Supporting Information

Documenting and Harnessing the Biological Potential of Molecules in Distributed Drug Discovery (D3) Virtual Catalogs

Milata M. Abraham[1] | Ryan E. Denton[1] | Richard W. Harper[1] | William L. Scott[1], | Martin J. O'Donnell[1,*] | Jacob D. Durrant[2,*]

[1]Department of Chemistry and Chemical Biology, Indiana University Purdue University Indianapolis, Indianapolis, Indiana 46202, United States

[2]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, 15260, United States

Correspondence: Jacob D. Durrant, durrantj@pitt.edu; Martin J. O'Donnell, modonnel@iupui.edu

Table of Contents

Workflow Diagrams

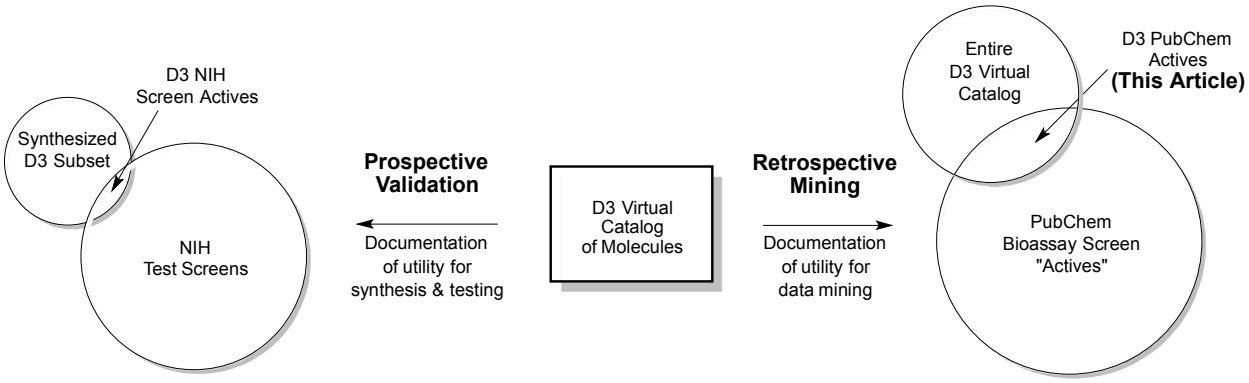


Figure SI-1. Two approaches for demonstrating the potential for biological activity of D3 virtual catalog members.

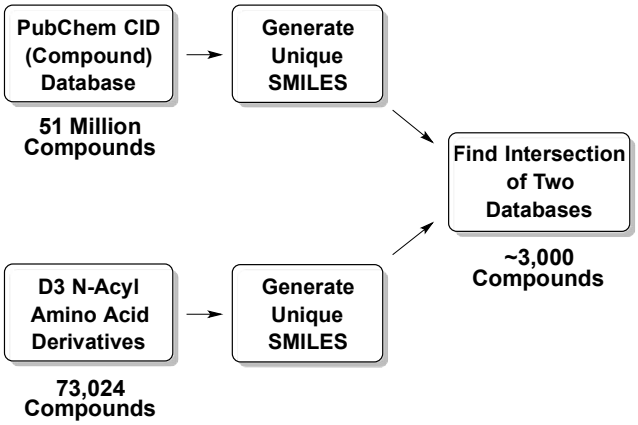


Figure SI-2. The chemoinformatics analysis. Common D3 virtual and bioactive PubChem compounds.

Stereochemical Issues in the Enumeration of N-Acyl α-Amino Acid Derivatives

Enumeration with the 100 electrophiles (alkyl halides and Michael acceptors) and 100 carboxylic acids gave 24,416 N-acyl amino acids **1** [1], 24,192 N-acyl amino acid methyl esters **2** [2], and 24,416 N-acyl amino acid amides **3** (presented here for the first time [3] (Figure SI-3). The reason these numbers exceed 10,000 (100 electrophiles x 100 carboxylic acids) is discussed below (taken in part from reference [1], Supporting Information, p. 4).

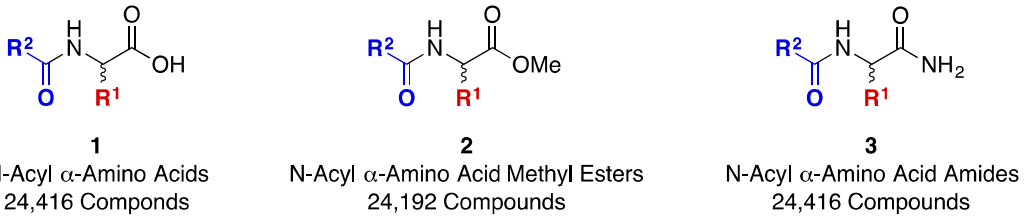


Figure SI-3 (**Figure 1** from article)**.** Generic structures of the 73,024 unnatural and natural N-acyl α-amino acid derivatives in the D3 virtual catalog

Issues of stereochemistry were addressed in this enumeration. Examples illustrating possible N-acylated amino acid products **1** are shown in Figure SI-4 The products are identified in three ways: row/column (e.g. **Aa**), Collaborative Drug Discovery (CDD) number (e.g. DDD-000002150), and the number used in our *J. Comb. Chem.* paper [e.g. (*R*)-{9}119] [1, 2]. The diversity elements (reactants R^1X and R^2CO_2H) include achiral (**a** and **A**), optically pure (**B**), racemic (**b** and **C**), and prochiral (**c**) examples. Only one of the possible stereoisomers for each product is shown in this figure.

(1) When achiral or optically pure reactants (R^1X and R^2CO_2H) were used, racemic products (2 stereoisomers) resulted. Examples: **Aa** and its enantiomer or **Ba** and its diastereomer at the α carbon. The database contains 15,308 unique molecules **1** (7,654 pairs) of this class.

(2) When one of the reactants (R^1X and R^2CO_2H) was racemic or, in the case of Michael acceptors, prochiral, 4 stereoisomers were obtained. Examples: **Ab** or **Ac** and their stereoisomers. Other cases yielding 4 stereoisomers: optically pure and racemic (**Bb**); optically pure and prochiral (**Bc**); or racemic and achiral (**Ca**). The database contains 8,068 such unique molecules **1** (2,017 sets of four stereoisomers) of this class.

(3) When both reactants (R^1X and R^2CO_2H) were racemic or prochiral and racemic, 8 stereoisomers were formed. Examples: **Cb** or **Cc** and their stereoisomers. The database contains 1,040 such unique molecules **1** (130 sets of eight stereoisomers) of this class.

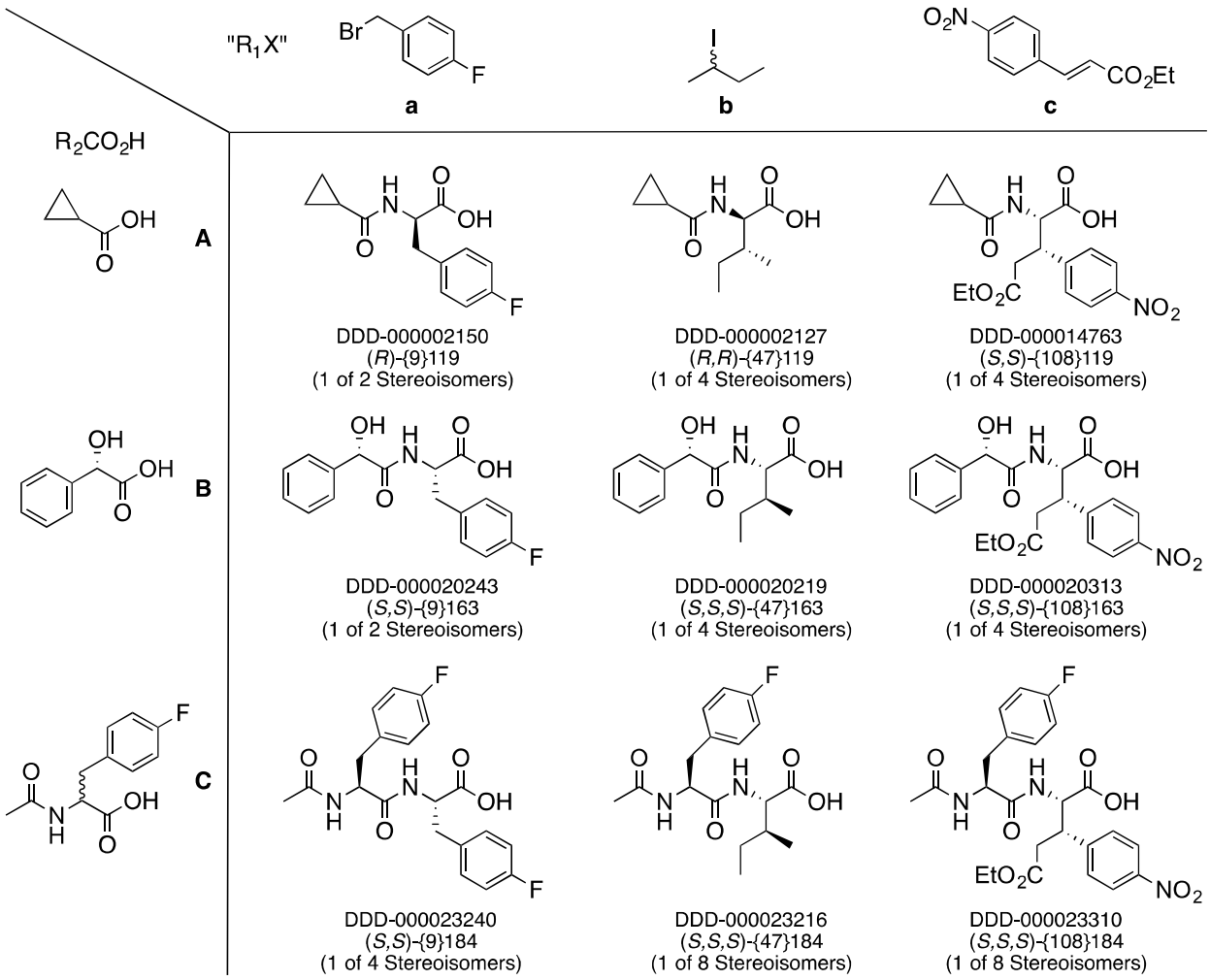


Figure SI-4. Sample enumerated compounds demonstrating different stereochemical issues.

Identifying Unique D3 Compounds Listed as Actives in PubChem

A series of Boolean substructure searches was used to identify all D3-like N-acyl α-amino acid derivatives (**A**, **B**, or **C**) and N-acyl dipeptide derivatives (**D**, **E**, or **F**) present in PubChem (i.e., "PC-CID" compounds in Supporting Information Figures SI-3, SI-4, SI-5, and SI-6). Current D3 catalog members do not contain glycine or α,α-disubstituted compounds, so the α-positions of substructures **A**-**F** were substituted with both an α-carbon and an α-hydrogen. Since the maximum molecular weight of the D3 catalog compounds is 685 daltons, the PubChem searches were limited to compounds with molecular weights less than 750 daltons. Low-weight dipeptides were not explicitly excluded from this search because they are accessible through existing D3 chemistry, when α-amino acid derivatives are used as N-acylating agents (R^2CO_2H).

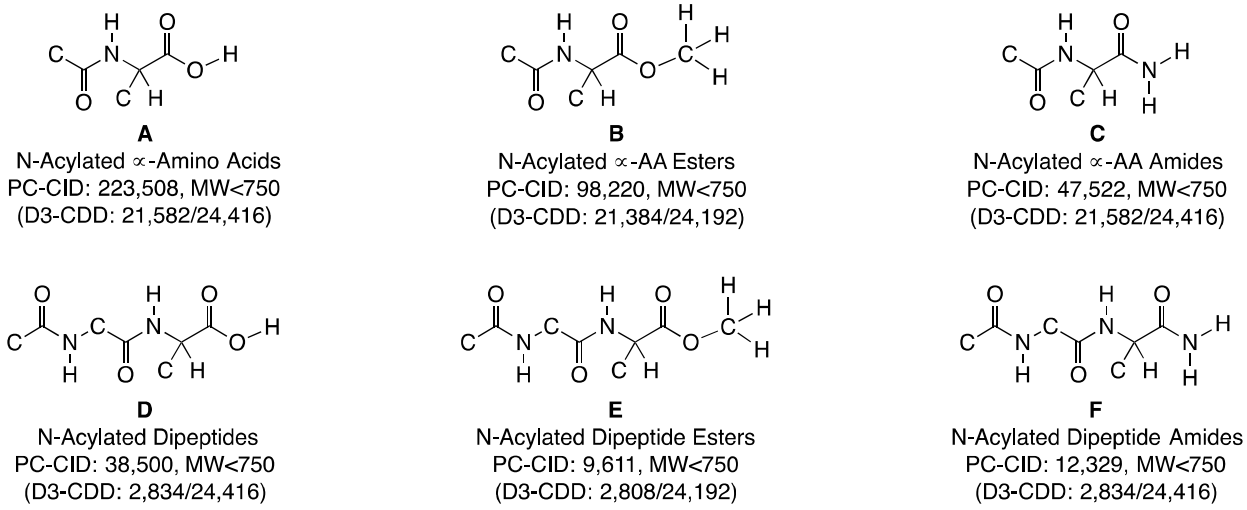
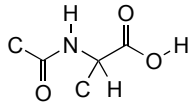


Figure SI-5. PubChem and D3-catalog substructure searches refined by subsequent Boolean searches to identify *only* N-acyl α-amino acid derivatives **A**-**C** and N-acyl dipeptide derivatives **D**-**F**. PC-CID is the number of PubChem compounds that meet the search parameters. D3-CDD indicates the fraction of D3 compounds (out of the total number of D3 compounds with the specified substructure) in the Collaborative Drug Discovery/D3 dataset (see the following Figures SI-6, SI-7, and SI-8 for Boolean search strategies).

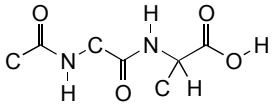
Substructure SMILES & Boolean Searches - Acids

Step 1
Substructures
SMILES
PubChem Searches



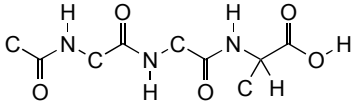
G

N-acylated ∝-amino acids (10/16/14):
[C]C(N([H])C([C])([H])C(O[H])=O)=O
422,803, MW<750



H

N-acylated dipeptides (10/16/14):
[H]OC(C([C])([H])N([H])C([C]N([H])C([C])=O)=O)=O
199,295, MW<750



I

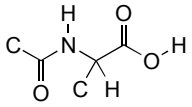
N-acylated tripeptides: (10/16/14):
[H]OC(C([C])([H])N([H])C([C]N([H])C([C]N([H])C([C])=O)=O)=O)=O
160,795, MW<750

Step 2
Boolean Searches/Calculations

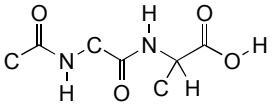
A = Only N-acylated ∝-amino acids:
A = G NOT H
(422,803 - 199,295) 223,508, MW<750

D = Only N-acylated dipeptides:
D = H NOT I
(199,295 - 160,795) 38,500, MW<750

Step 3
Search Results



A
N-Acylated ∝-Amino Acids
PC-CID: 223,508, MW<750
(D3-CDD: 21,582/24,416)



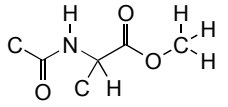
D
N-Acylated Dipeptides
PC-CID: 38,500, MW<750
(D3-CDD: 2,834/24,416)

Results that appear
Figure SI-3

Figure SI-6. Boolean substructure searches using SMILES strings identified all D3-like N-acyl α-amino acids **A** and N-acyl dipeptides **D** in PubChem (Figure SI-5). Current D3 catalog members do not contain glycine or α,α-disubstituted compounds, so the α-positions of substructures **A** and **D** were substituted with both an α-carbon and an α-hydrogen. Since the maximum molecular weight of the D3 catalog compounds is 685 daltons, the PubChem searches were limited to compounds with molecular weights less than 750 daltons.

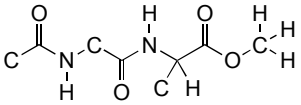
Substructure SMILES & Boolean Searches - Esters

Step 1
Substructures
SMILES
PubChem Searches



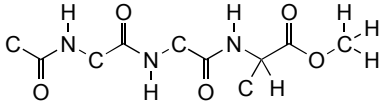
J

N-acylated ∝-AA Me Esters (11/14/14):
[C]C(N([H])C([C])([H])C(OC([H])([H])[H])=O)=O
109,866, MW<750



K

N-acylated dipeptide Me Esters (11/14/14):
O=C(N([H])C([C])([H])C(OC([H])([H])[H])=O)[C]N([H])C([C])=O
11,646, MW<750



L

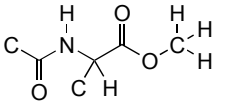
N-acylated tripeptide Me Esters: (11/14/14):
O=C(N([H])C([C])([H])C(OC([H])([H])[H])=O)[C]N([H])C([C]N([H])C([C])=O)=O
2,035, MW<750

Step 2
Boolean Searches/Calculations

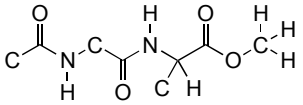
B = Only N-acylated ∝-AA esters:
B = J NOT K
(109,866 - 11,646) 98,220, MW<750

E = Only N-acylated dipeptide esters:
E = K NOT L
(11,646 - 2,035) 9,611, MW<750

Step 3
Search Results



B
N-Acylated ∝-AA Esters
98,220, MW<750
(D3-CDD: 21,384/24,192)



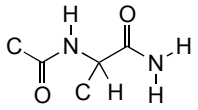
E
N-Acylated Dipeptide Esters
9,611, MW<750
(D3-CDD: 2,808/24,192)

Results that appear
Figure SI-3

Figure SI-7. Boolean substructure searches using SMILES strings were conducted to identify all D3-like N-acyl α-amino acids methyl esters **B** and N-acyl dipeptide methyl esters **E** in PubChem (Figure SI-5). Current D3 catalog members do not contain glycine or α,α-disubstituted compounds, so the α-positions of substructures **B** and **E** were substituted with both an α-carbon and an α-hydrogen. Since the maximum molecular weight of the D3 catalog compounds is 685 daltons, the PubChem searches were limited to compounds with molecular weights less than 750 daltons.

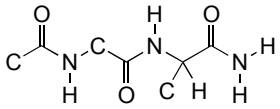
Substructure SMILES & Boolean Searches - Amides

Step 1
Substructures
SMILES
PubChem Searches



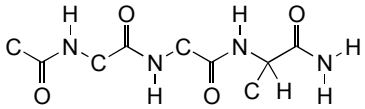
M

N-acylated ∝-AA Me amides (11/14/14):
[C]C(N([H])C([C])([H])C(N([H])[H])=O)=O
68,883, MW<750



N

N-acylated dipeptide Me amides (11/14/14):
[H]N([H])C(C([C])([H])N([H])C([C]N([H])C([C])=O)=O)=O
21,361, MW<750



O

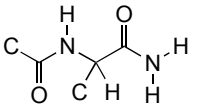
N-acylated tripeptide amides: (11/14/14):
[H]N([H])C(C([C])([H])N([H])C([C]N([H])C([C]N([H])C([C])=O)=O)=O)=O
9,032, MW<750

Step 2
Boolean Searches/Calculations

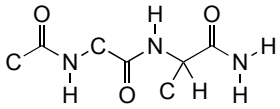
C = Only N-acylated ∝-AA amides:
C = M NOT N
(68,883 - 21,361) 47,522, MW<750

F = Only N-acylated dipeptide amides:
F = N NOT O
(21,361 - 9,032) 12,329, MW<750

Step 3
Search Results



C
N-Acylated ∝-AA Amides
47,522, MW<750
(D3-CDD: 21,582/24,416)



F
N-Acylated Dipeptide Amides
12,329, MW<750
(D3-CDD: 2,834/24,416)

Results that appear
Figure SI-3

Figure SI-8. Boolean substructure searches using SMILES strings were conducted to identify all D3-like N-acyl α-amino acid amides **C** and N-acyl dipeptide methyl amides **F** in PubChem (Figure SI-5). Current D3 catalog members do not contain glycine or α,α-disubstituted compounds, so the α-positions of substructures **C** and **F** were substituted with both an α-carbon and an α-hydrogen. Since the maximum molecular weight of the D3 catalog compounds is 685 daltons, the PubChem searches were limited to compounds with molecular weights less than 750 daltons.

This limited set of D3-like PubChem compounds was next searched for exact D3 matches. A list of all D3 compounds can be obtained from either the Collaborative Drug (CDD) or Durrant Lab websites (https://www.collaborativedrug.com [4, 5], accessed February 27, 2017; or http://durrantlab.com/liglib/iupui/d3_docking/, accessed February 27, 2017). Included in this comparison were the 2,834, 2,808, and 2,834 dipeptide derivatives present in the "D3-CDD" N-acyl dipeptide (**D**), D3 N-acyl dipeptide methyl ester (**E**), and D3 N-acyl dipeptide amide (**F**) sets, respectively.

As PubChem does not consistently report chirality, structural information related to the stereochemistry of both the PubChem and D3 compounds was ignored for the purpose of this comparison. Since epimeric pairs are always present in the D3 catalog, ignoring chirality did not impact the total number of compounds ultimately identified. By all these search criteria, and ignoring chirality, ~3,000 PubChem compounds were present in the D3 virtual catalog, excluding those that had been specifically uploaded by D3 or CDD (June 2014).

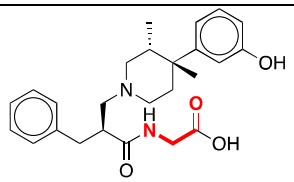
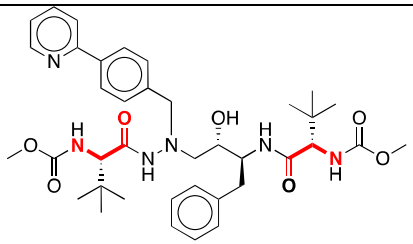
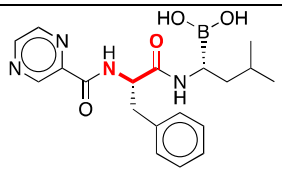
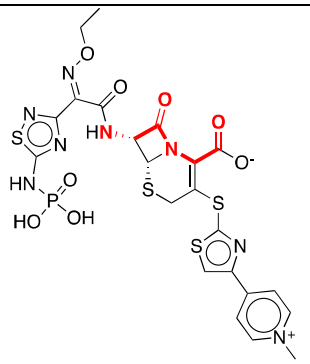
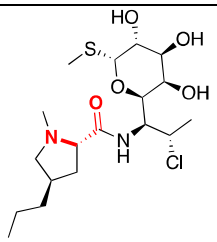
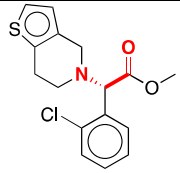
PubChem substructure searches were performed using the SMILES strings derived from substructures **1** - **3** (Figure SI-3) "CC(N([H])CC(O[H])=O)=O" (**1**), "CC(N([H])CC(OC([H])([H])[H])=O)=O" (**2**), and "CC(N([H])CC(N([H])[H])=O)=O" (**3**). The SMILES strings of these compounds were downloaded from the PubChem servers. Structural information related to stereochemistry was removed by applying the regex expressions "s/@[@H]*]/]/g" and "s/@//g" to the SMILES strings. If appropriate, isotopic information and explicitly specified heavy hydrogen atoms were removed using the "s/\[[0-9]*/[/g", "s/\[H]//g", and "s/\[\(.\)]/\1/g" expressions. To ensure consistent charge assignments, nitrogen, carbon, oxygen, sulfur, and phosphorus atoms with non-neutral formal charges were converted to simple atoms with charge unspecified, using the "s/\[NH[0-9]?[+-]]/N/g", "s/\[N-]/N/g", "s/\[n-]/n/g", "s/\[CH[0-9]?[+-]]/C/g", "s/\[C-]/C/g", "s/\[C+]/C/g", "s/\[c+]/c/g", "s/\[c-]/c/g", "s/\[OH[0-9]?[+-]]/O/g", "s/\[O-]/O/g", "s/\[S-]/S/g", "s/\[S+]/S/g", "s/\[SH[0-9]?[+-]]/S/g", and "s/\[P+]/P/g" expressions.

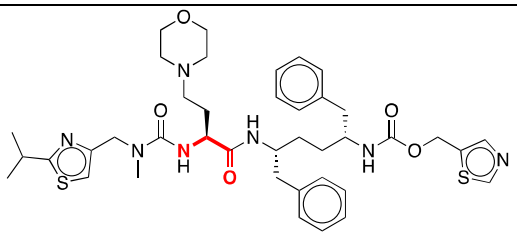
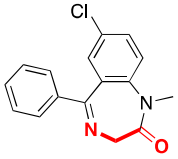
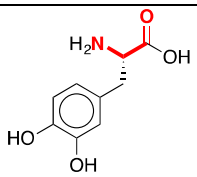
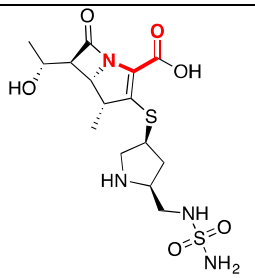
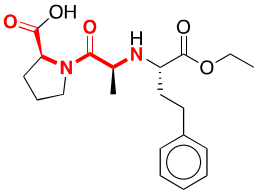
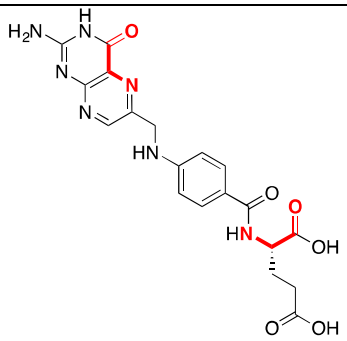
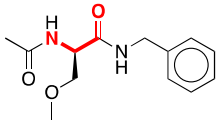
As a given compound can have multiple valid SMILES representations, we used the Open Babel computer program [6] to convert each SMILES string into a unique canonical form. Open Babel was also used to remove hydrogen atoms, convert dative bonds (e.g. [N+]([O-])=O to N(=O)=O), and discard all but the largest molecule associated with each

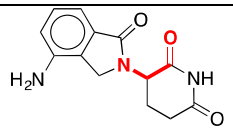
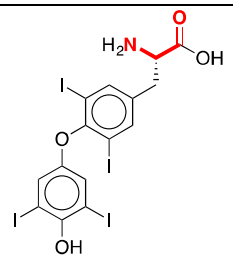
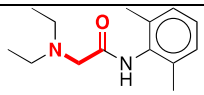
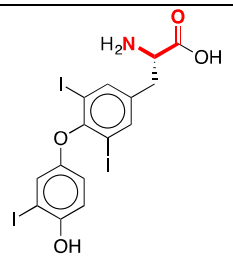
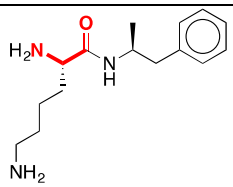
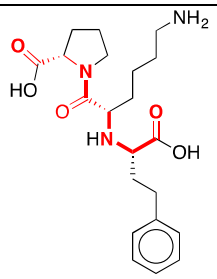
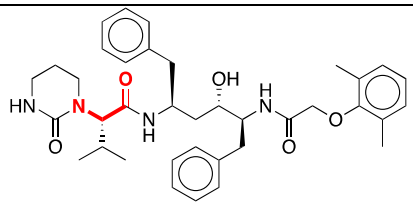
string. The N-acyl α-amino acids, esters, and amides of the D3 enumerated virtual catalog were subjected to the same protocol.

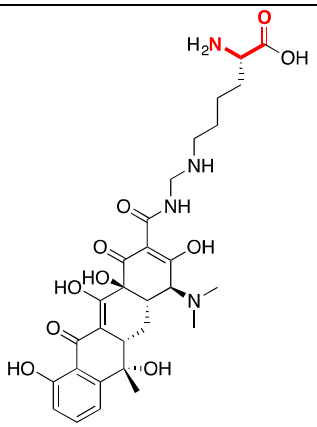
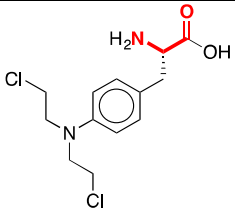
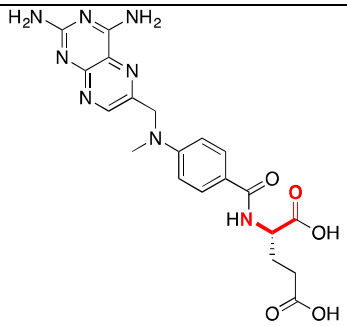
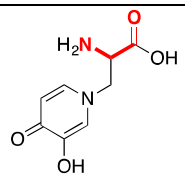
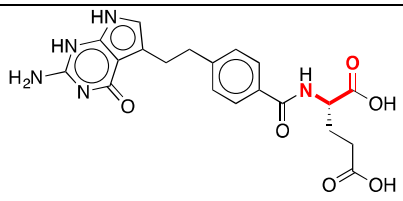
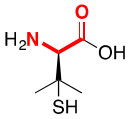
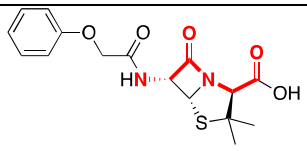
Since there is only one valid canonical SMILES string for a given compound, SMILES strings in this form serve as unique molecular identifiers and so can be used to identify compounds common to both the PubChem and D3 lists. We used the Unix command line for this purpose, but a similar protocol is possible with popular spreadsheet programs. For example, the unique SMILES strings in the PubChem and D3 lists could be identified using Microsoft Excel's advanced data filter ("Unique records only"). The lists of unique PubChem and D3 SMILES strings could then be combined into a single column, and the COUNTIF function could be used to identify any entries that appear more than once and must therefore be common to both lists. Computer programs such as MONA [7] and ChemCom [8] may also facilitate this type of molecular searching strategy.

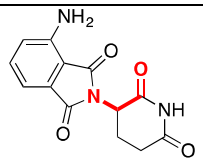
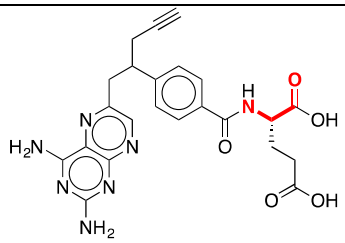
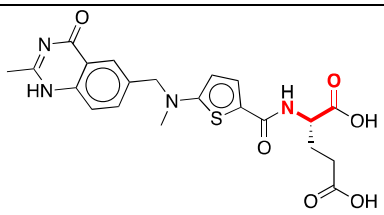
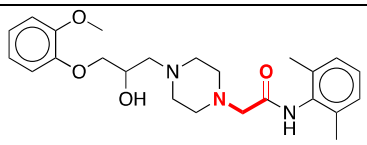
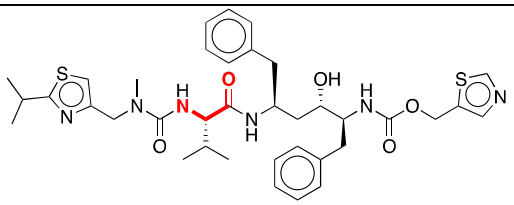
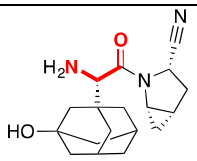
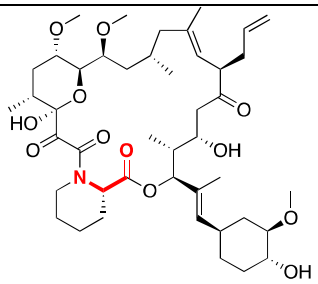
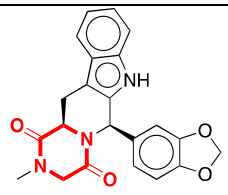
Table SI-1. 40 approved drugs containing substructures of various D3-like α-amino acid and dipeptide derivatives. The drug name (with reference), structure, classification, and therapeutic indication (adapted from DrugBank [9]) are shown.

#	Compound	Structure	Classification	Indication
1	Alvimopan [10]		N-Acyl α-Amino Acid / Synthetic	Gastrointestinal recovery following bowel resection surgery with primary anastomosis
2	Atazanavir [11]		N-Acyl α-Amino Acid Amide / Synthetic	HIV-1
3	Bortezomib [12]		N-Acyl Dipeptide Boronic Acid / Synthetic	Multiple myeloma
4	Ceftaroline Fosamil [13]		N-Acyl Dipeptide (beta-Lactam) / Synthetic	Bacterial infections
5	Clindamycin [14]		N,N-Dialkyl Amino Acid Amide / Synthetic	Anaerobic bacteria
6	Clopidogrel (Plavix) [15]		N,N-Dialkyl Amino Acid Ester / Synthetic	Atherosclerotic events

7	Cobicistat [16]		N-Acyl α-Amino Acid Amide / Synthetic	HIV
8	Diazepam (Valium) [17]		Imino Acid Amide / Synthetic	Anxiety disorders, insomnia, anticonvulsant, alcohol withdrawal syndrome
9	L-DOPA [18]		α-Amino Acid / Natural Product	Parkinsonism and dopa-responsive dystonia
10	Doripenem [19]		N-Acyl α-Amino Acid (beta-Lactam) / Synthetic	Antibiotic
11	Enalapril [20]		N-Alkyl Dipeptide, N-Alkyl Amino Acid Ester / Synthetic	Hypertension, congestive heart failure
12	Folic Acid [21]		N-Acyl α-Amino Acid / Natural Product	Vitamin B_9
13	Lacosamide [22]		N-Acyl α-Amino Acid Amide / Synthetic	Epilepsy

14	Lenalidomide [23]		N-Acyl α-Amino Acid Imide / Synthetic	Multiple myeloma, some forms of transfusion-dependent anemia
15	Levothyroxine [24]		α-Amino Acid / Natural Product	Hypothyroidism, goiter, chronic lymphocytic thyroiditis, myxedema coma, and stupor
16	Lidocaine [25]		N,N-Dialkyl α-Amino Acid / Synthetic	Anesthesia
17	Liothyronine [26]		α-Amino Acid / Synthetic	Hypothyroidism
18	Lisdexamfetamine [27]		α-Amino Acid Amide / Synthetic	Attention Deficit/Hyperactivity Disorder
19	Lisinopril [28]		N-Alkyl Dipeptide, N-Alkyl α-Amino Acid / Synthetic	Hypertension, congestive heart failure
20	Lopinavir [29]		N-Acyl-Type α-Amino Acid Amide / Synthetic	HIV

21	Lymecycline [30]		α-Amino Acid / Synthetic	Acne, some bacterial infections
22	Melphalan [31]		α-Amino Acid / Synthetic	Multiple myeloma, epithelial carcinoma of the ovary
23	Methotrexate [32]		N-Acyl α-Amino Acid / Synthetic	Some cancers
24	Mimosine [33]		α-Amino Acid / Natural Product	Some cancers
25	Pemetrexed [34]		N-Acyl α-Amino Acid / Synthetic	Malignant pleural mesothelioma, non-small cell lung cancer
26	Penicillamine [35]		α-Amino Acid / Synthetic	Wilson's disease, cystinuria, rheumatoid arthritis
27	Penicillin V [36]		N-Acyl Dipeptide (beta-Lactam) / Synthetic	Bacterial infections

28	Pomalidomide [37]		N-Imido α-Amino Acid Imide / Synthetic	Multiple myeloma
29	Pralatrexate [38]		N-Acyl α-Amino Acid / Synthetic	Peripheral T-cell lymphoma
30	Raltitrexed [39]		N-Acyl α-Amino Acid / Synthetic	Malignant neoplasm of colon and rectum
31	Ranolazine [40]		N,N-Dialkyl α-Amino Acid Amide / Synthetic	Chronic angina
32	Ritonavir [29]		α-Amino Acid Amide / Synthetic	HIV
33	Saxagliptin [41]		α-Amino Acid Amide / Synthetic	Type 2 diabetes mellitus
34	Tacrolimus [42]		N-Acyl α-Amino Acid Ester / Synthetic	Organ transplantation, immunosuppression
35	Tadalafil (Cialis) [43]		Cyclic Dipeptide / Synthetic	Erectile dysfunction

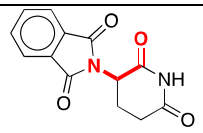
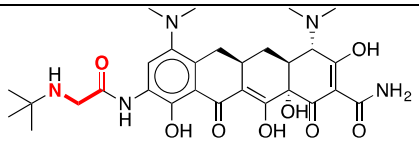
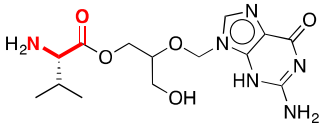
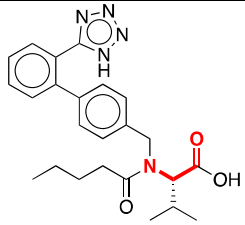
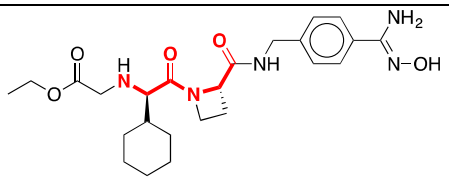
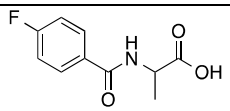
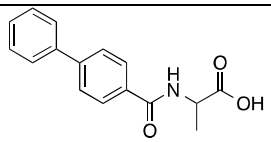
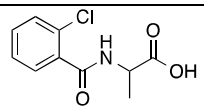
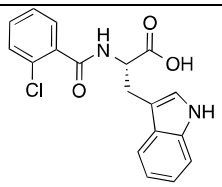
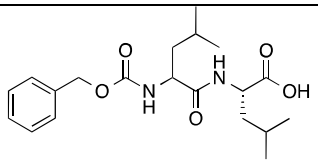
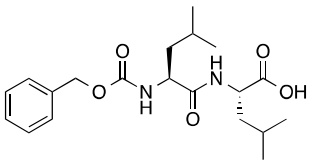
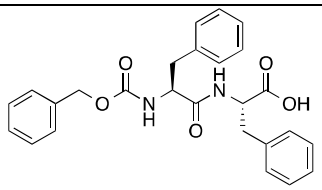
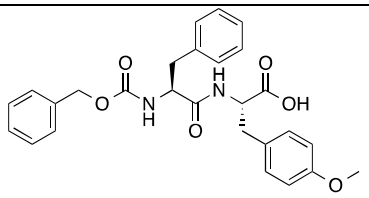
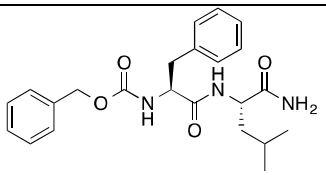
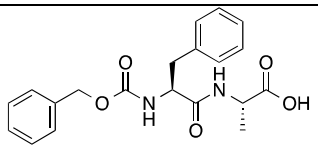
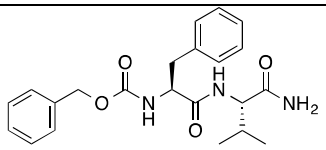
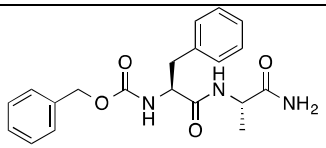
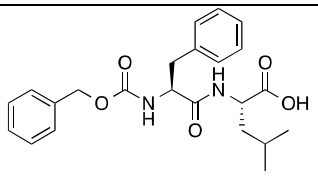
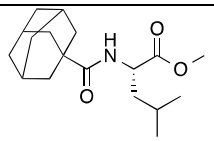
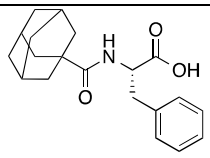
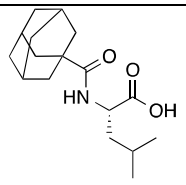
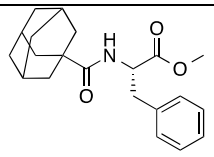
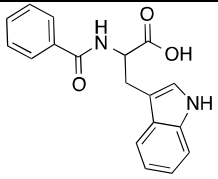
36	Thalidomide [44]		N-Imido α-Amino Acid Imide / Synthetic	Erythema nodosum leprosum
37	Tigecycline [45]		N-Alkyl α-Amino Acid Amide / Synthetic	Bacterial infections
38	Valganciclovir [46]		α-Amino Acid Ester / Synthetic	Cytomegalovirus infections
39	Valsartan [47]		N-Acyl α-Amino Acid / Synthetic	Hypertension, left ventricular hypertrophy, diabetic nephropathy, other cardiac problems
40	Ximelagatran [48]		N-Alkyl Dipeptide Amide / Synthetic	Deep vein thrombosis

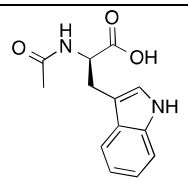
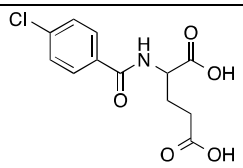
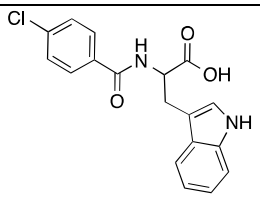
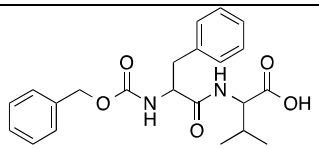
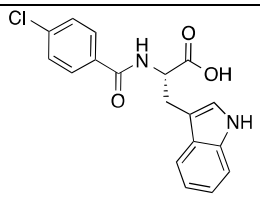
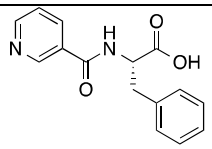
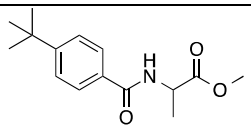
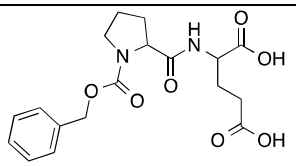
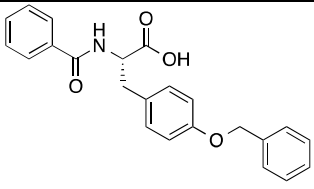
Table SI-2. 32 protein drug targets with orthosteric pockets known to bind at least one D3 α-amino acid derivative.

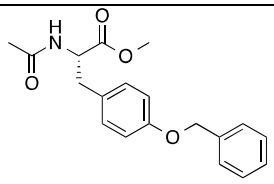
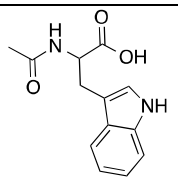
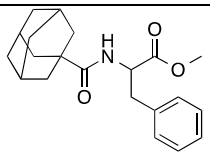
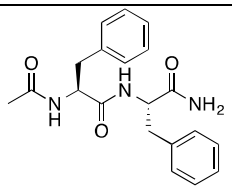
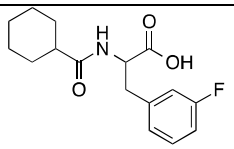
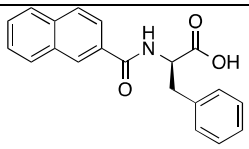
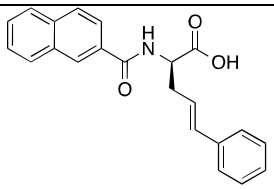
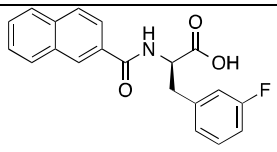
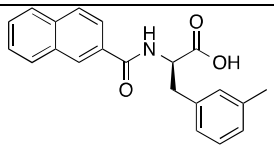
#	Target	Relevant Medical Conditions	Known Ligand(s)
1	Histone deacetylase	Psychiatric and neurological conditions [49]; cancer [50]	α-Amino acid (acetylated lysine residues [51])
2	Histone lysine methyltransferase G9a	Cancer [52]	α-Amino acid (histone tail lysine residues [53])
3	Methionine sulfoxide reductase A	Aging [54]	α-Amino acid (methionine [55])
4	Dopamine receptor	Parkinson's disease; dopa-responsive dystonia; psychiatric conditions	α-Amino acid metabolite (dopamine)
5	Luciferase	None	α-Amino acid metabolite (luciferin, from cysteine [56])
6	Thioredoxin glutathione reductase	Schistosomiasis [57]	Peptide (glutathione) and protein (thioredoxin) [58]
7	Oxytocin receptor	Labor induction [59]	Peptide (oxytocin) [60]
8	GLP-1 receptor	Diabetes mellitus type 2 [61]	Peptide (peptide hormone glucagon-like peptide-1, exendin-4, etc. [62])
9	NK1 receptor	Chemotherapy-induced nausea [63]	Peptide [64]
10	Polo-like kinase 1 - polo-box domain	Cancer [65]	Peptides [66]
11	Ribosomal peptidyl transferase center	Bacterial infections	Peptides [67]
12	MAS-related GPR member X1 receptor	Chronic pain [68]	Peptides [69]
13	Mitogen-activated protein kinase kinase kinase 3 isoform 1	Cancer [70]	Protein [71]
14	Peptidylprolyl cis/trans isomerase, NIMA-interacting 1	Cancer [72]	Protein residues [72]
15	Protein arginine deiminase 4	Rheumatoid arthritis [73]	Proteins (arginines) [74]
16	Ras-converting enzyme	Cancer [75]	Proteins [76]
17	Factor XIa	Hemophilia C [77]	Proteins and peptides [78]
18	Matrix metalloproteinase 2	Cancer [79]	Proteins and peptides [51]
19	Calpain	Alzheimer's disease, stroke, amyotrophy, muscular dystrophy, motor	Proteins and peptides [80-82]

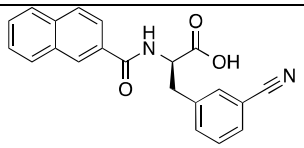
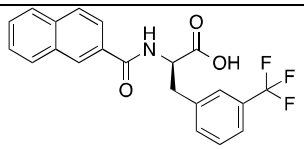
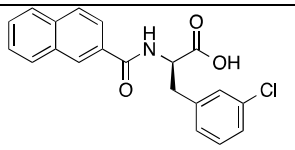
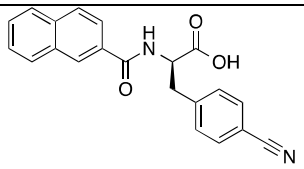
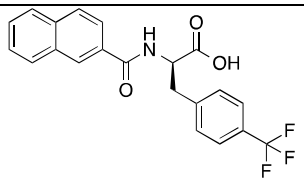
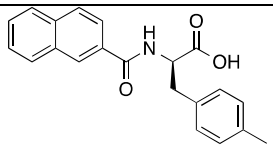
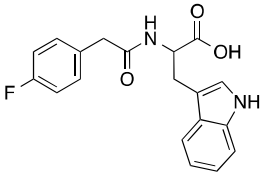
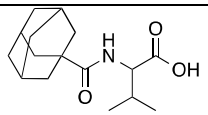
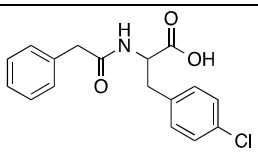
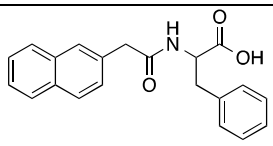
20	Cathepsin	Cancer [83], neurodegenerative disorders, rheumatoid arthritis [84]	Proteins and peptides [83]
21	Platelet-activating factor acetylhydrolase 1b, catalytic subunit 2	Atherosclerosis [85]	1-O-alkyl-2-acetyl-sn-glycero-3-phosphocholine (PAF or platelet-activating factor [86])
22	Calcium-activated chloride channels (TMEM16A)	Diarrhea [87]	Cations (Calcium)
23	Renal outer medullary K+ channel	Bartter syndrome [88]	Cations (Potassium)
24	Apoptotic protease activating factor 1	Cancer [89]	Nucleic acid (ATP) and various protein partners [90]
25	Heat shock factor-1	Cancer [91]	Nucleic acid (DNA, though the α-amino acid glutamine may increase HSF1 transcription) [92]
26	Shiga toxin	Dysentery [93]	Nucleic acids [94]
27	1,4-dihydroxy-2-naphthoyl-CoA synthase (MenB)	Bacterial infections [95]	O-succinylbenzoyl-CoA [96]
28	Vitamin D receptor	Rickets [97]	Secosteroid [98]
29	DAF-12	Parasitic helminthes [99]	Steroidal ligand [100]
30	Fibroblast growth factor 22	Epilepsy [101]	Uncertain
31	Dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1	X-linked adrenal hypoplasia congenital [102]	Uncertain
32	CD81 receptor	Hepatitis C [103]	Various protein partners

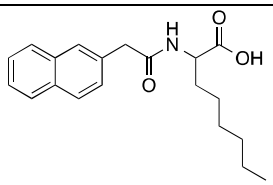
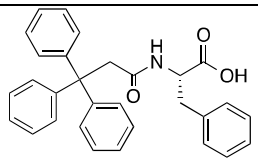
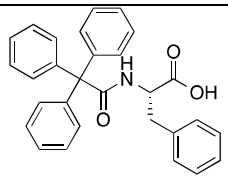
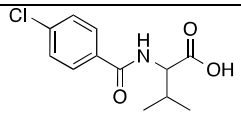
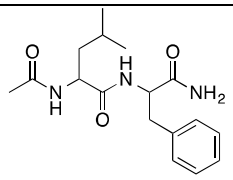
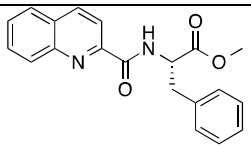
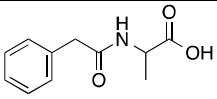
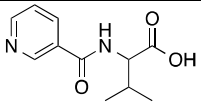
1		**9**	Unspecified	2787610	540299	1,4-dihydroxy-2-naphthoyl-CoA synthase (MenB)
2		**10**	Unspecified	3153153	540299	1,4-dihydroxy-2-naphthoyl-CoA synthase (MenB)
3		**11**	Unspecified	2787609	489031	Apoptotic protease activating factor 1
4		**12**	S (natural)	2304949	588511	Calcium-activated chloride channels (TMEM16A)
5		**13**	S (natural)	44350733	46724	Calpain
6		**14**	S (natural), S (natural)	7299379	262308 262309	Calpain
7		**15**	S (natural), S (natural)	114619	46713 50461 51388	Calpain; Cathepsin
8		**16**	S (natural), S (natural)	15017709	51388 46714 50461	Calpain; Cathepsin

9		**17**	S (natural), S (natural)	13386706	746599	Cathepsin
10		**18**	S (natural), S (natural)	152128	50461 51388	Cathepsin
11		**19**	S (natural), S (natural)	40522013	746599	Cathepsin
12		**20**	S (natural), S (natural)	7020268	746599	Cathepsin
13		**21**	S (natural), S (natural)	7272330	51388 50461	Cathepsin
14		**22**	S (natural)	10990457	37849	CD81 receptor
15		**23**	S (natural)	510498	42940	CD81 receptor
16		**24**	S (natural)	510499	42940	CD81 receptor
17		**25**	S (natural)	7116346	42940	CD81 receptor
18		**26**	Unspecified	294891	652126	DAF-12

19		**27**	R (unnatural)	439917	652067	DAF-12
20		**28**	Unspecified	93671	50465	Dopamine receptor
21		**29**	Unspecified	2787526	50465	Dopamine receptor
22		**30**	Unspecified, Unspecified	100050	798	Factor XIa
23		**31**	S (natural)	2060890	651658 588342	Fibroblast growth factor 22; Luciferase
24		**32**	S (natural)	5307679	504332	Histone lysine methyltransferase G9a
25		**33**	Unspecified	2918890	588342 2314 2315	Luciferase; Shiga Toxin
26		**34**	Unspecified	3136528	588676 1918	MAS-Related GPR Member X1 receptor; Renal outer medullary K+ channel
27		**35**	S (natural)	13746652	689654	Matrix metalloproteinase 2

28		**36**	S (natural)	71452972	689654	Matrix metalloproteinase 2
29		**37**	Unspecified	2002	602163 652067 504408 2563 504766	Methionine sulfoxide reductase A; DAF-12; Heat Shock Factor-1; Ras-converting enzyme; Dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1
30		**38**	Unspecified	466382	1529 720706	Mitogen-activated protein kinase kinase kinase 3 isoform 1; Ribosomal peptidyl transferase center
31		**39**	S (natural), S (natural)	11100291	462888	NK1 receptor
32		**40**	Unspecified	44142844	2445	Oxytocin receptor
33		**41**	R (unnatural)	14186396	474989	Peptidylprolyl cis/trans isomerase, NIMA-interacting 1
34		**42**	R (unnatural)	45100499	474989	Peptidylprolyl cis/trans isomerase, NIMA-interacting 1
35		**43**	R (unnatural)	45114305	474989 761669	Peptidylprolyl cis/trans isomerase, NIMA-interacting 1
36		**44**	R (unnatural)	46884012	474989	Peptidylprolyl cis/trans isomerase, NIMA-interacting 1

37		**45**	R (unnatural)	46884046	474989	Peptidylprolyl cis/trans isomerase, NIMA-interacting 1
38		**46**	R (unnatural)	46884048	474989	Peptidylprolyl cis/trans isomerase, NIMA-interacting 1
39		**47**	R (unnatural)	46884050	474989	Peptidylprolyl cis/trans isomerase, NIMA-interacting 1
40		**48**	R (unnatural)	46884146	474989	Peptidylprolyl cis/trans isomerase, NIMA-interacting 1
41		**49**	R (unnatural)	46884147	474989	Peptidylprolyl cis/trans isomerase, NIMA-interacting 1
42		**50**	R (unnatural)	46884149	474989	Peptidylprolyl cis/trans isomerase, NIMA-interacting 1
43		**51**	Unspecified	16327462	493034 492953	Platelet-activating factor acetylhydrolase 1b, catalytic subunit 2
44		**52**	Unspecified	2999992	492953	Platelet-activating factor acetylhydrolase 1b, catalytic subunit 2
45		**53**	Unspecified	41410	492953 493034	Platelet-activating factor acetylhydrolase 1b, catalytic subunit 2
46		**54**	Unspecified	44601509	492953 493034	Platelet-activating factor acetylhydrolase 1b, catalytic subunit 2

47		**55**	Unspecified	44601510	493034 492953	Platelet-activating factor acetylhydrolase 1b, catalytic subunit 2
48		**56**	S (natural)	16066615	492953	Platelet-activating factor acetylhydrolase 1b, catalytic subunit 2
49		**57**	S (natural)	16066508	720504 624417 720706	Polo-like kinase 1 - polo-box domain; GLP-1 receptor; Ribosomal peptidyl transferase center
50		**58**	Unspecified	2770855	485272	Protein arginine deiminase 4
51		**59**	Unspecified, Unspecified	646621	2314 652051 652048	Shiga toxin; Dopamine receptor
52		**60**	S (natural)	9439964	2315 588342	Shiga toxin; Luciferase
53		**61**	Unspecified	564251	485364	Thioredoxin glutathione reductase
54		**62**	Unspecified	299354	504847	Vitamin D receptor