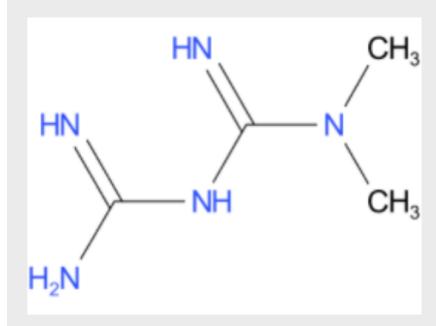


Module 8 – Cheminformatics

- Find drug "metformin" at [Drug Central](#) or elsewhere and copy the image of this molecular structure into your answer. From this image, draw the same molecule using the molecular editor JSME, at <http://pasilla.health.unm.edu/jsme.html>. Use the "Get SMILES" button to obtain the SMILES for the molecule, and include in answer.

Metformin molecular structure:



SMILES for the molecule: CN(C)C(=N)NC(N)=N

Note: I couldn't get JSME to work properly on my Mac. I spent a lot of time trying to create the molecule to no success. So I got the SMILES for the molecule by clicking the SMILES link for metformin in Drug Central.

- Using this SMILES, search PubChem, the structural search tool, at <https://pubchem.ncbi.nlm.nih.gov>, in similarity mode. Describe results. Was the drug found as the top hit? Download the hits in CSV format (maximum 100).

Yes, the drug was found as the top hit:

The screenshot shows the PubChem search interface. The search bar contains the SMILES string: CN(C)C(=N)NC(N)=N. The results page indicates 1 to 20 of 48 hits. The top hit is metformin, with the following details:

- PubChem CID: 4091
- MF: C₆H₁₁N₅
- MW: 129.163 g/mol
- IUPAC Name: 3-(diaminomethylidene)-1,1-dimethylguanidine

Downloaded file:

DSDHT FALL 2017
Carlos Sathler (cssathler@gmail.com)

mw	cid	cmpdname	cmpdiupacname	mf
129.163	4091	metformin	3-(diaminomethylidene)-1,1-dimethylguanidine	C4H11N5
129.163	3043660	AC1MIFQL	1-carbamimidoyl-1,2-dimethylguanidine	C4H11N5
157.216	4306178	AC1N7PIY	2-(N,N-dimethylcarbamimidoyl)-1,1-dimethylguanidine	C6H15N5
157.216	10313238	SCHEMBL471123	2-(N,N-dimethylcarbamimidoyl)-1,1-dimethylguanidine	C6H15N5
130.171	60003225	AGN-PC-0D7U04	[N'-(N,N-dimethylcarbamimidoyl)carbamimidoyl]azanium	C4H12N5+
128.164	73388104	CHEMBL3094198	3-(diaminomethylidene)-1,1-di(methyl)guanidine	C4H11N5
130.169	90702925	GLUCOPHAGE-D	2-(N'-deuteriocarbamimidoyl)-1,1-dimethylguanidine	C4H11N5
135.2	45039708		3-(diaminomethylidene)-1,1-bis(trideuteriomethyl)guanidine	C4H11N5
165.624	14219	Metformin hydrochloride	3-(diaminomethylidene)-1,1-dimethylguanidine;hydrochloride	C4H12CIN5
140.146	222300	(6e)-6-imino-1-methyl-1,6-dihydro-1,3,5-triazine-2,4-diamine	6-imino-1-methyl-1,3,5-triazine-2,4-diamine	C4H8N6
164.616	21177118	HMS2093F08	3-(diaminomethylidene)-1,1-dimethylguanidine;chloride	C4H11CIN5-
165.624	44573417	metformin XR	1-carbamimidoyl-1,2-dimethylguanidine;hydrochloride	C4H12CIN5
194.543	86605940	Dimethylbiguanide Zinc	3-(diaminomethylidene)-1,1-dimethylguanidine;zinc	C4H11N5Zn
210.075	11310327	AGN-PC-006IPI	3-(diaminomethylidene)-1,1-dimethylguanidine;hydrobromide	C4H12BrN5
171.661	45039707	Metformin-d6, Hydrochloride	3-(diaminomethylidene)-1,1-bis(trideuteriomethyl)guanidine;hydrochloride	C4H12CIN5
192.669	49876377	AKOS001015598	2-(N,N-dimethylcarbamimidoyl)-1,1-dimethylguanidine;chloride	C6H15CIN5-
165.624	53630812	177024-EP2275427A1	3-(diaminomethylidene)-1,1-dimethylguanidine;hydron;chloride	C4H12CIN5
336.363	87223577	SCHEMBL1290484	3-(diaminomethylidene)-1,1-dimethylguanidine;lead	C4H11N5Pb
237.031	87259835	SCHEMBL1560020	3-(diaminomethylidene)-1,1-dimethylguanidine;silver	C4H11AgN5
157.216	86292532	SCHEMBL11889600	1-carbamimidoyl-1,2,3,3-tetramethylguanidine	C6H15N5
157.216	86292533	SCHEMBL14868262	1,1,3-trimethyl-3-(N'-methylcarbamimidoyl)guanidine	C6H15N5
141.174	89946131	SCHEMBL15444267	3-[amino-(methylideneamino)methylidene]-1,1-dimethylguanidine	C5H11N5
143.19	53307485	SCHEMBL471124	1,1-dimethyl-2-(N'-methylcarbamimidoyl)guanidine	C5H13N5
143.19	73734534	AGN-PC-034KWZ	1,1-dimethyl-2-(N'-methylcarbamimidoyl)guanidine	C5H13N5
143.19	21027224		1,1-dimethyl-2-(N'-methylcarbamimidoyl)guanidine	C5H13N5
143.19	86292343	SCHEMBL14868271	3-(diaminomethylidene)-1,1,2-trimethylguanidine	C5H13N5
143.19	86292531	SCHEMBL14868263	1,2-dimethyl-1-(N'-methylcarbamimidoyl)guanidine	C5H13N5
232.101	15949929	AGN-PC-00QBLJ	chromium;3-(diaminomethylidene)-1,1-dimethylguanidine;vanadium	C4H11CrN5V
193.677	16235945	T0501-8461	2-(N,N-dimethylcarbamimidoyl)-1,1-dimethylguanidine;hydrochloride	C6H16CIN5
202.085	23615830	SCHEMBL1547913	3-(diaminomethylidene)-1,1-dimethylguanidine;dihydrochloride	C4H13Cl2N5
240.045	89038103	SCHEMBL13206000	N'-(N,N-dimethylcarbamimidoyl)carbamimidoyl iodide	C4H9IN4
221.469	87281700	SCHEMBL1714633	3-(diaminomethylidene)-1,1-dimethylguanidine;iron;hydrochloride	C4H12ClFeN5

DSDHT FALL 2017
Carlos Sathler (cssathler@gmail.com)

mw	cid	cmpdname	cmpdipacname	mf
230.274	87088497	SCHEMBL292820	3-(diaminomethylidene)-1,1-dimethylguanidine;1-(diaminomethylidene)guanidine	C6H18N10
112.133	121135	N'-Cyano-N,N-dimethylguanidine	2-cyano-1,1-dimethylguanidine	C4H8N4
112.133	21023741	SCHEMBL2379315	1-cyano-1,2-dimethylguanidine	C4H8N4
127.147	55281088	SCHEMBL6248347	1-methyl-2H-1,3,5-triazine-4,6-diamine	C4H9N5
114.149	73676094	SCHEMBL4097383	N'-carbamimidoyl-N,N-dimethylmethanimidamide	C4H10N4
127.167	21122269	AGN-PC-03655D	cyano-dimethyl-(N'-methylcarbamimidoyl)azanium	C5H11N4+
114.149	20314408		N'-carbamimidoyl-N,N-dimethylmethanimidamide	C4H10N4
158.224	29530186	ZINC21487152	[amino-[[amino(dimethylamino)methylidene]amino]methylidene]-dimethylazanium	C6H16N5+
179.651	53307484	SCHEMBL13584677	1,1-dimethyl-2-(N'-methylcarbamimidoyl)guanidine;hydrochloride	C5H14CIN5
131.179	88591075	SCHEMBL10310520	2-(diaminomethyl)-1,1-dimethylguanidine	C4H13N5
171.243	86292536	SCHEMBL14868259	(3E)-3-[amino(dimethylamino)methylidene]-1,1,2-trimethylguanidine	C7H17N5
127.147	89506965	SCHEMBL14869901	5-methyl-2H-1,3,5-triazine-4,6-diamine	C4H9N5
101.15	141619	N,N,N'-Trimethylguanidine	1,1,2-trimethylguanidine	C4H11N3
116.184	15603	ZINC00164734	[amino(dimethylamino)methylidene]-dimethylazanium	C5H14N3+
116.184	18921325	SCHEMBL548143	trimethyl-(N'-methylcarbamimidoyl)azanium	C5H14N3+
128.155	58635196	AGN-PC-0CI37G	methyl-[N'-(N'-methylcarbamimidoyl)carbamimidoyl]azanide	C4H10N5-

3. Using [KEGG](#), find protein targets associated with a disease of interest. You may select your own, or use "diabetes". We can hypothesize that these targets could be modulated by drugs therapeutically as "drug targets". Report your findings and provide illustrative screenshots of KEGG.

I'm selecting "asthma". We can find protein targets for a possible drug by exploring the genes that are in the asthma pathway in KEGG. They are:

Pathway	hsa05310 Asthma
Gene	IL4 (polymorphism) [HSA:3565] [KO:K05430] IL4RA (polymorphism) [HSA:3566] [KO:K05071] IL13 (polymorphism) [HSA:3596] [KO:K05435] FCER1B (polymorphism) [HSA:2206] [KO:K08090] TNFA (polymorphism) [HSA:7124] [KO:K03156] ADAM33 (polymorphism) [HSA:80332] [KO:K08616] CD14 (polymorphism) [HSA:929] [KO:K04391] HLA-DRB1 (polymorphism) [HSA:3123] [KO:K06752] HLA-DQB1 (polymorphism) [HSA:3119] [KO:K06752] ADRB2 (polymorphism) [HSA:154] [KO:K04142]

We need to find the gene products of these genes. By clicking on the links for each of them, we see KEGG's page for the specific gene, for example for IL13:

DSDHT FALL 2017
Carlos Sathler (cssathler@gmail.com)

The Definition field on this page (“Description of the ortholog group followed by the EC number, if any.” per KEGG help) identifies the protein associated with this gene. I confirmed that in UniProt database for several genes in the “asthma” pathway. For example, for IL13:

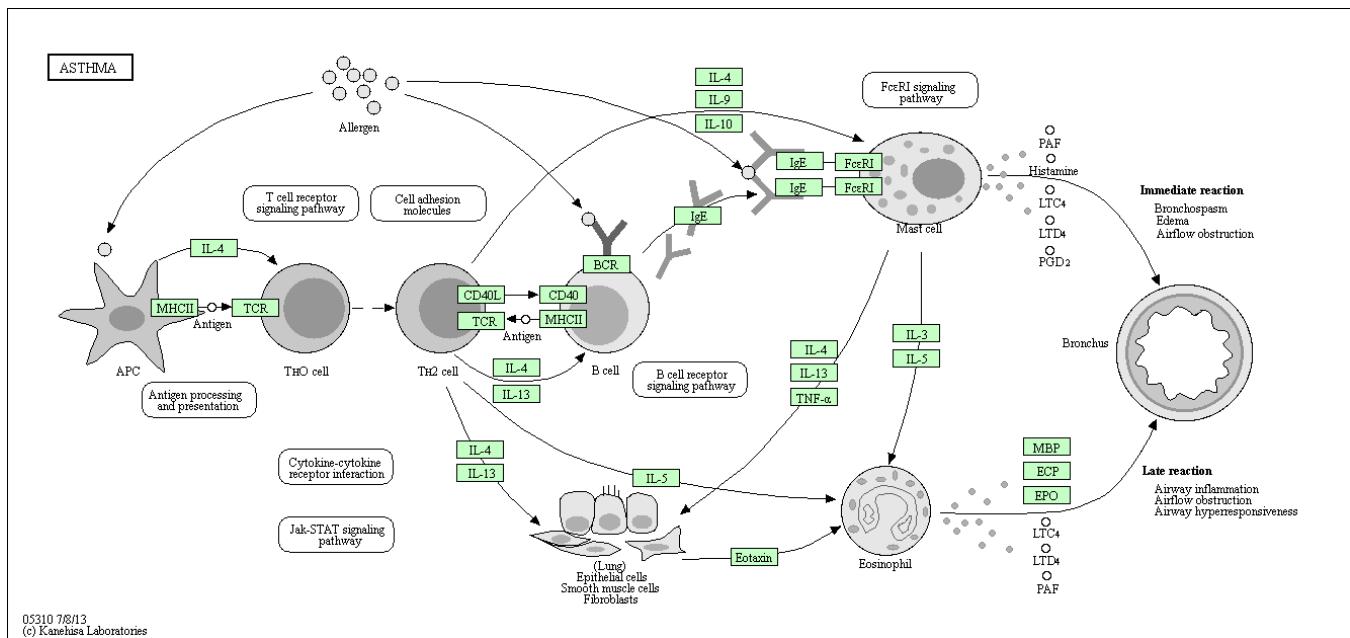
By exploring gene products in the asthma pathway following this process we are able to identify “drug targets” for asthma. The list includes Interleukin-4, Interleukin-13, interleukin 4 receptor, high affinity immunoglobulin epsilon receptor subunit beta, tumor necrosis factor superfamily, member 2, disintegrin and metalloproteinase domain-containing protein 33, monocyte differentiation antigen CD14, major histocompatibility complex, class II and adrenergic receptor beta-2, among a few others.

4. Using [PubChem](#), find bioassays involving one of the protein targets from KEGG. Find the compounds which are most active against this target. Report your findings and provide illustrative screenshots of PubChem.

None of the “drug targets” identified by searching the immediate gene products in the asthma pathway has yielded a match in PubChem. So, I looked again at the asthma pathway in KEGG (next page) and searched for downstream effects of the genes products in the pathway. Asthma is an inflammation of the bronchus, per the KEGG description of the disease. There are many substances in the pathway that affect the bronchus directly causing the symptoms of asthma. These substances appear in the arrows connected directly to the bronchus at the end of the pathway. I searched for them and the following yielded a direct match in PubChem: Histamine, LTC4 and LTD4.

DSDHT FALL 2017

Carlos Sathler (cssathler@gmail.com)



We will look for compounds that are most active against “histamine”.

PubMed results for “antihistamine”:

Screenshot of a web browser showing search results for "Antihistamine" on NCBI's PubChem Compound database. The search results list four compounds:

- diphenhydramine; Benadryl; Benzhydryamine**
MW: 255.361 g/mol MF: C₁₇H₂₂NO
IUPAC name: 2-benzhydryloxy-N,N-dimethylethanamine
Create Date: 2005-03-25
CID: 3100
[Summary](#) [Similar Compounds](#) [Same Parent_Connectivity](#) [Mixture/Component Compounds](#) [PubMed \(MeSH Keyword\)](#)
- Diphenylpyraline hydrochloride; 132-18-3; Allerzine**
MW: 317.857 g/mol MF: C₁₉H₂₂CINO
IUPAC name: 4-benzhydryloxy-1-methylpiperidine;hydrochloride
Create Date: 2005-03-27
CID: 66063
[Summary](#) [Similar Compounds](#) [Same Parent_Connectivity](#) [Mixture/Component Compounds](#) [PubMed \(MeSH Keyword\)](#)
- pyramilene; Mepyramine; Pyranisamine**
MW: 265.391 g/mol MF: C₁₇H₂₃N₃O
IUPAC name: N-[(4-methoxyphenyl)methyl]-N,N-dimethyl-N'-pyridin-2-yleth...
Create Date: 2005-03-25
CID: 4992
[Summary](#) [Similar Compounds](#) [Same Parent_Connectivity](#) [Mixture/Component Compounds](#) [PubMed \(MeSH Keyword\)](#)
- loratadine; 79794-75-5; Claritin**
MW: 382.888 g/mol MF: C₂₂H₂₃CIN₂O₂
IUPAC name: ethyl 4-(8-chloro-5,6-dihydrobenzo[1,2]cyclohepta[2,4-b]pyri...
Create Date: 2005-03-25
CID: 3957
[Summary](#) [Similar Compounds](#) [Same Parent_Connectivity](#) [Mixture/Component Compounds](#) [PubMed \(MeSH Keyword\)](#)

The right side of the interface includes filters for bioactivity analysis, structure clustering, and structure download. It also provides links to refine results by chemical properties, bioactivity experiments, and biomedical annotation.

I consider all but one of the eight antihistamine compounds found effective. Except for the second one in the screenshot, all compounds are said to be Histamine-1 receptor antagonists and I recognize some of the drug names associated with them, such as Benadryl and Claritin, for example.

5. Consider your "hitlist" of compounds and criteria for their potential as drug leads. Download the compounds and all the properties available from PubChem. Assess these compounds using the "Lipinsky Rule of Five". This can be done using the web app at <http://pasilla.health.unm.edu/tomcat/biocomp/ro5>

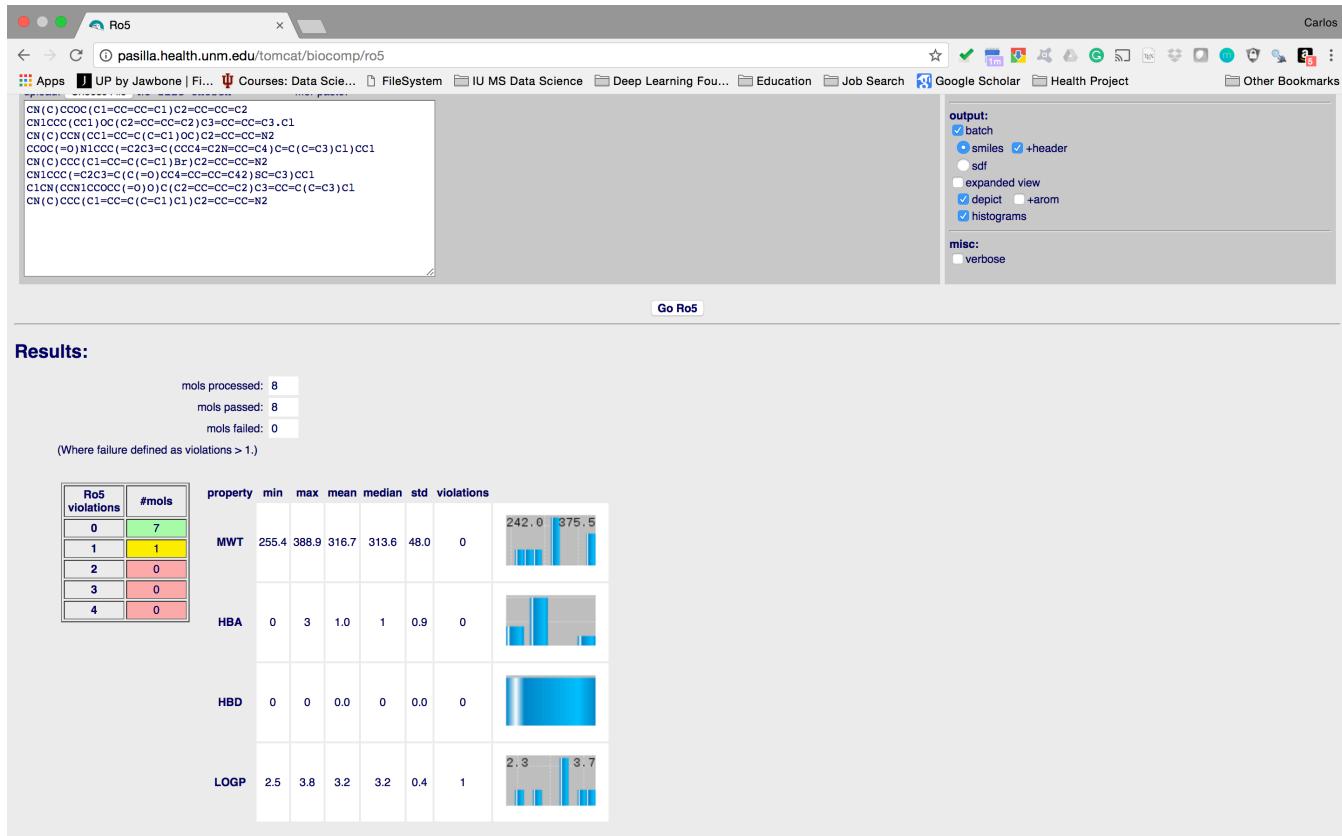
I downloaded a SMILES formatted file with 8 antihistamine compounds:

```

3100 CN (C) CCOC (C1=CC=CC=C1) C2=CC=CC=C2
66063 CN1CCC (CC1) OC (C2=CC=CC=C2) C3=CC=CC=C3 . C1
4992 CN (C) CCN (CC1=CC=C (C=C1) OC) C2=CC=CC=N2
3957 CCOC (=O) N1CCC (=C2C3=C (CCC4=C2N=CC=C4) C=C (C=C3) C1) CC1
6834 CN (C) CCC (C1=CC=C (C=C1) Br) C2=CC=CC=N2
3827 CN1CCC (=C2C3=C (C (=O) CC4=CC=CC=C42) SC=C3) CC1
2678 C1CN (CCN1CCOC ( =O) O) C (C2=CC=CC=C2) C3=CC=C (C=C3) C1
2725 CN (C) CCC (C1=CC=C (C=C1) C1) C2=CC=CC=N2

```

Results in ro5:



The Lipinski Rule of Five (Ro5) specifies that compounds should violate no more than one of four criteria measured against the properties seen in the screenshot (MWT, HBA, HBD and LOGP). The screenshot shows 7 of the 8 compounds violate 0 rules, and 1 out of 8 compounds violate a single rule. So, the compounds meet the requirements for “drug-likeness”.