, they will be defined as Agonists			

29 Table 15. The performance of NRMEA

Active/Inactive prediction ^a								
	TP	FP	TN	FN	Precision	Sensitivity	Specificity	Accuracy
ERa- Training Set	700	23	1156	83	0.97	0.89	0.98	0.95
ERa- Test Set	172	13	284	27	0.93	0.86	0.96	0.92
AR- Training Set	663	14	1573	35	0.98	0.95	0.99	0.98
AR- Test Set	161	9	391	12	0.95	0.93	0.98	0.96
ERb- Training Set	439	33	1594	18	0.93	0.96	0.98	0.98
ERb- Test Set	98	21	479	17	0.82	0.85	0.96	0.94
GR- Training Set	641	12	1482	51	0.98	0.93	0.99	0.97
GR- Test Set	163	11	363	13	0.94	0.93	0.97	0.96
MR- Training Set	354	4	1990	29	0.99	0.92	0.997	0.99
MR- Test Set	90	1	498	7	0.99	0.93	0.997	0.99
PR- Training Set	378	1	1940	24	0.997	0.94	0.999	0.98
PR- Test Set	91	0	486	11	1	0.90	1	0.98
RARa- Training Set	278	29	2357	219	0.91	0.56	0.99	0.91
RARa - Test Set	58	14	583	68	0.81	0.46	0.98	0.89
RARb - Training Set	235	3	2514	136	0.99	0.63	0.998	0.95
RARb - Test Set	59	2	626	36	0.97	0.62	0.997	0.95
RARg- Training Set	199	10	2538	102	0.95	0.66	0.996	0.96
RARg - Test Set	50	7	630	28	0.88	0.64	0.99	0.95
TRa- Training Set	91	1	5443	10	0.99	0.90	0.999	0.998
TRa - Test Set	22	2	1355	4	0.92	0.85	0.998	0.995
TRb - Training Set	101	6	5438	21	0.94	0.83	0.998	0.995
TRb - Test Set	22	4	1353	9	0.85	0.71	0.997	0.99
VDR - Training Set	452	32	5241	276	0.93	0.62	0.99	0.95
VDR - Test Set	116	17	1301	70	0.87	0.62	0.99	0.94
Agonist/A-Anta/Antagonist prediction								
	Accuracy-Training Set				Accuracy-Test Set			
ERa	0.88				0.84			
AR	0.88				0.87			
ERb	0.92				0.98			
GR	0.87				0.80			
MR	0.96				0.92			
PR	0.80				0.74			
RARa	0.94				0.90			
RARb	0.96				0.93			
RARg	0.96				0.92			
TRa	0.99				0.95			
TRb	0.98				1			
VDR	0.94				0.91			

30 α : TP: True Positive; FP: False Positive; TN: True Negative; FN: False Negative.