Cheminformatic Analysis of Natural Product-based Drugs and Chemical Probes

Samantha Stone[a], David J. Newman[b], Steven L. Colletti[c], Derek S. Tan[a,d]*

[a]*Chemical Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA*

[b]*NIH Special Volunteer, Wayne, Pennsylvania 19087, USA*

[c]*Zymergen, Inc., 430 E 29th St, New York, New York 10016, USA*

[d]*Tri-Institutional Research Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA*

Supporting Information

A. SUPPLEMENTARY FIGURES S1–S15

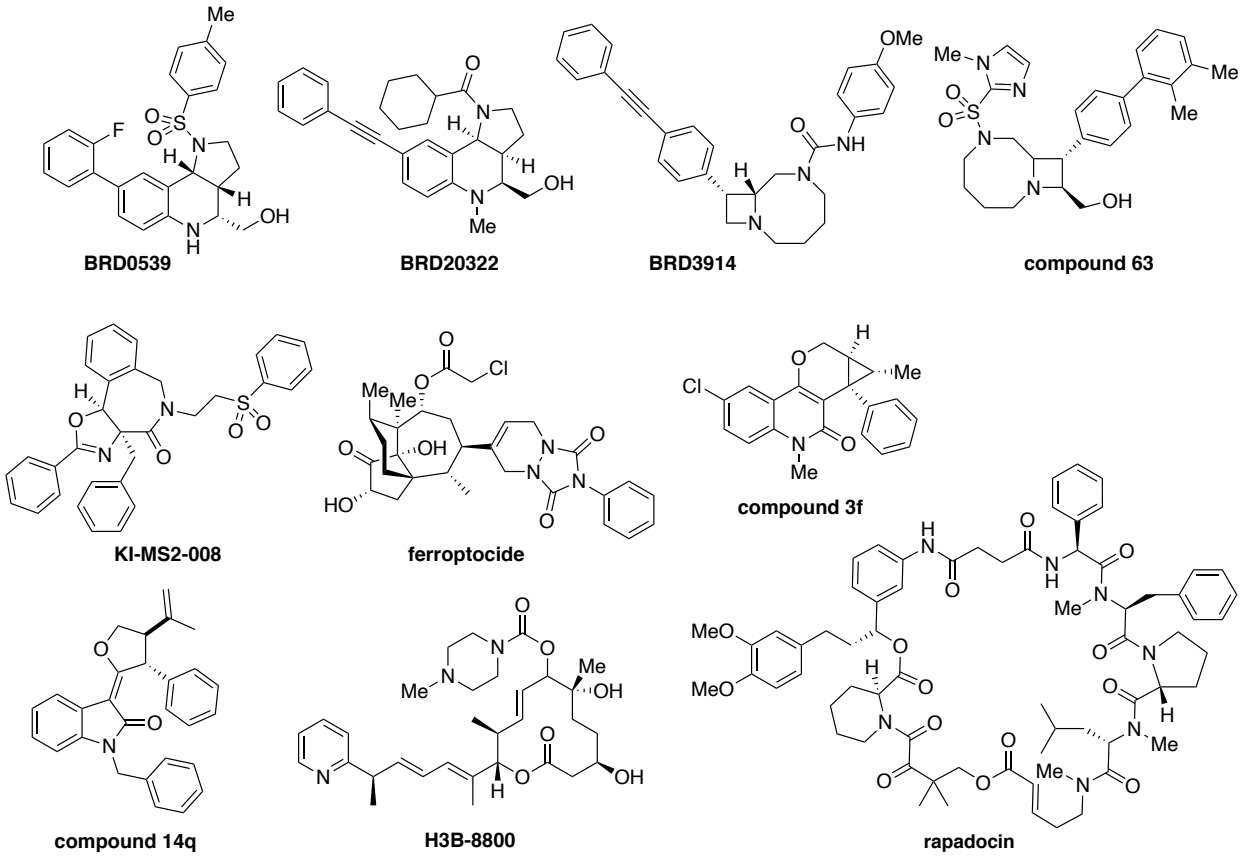

Figure S1. Structures of chemical probes recently discovered from DOS libraries.

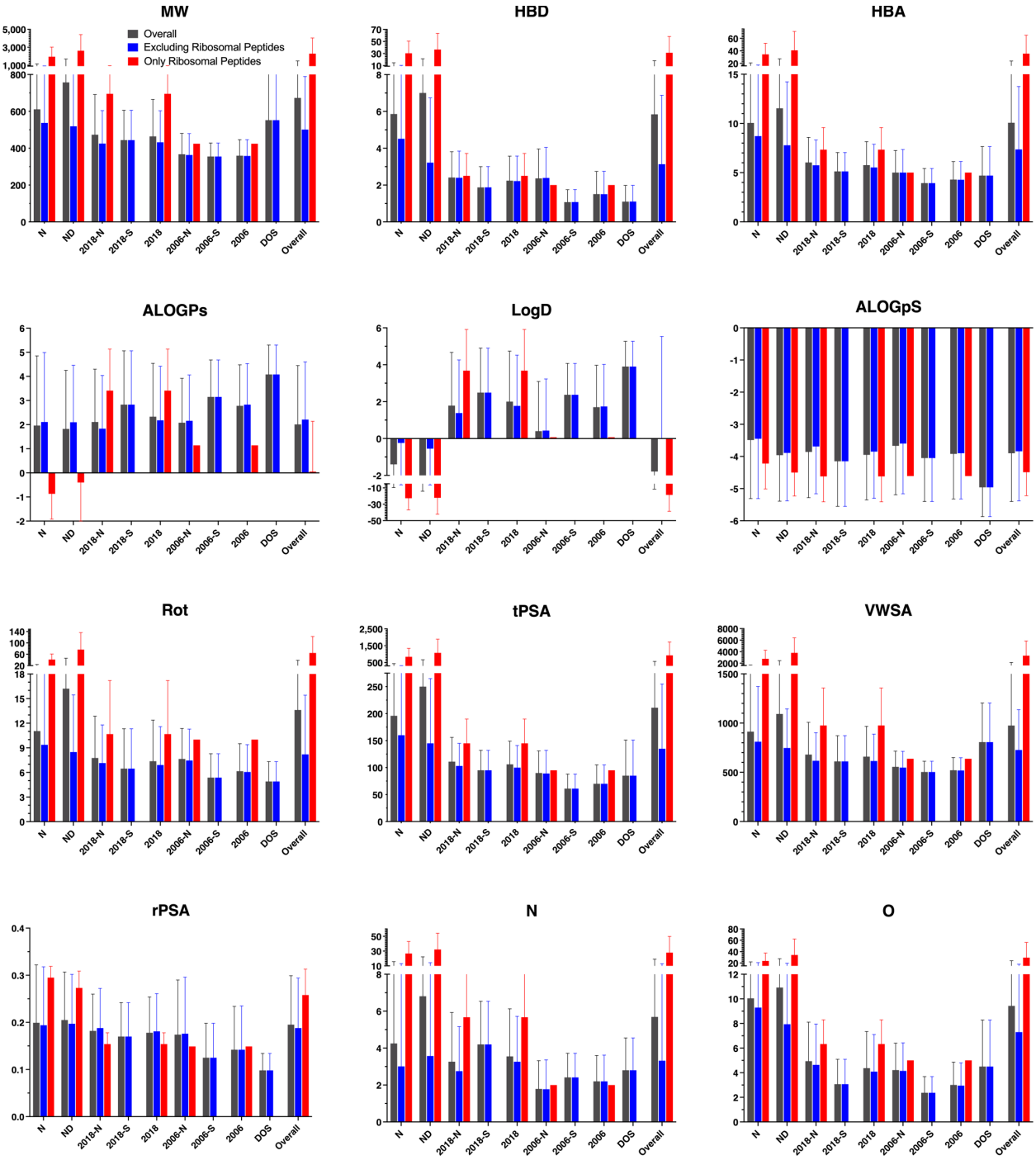

Figure S2. Complete bar graphs of all 20 structural and physicochemical properties of natural product drugs (N), natural product-derived drugs (ND), top 40 brand-name drugs from 2006 and 2018 from natural product (-N) and purely synthetic (-S) origins, and recently discovered chemical probes from diversity-oriented synthesis libraries (DOS). **(continues on following page)**

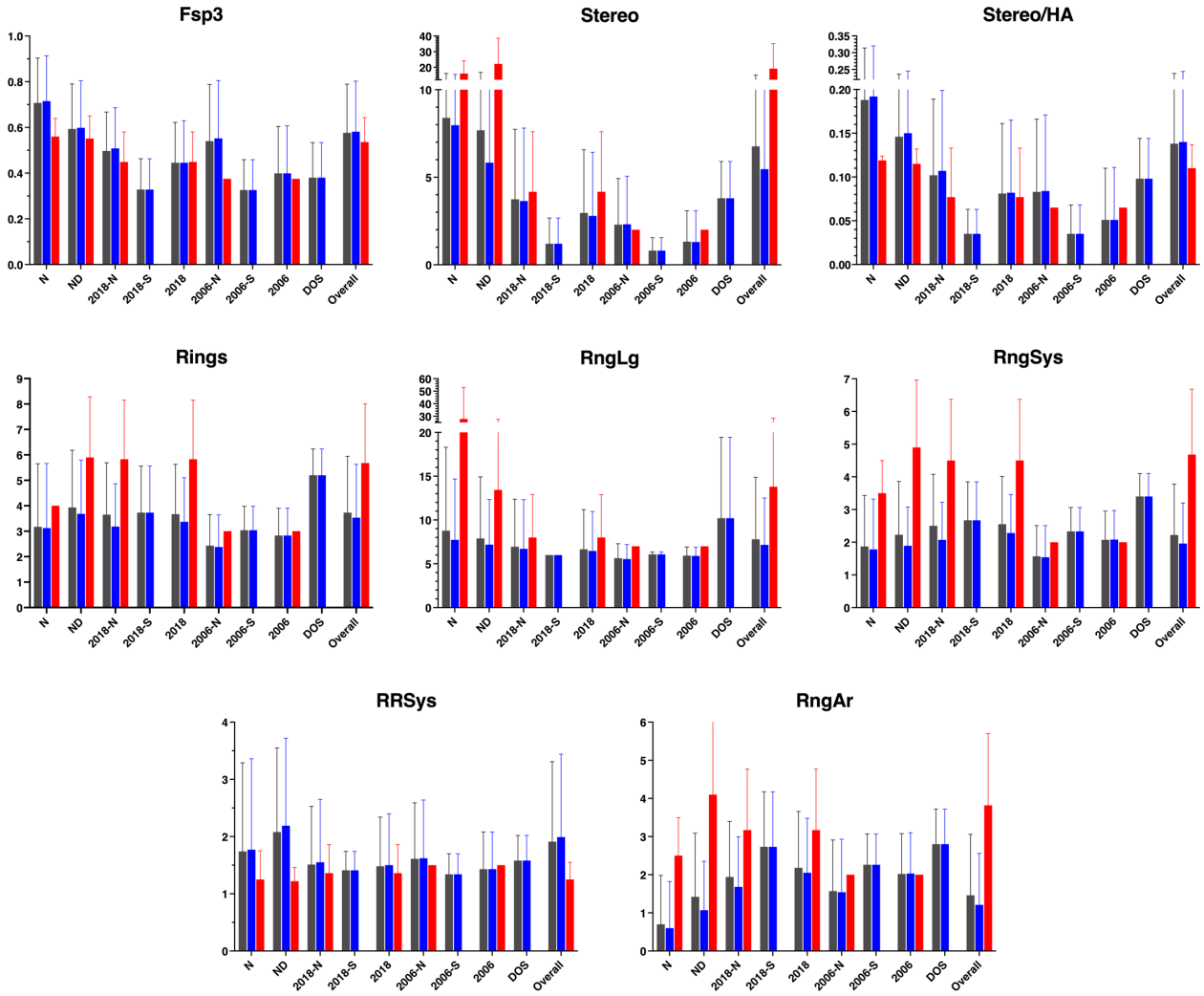

Figure S2 (continued).

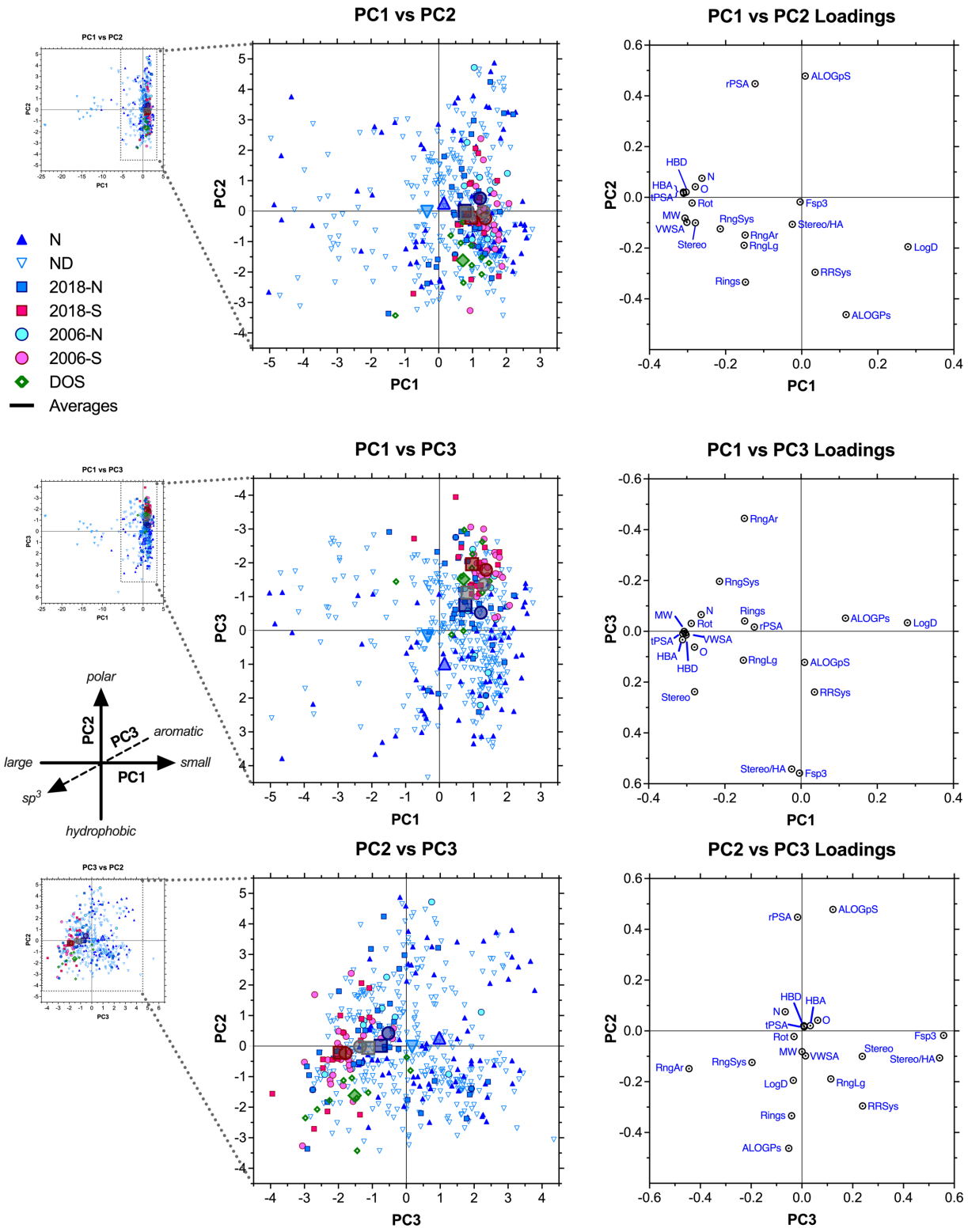

Figure S3. Complete PCA and loading plots based on 20 structural and physicochemical parameters for natural product drugs (N), natural product-derived drugs (ND), top 40 brand name drugs in 2006 and 2018 derived from natural products (-N) or purely synthetic (-S), and recently discovered chemical probes from diversity-oriented synthesis libraries (DOS).

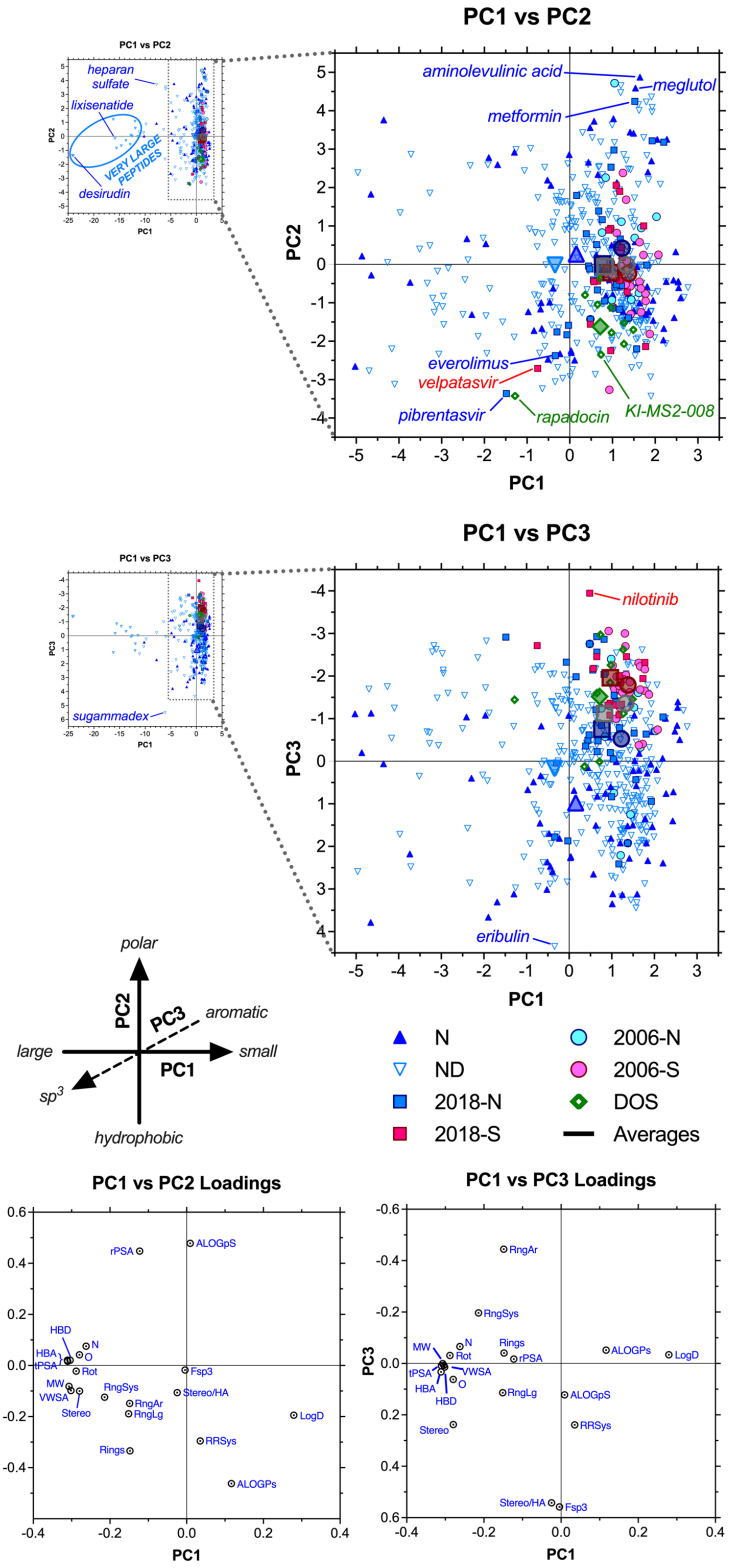

Figure S4. PCA plots of drugs coded by category with selected compounds labeled.

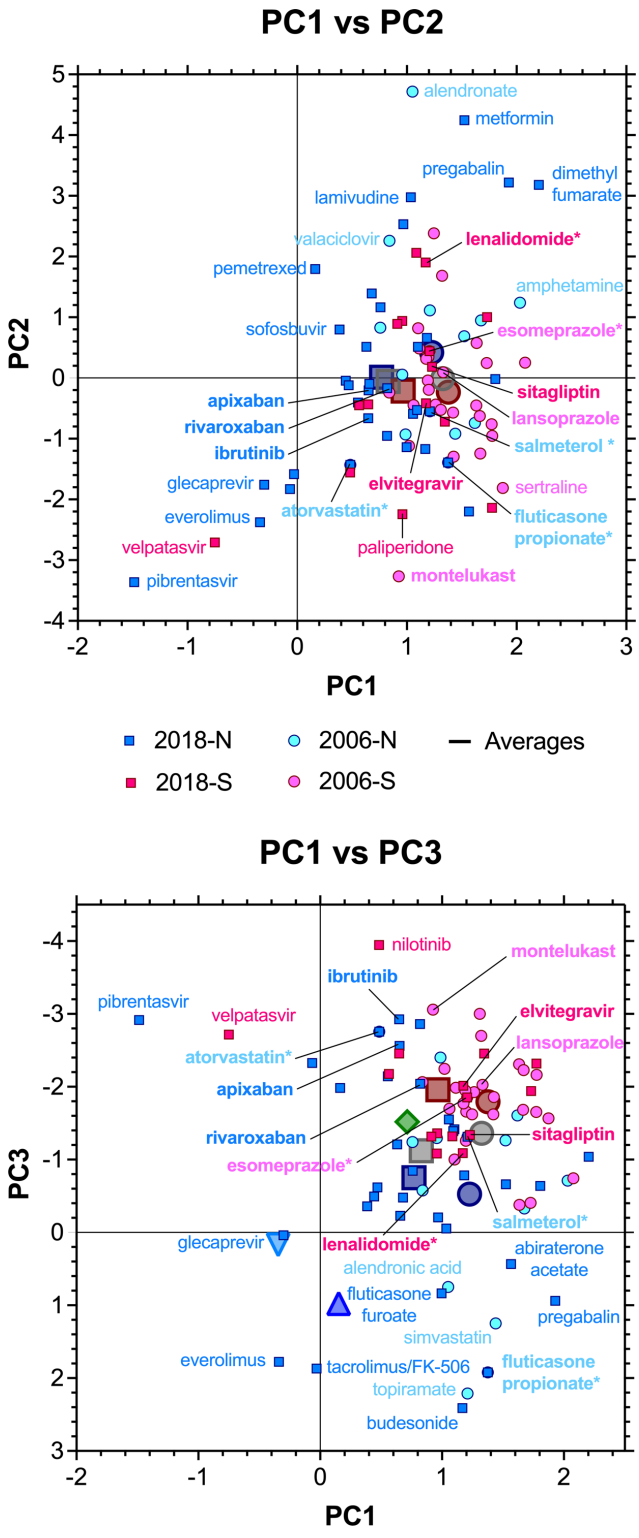

Figure S5. PCA plots of drugs coded by category showing only top 40 brand name drugs in 2006 and 2018 derived from natural products (-N) or purely synthetic (-S).

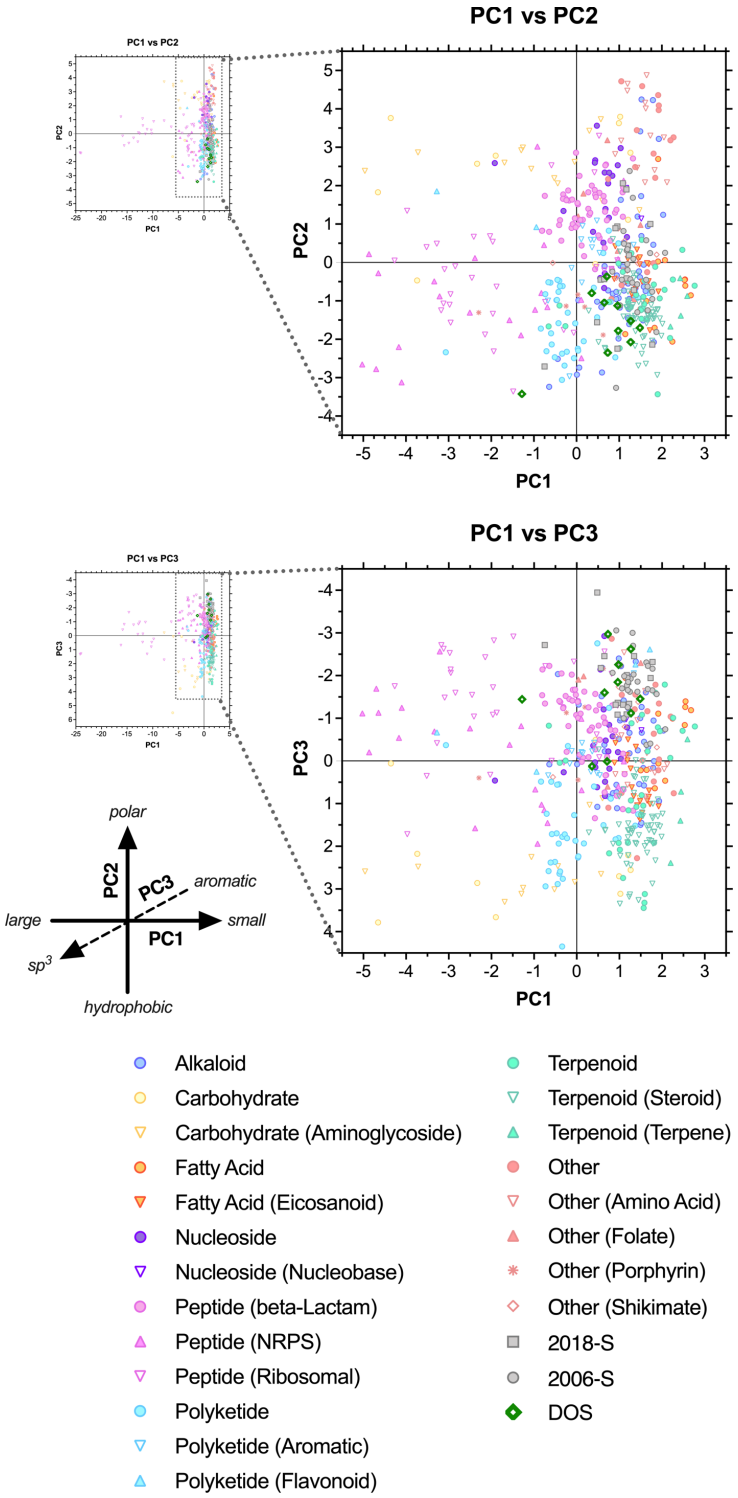

Figure S6. PCA plots of drugs coded by biosynthetic origin (without annotation of macrocycles).

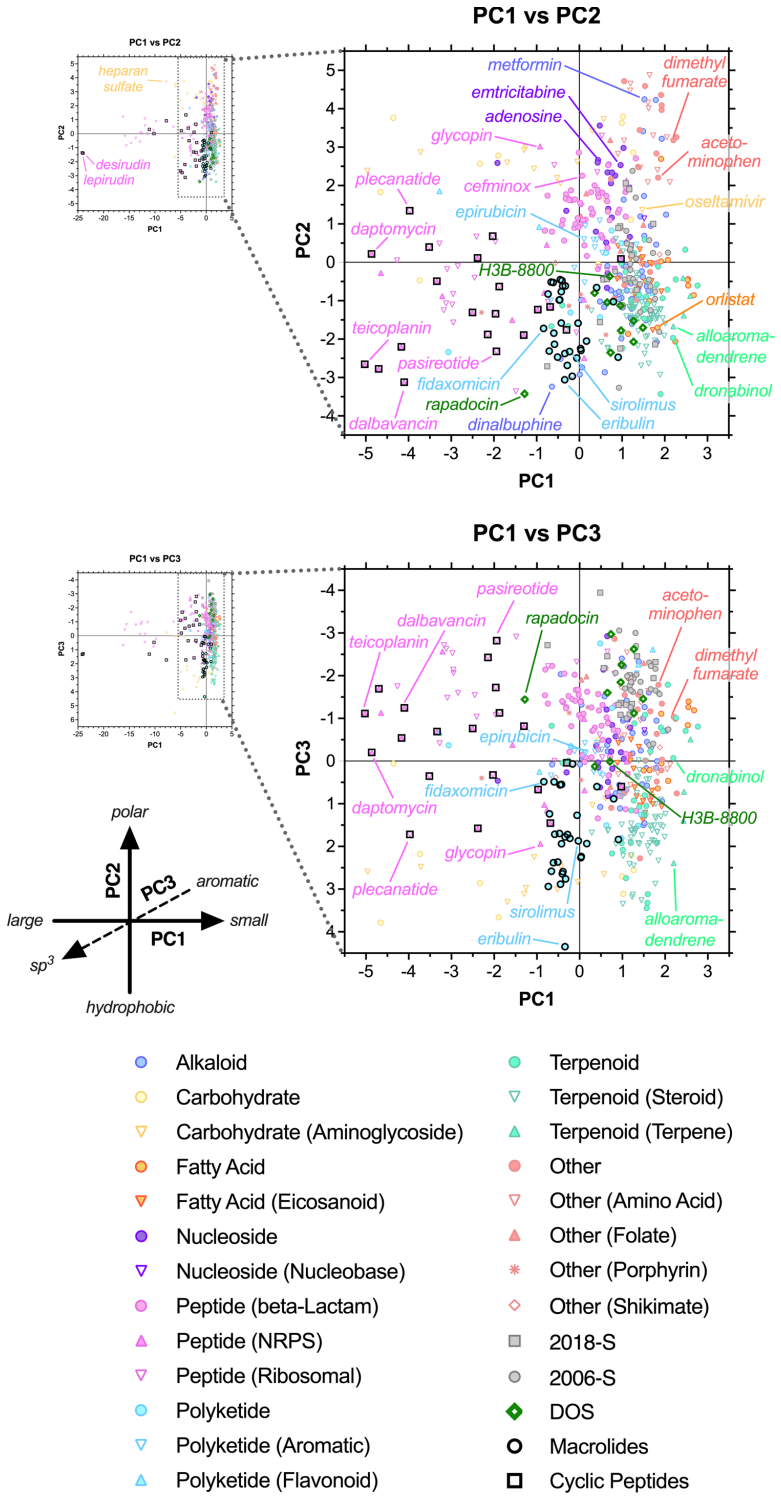

Figure S7. PCA plots of drugs coded by biosynthetic origin with selected compounds labeled.

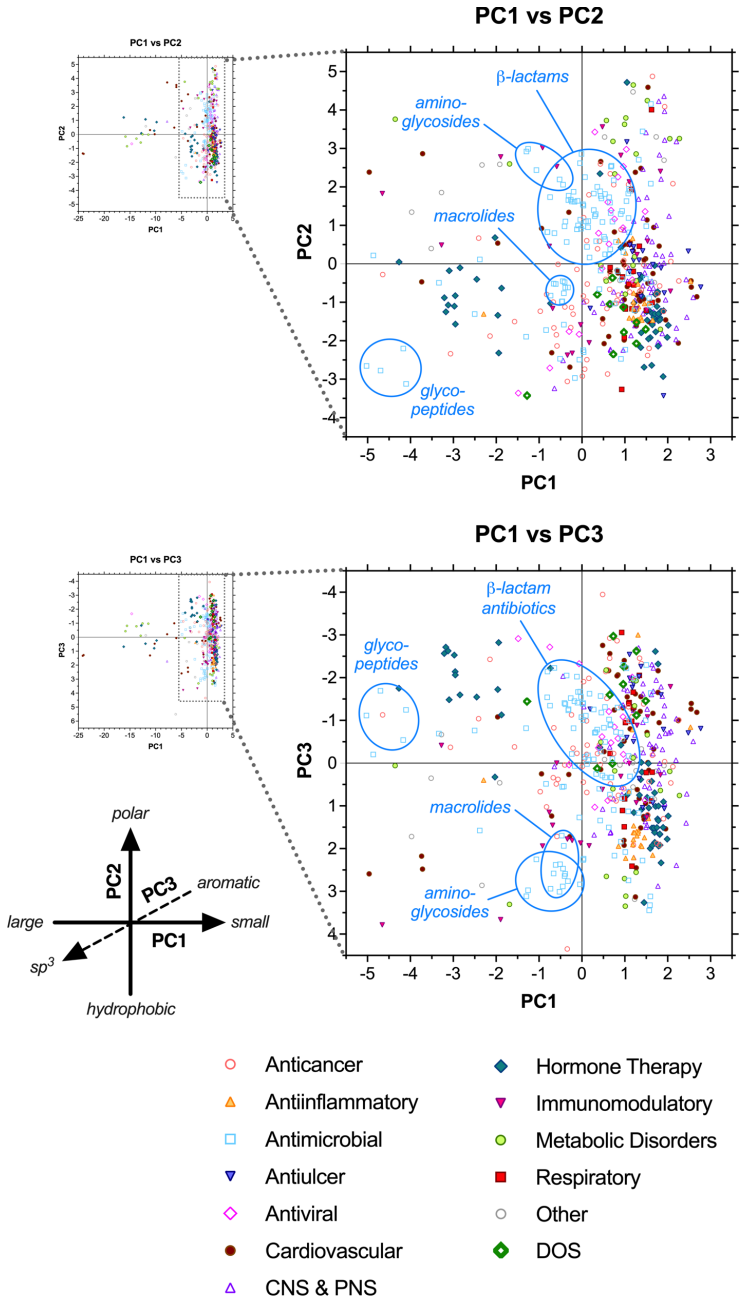

Figure S8. PCA plots of drugs coded by therapeutic class with selected antimicrobial drug subclasses labeled.

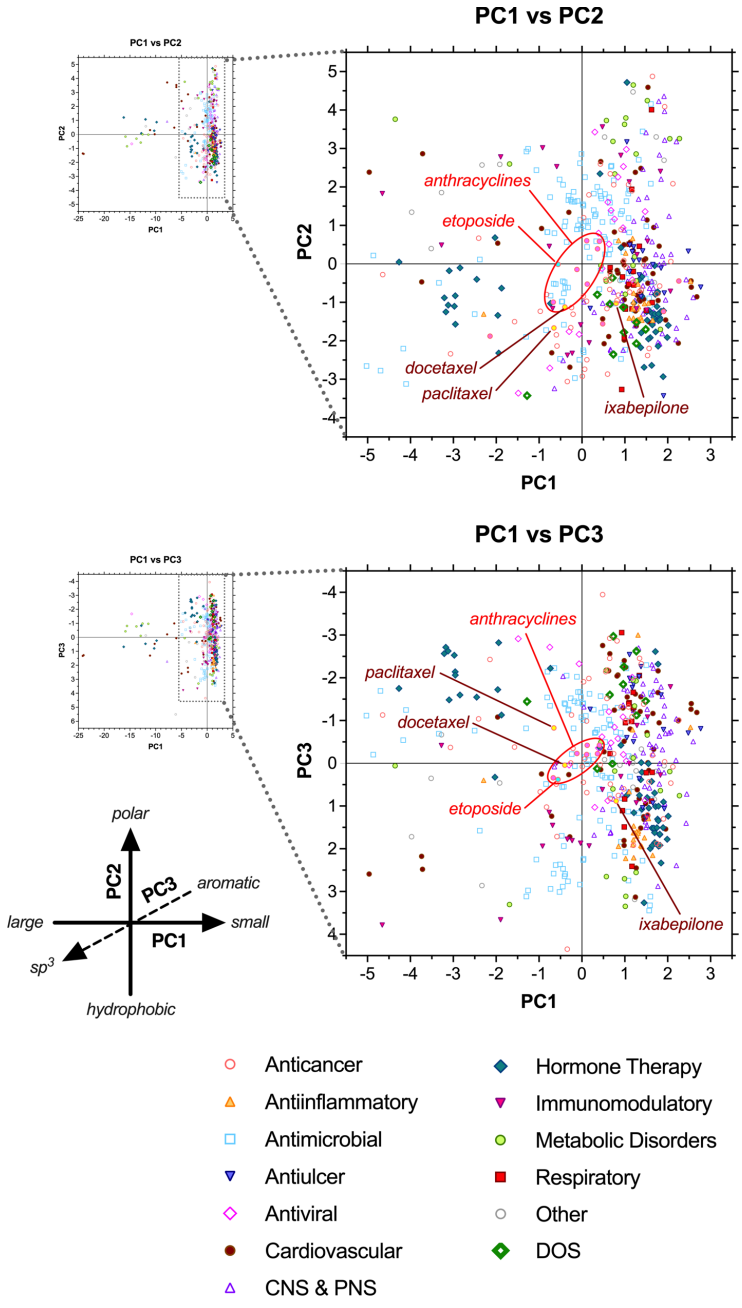

Figure S9. PCA plots of drugs coded by therapeutic class with selected anticancer drugs labeled. The anthracyclines and etoposide (red) target topoisomerase II, and fall in overlapping regions. Similarly, paclitaxel, docetaxel, and ixebpilone (maroon) target tubulin polymerization and fall in distinct but nearby regions.

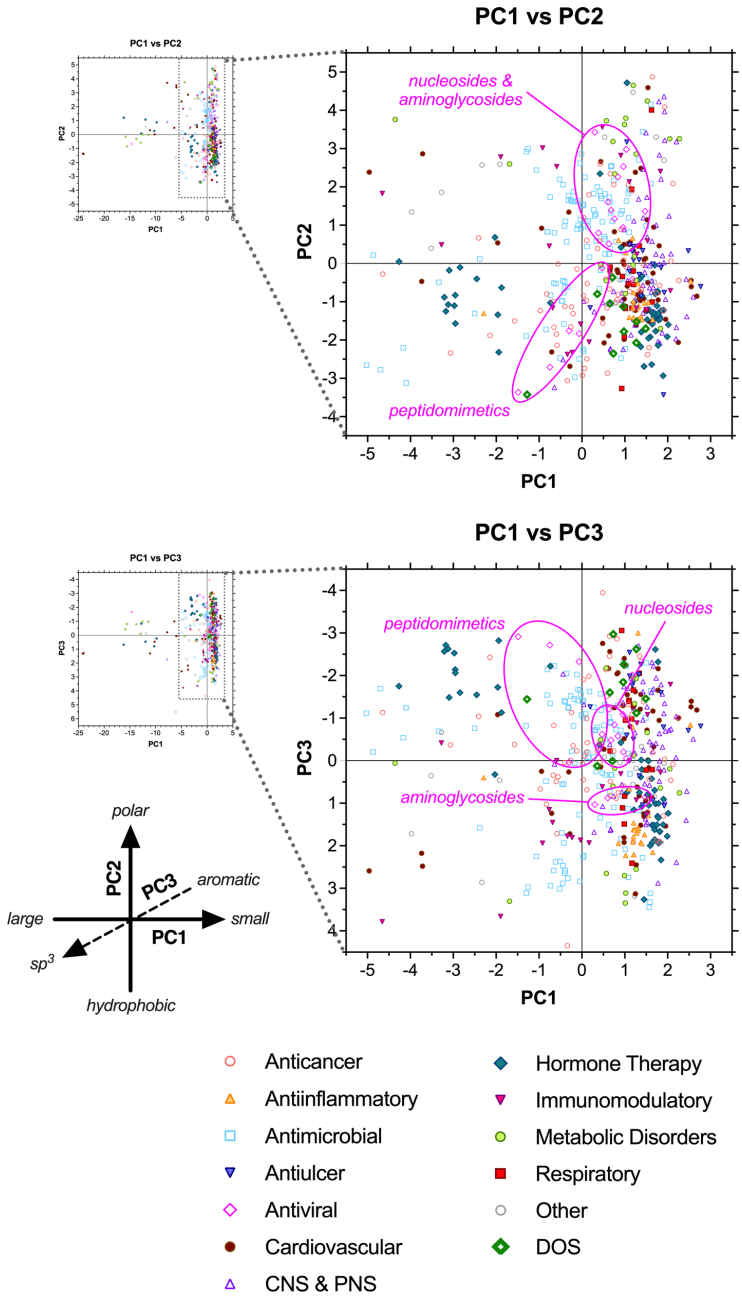

Figure S10. PCA plots of drugs coded by therapeutic class with selected antiviral drug subclasses labeled.

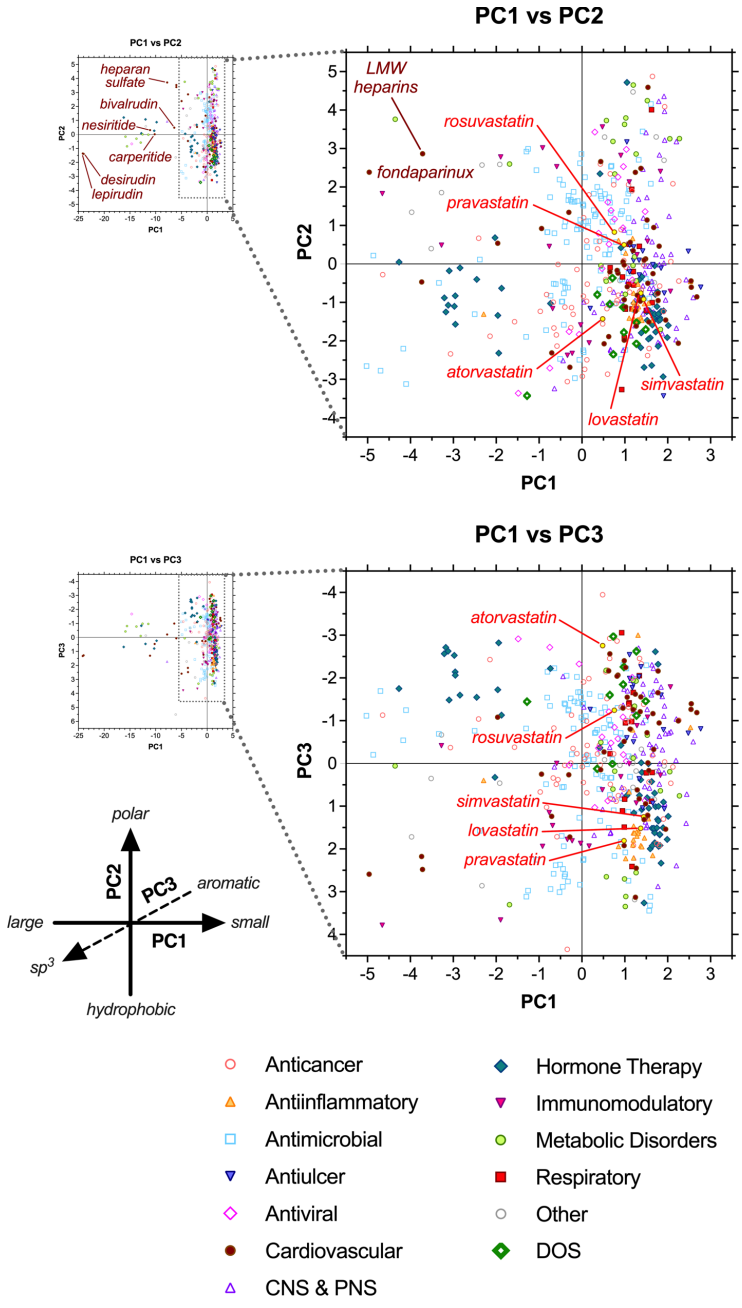

Figure S11. PCA plots of drugs coded by therapeutic class with selected cardiovascular drugs labeled. The statin drugs (red) all share the same molecular target, HMG-CoA reductase, but the natural products (N: simvastatin, lovastatin, pravastatin) differentiate from the natural product-mimetic drugs (S*/NM: atorvastatin, rosuvastatin) along PC3.

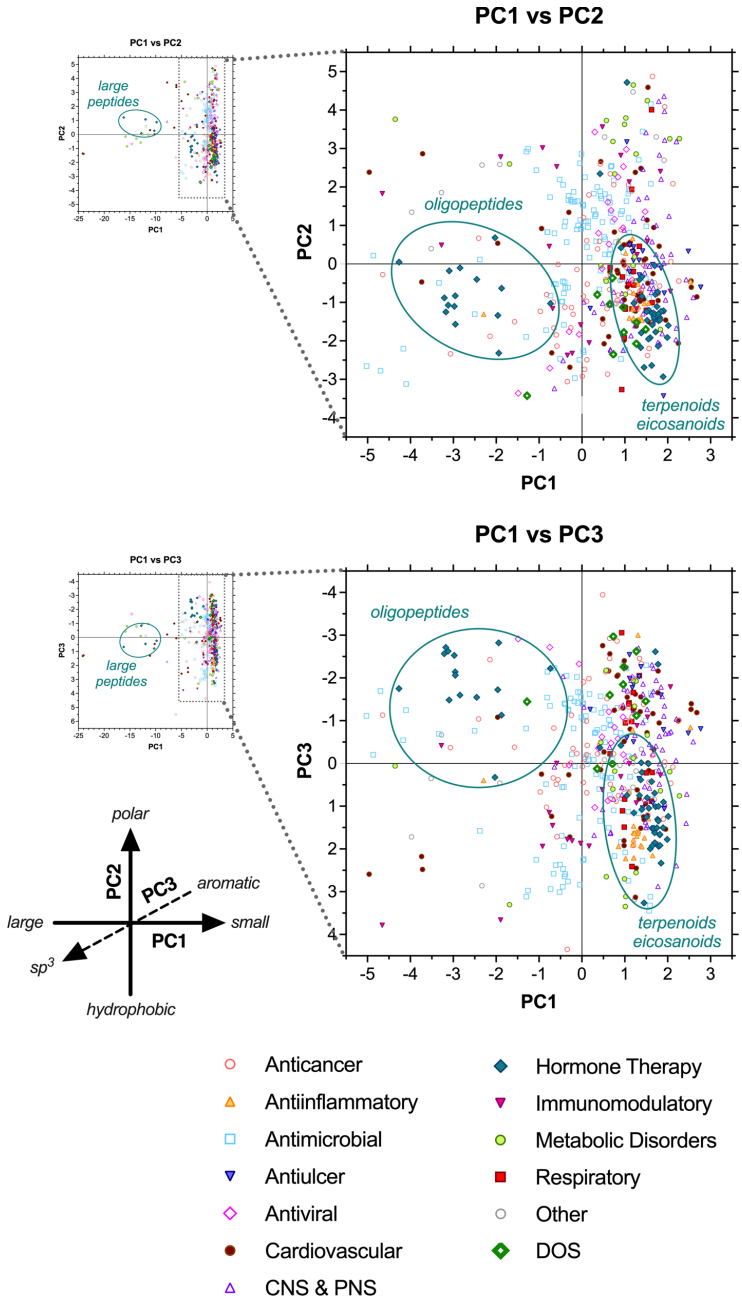

Figure S12. PCA plots of drugs coded by therapeutic class with selected hormone therapy drug subclasses labeled.

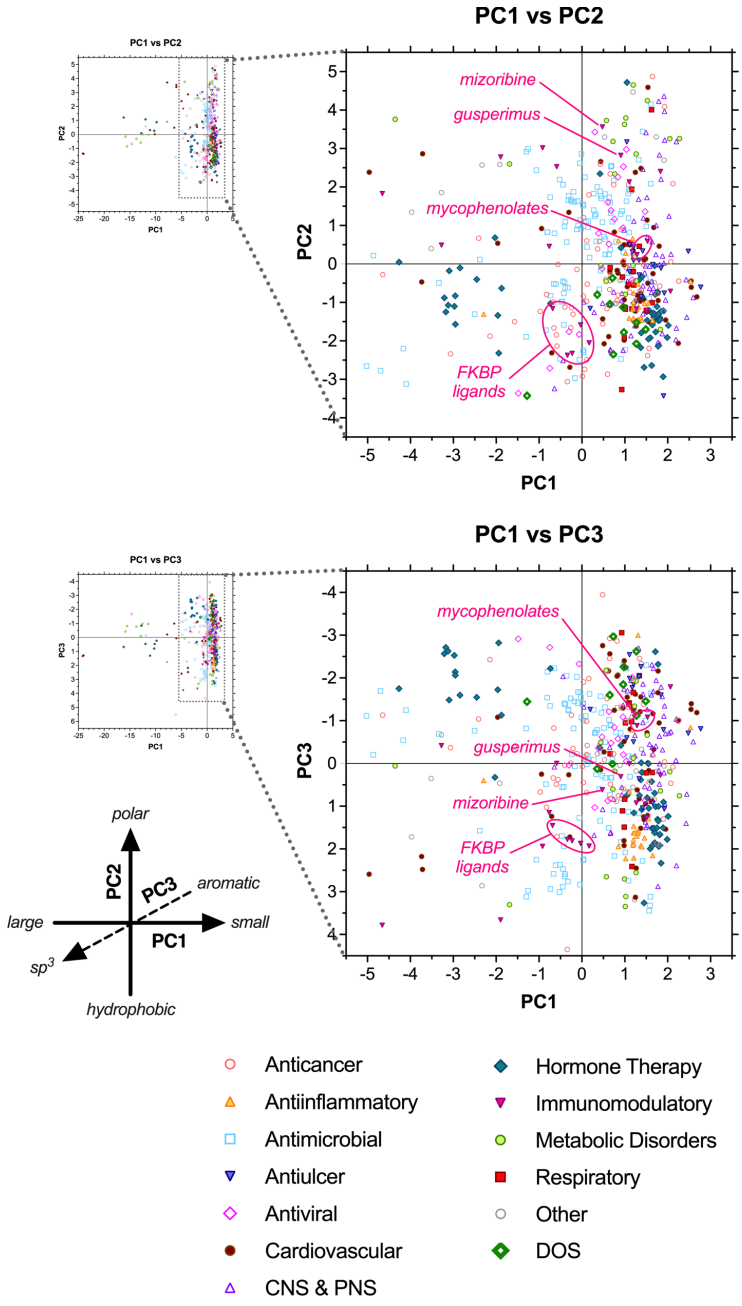

Figure S13. PCA plots of drugs coded by therapeutic class with selected immunomodulatory drugs and subclasses labeled.

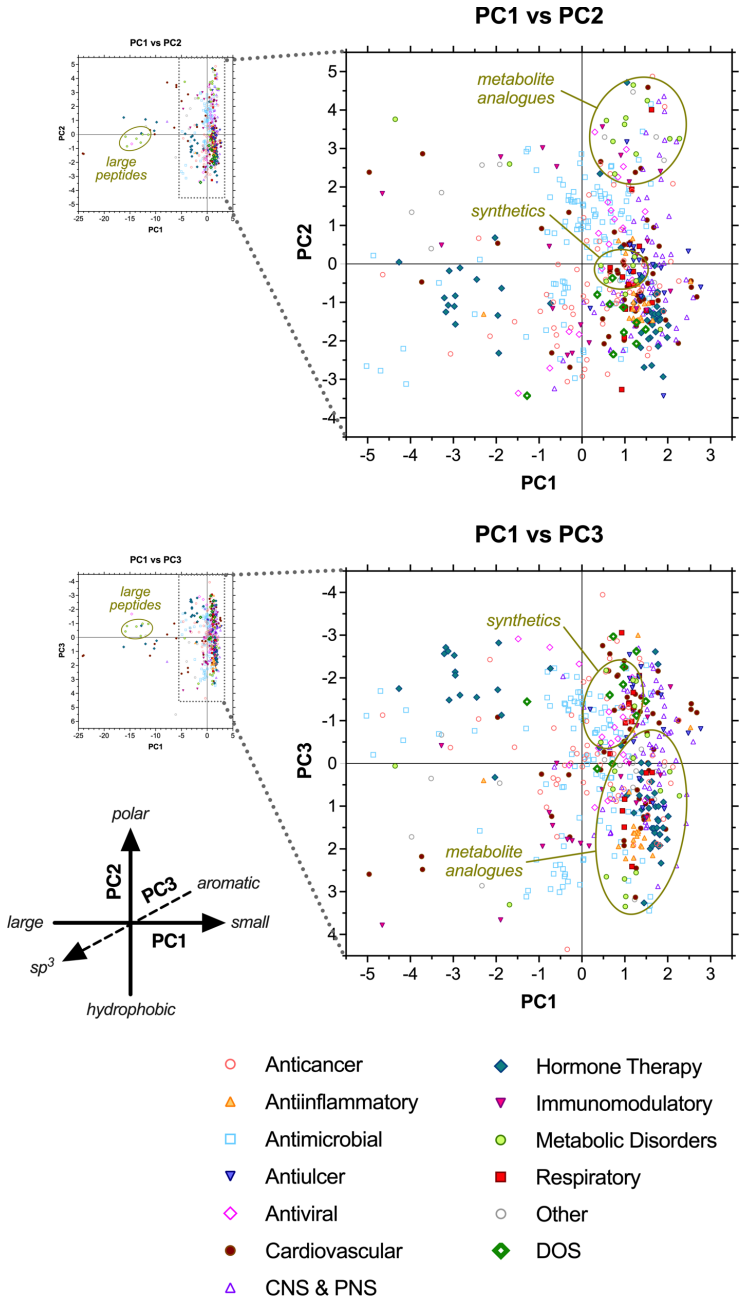

Figure S14. PCA plots of drugs coded by therapeutic class with selected metabolic disorder drug subclasses labeled.

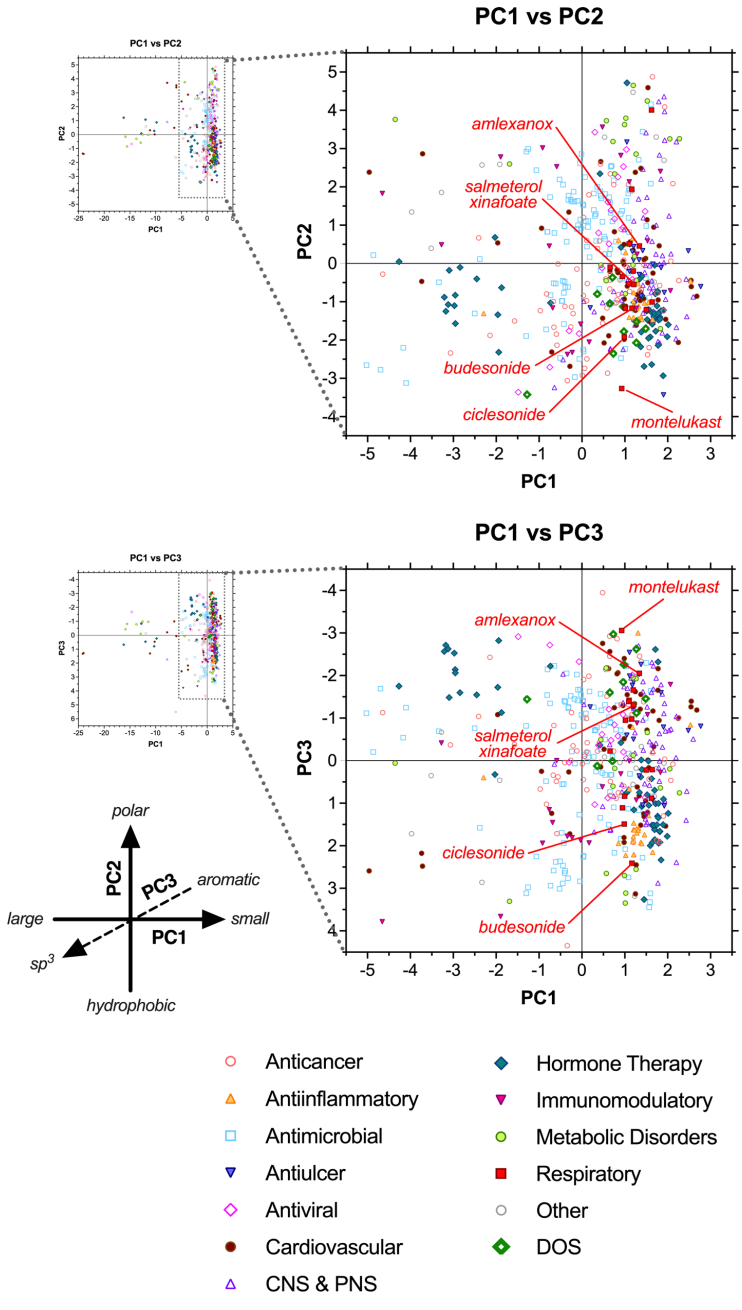

Figure S15. PCA plots of drugs coded by therapeutic class with selected respiratory drugs labeled.

B. SUPPLEMENTARY TABLES S1–S16

*See **Supplementary Dataset** spreadsheet for complete data and calculations.*

Table S1. Category counts for top 40 brand name drugs in 2006 and 2018, indicating increase in prevalence of natural product-based (NP-based) drugs between the two compound sets. Both sets of 40 drug products includes combination therapies with multiple components, resulting in more than 40 unique structures in each collection.

Category	2006 (unique structures)	%	2018 (unique structures)	%	change (2018 vs 2006)
N	**1**	2.4%	**1**	2.0%	-0.4%
ND	**3**	7.3%	**10**	20.4%	13.1%
S*	**2**	4.9%	**5**	10.2%	5.3%
S*/NM	**8**	19.5%	**18**	36.7%	17.2%
NP-based	***14***	*34.1%*	***34***	*69.4%*	*35.2%*
S	**16**	39.0%	**7**	14.3%	-24.7%
S/NM	**11**	26.8%	**8**	16.3%	-10.5%
synthetic	***27***	*65.9%*	***15***	*30.6%*	*-35.2%*
total	**41**	**100%**	**49**	**100%**	

Table S2. Complete list of parameter averages by category. MW = molecular weight; HBD = hydrogen-bond donors; HBA = hydrogen-bond acceptors; ALOGPs = 1-octanol/water partition coefficient; LogD = 1-octanol/water distribution coefficient at pH 7.4; ALOGpS = log aqueous solubility (mol/L); Rot = rotatable bonds; tPSA = topological polar surface area; VWSA = vander Waals surface area; rPSA = relative polar surface area; N = nitrogen atoms; O = oxygen atoms; Fsp3 = fraction sp^3 carbons; Stereo = stereocenter count; Stereo/HA = stereocenters ÷ heavy atoms; Rings = ring count; RngLg = largest ring; RngSys = ring systems; RRSys = rings per ring system; RngAr = aromatic rings. **(continues on following page)**

Category	Total Cmpds	Unique Structs	MW	HBD	HBA	ALOGPs	LogD	ALOGpS	Rot	tPSA	VWSA
N	**88**	**77**	**611**	**5.9**	**10.1**	**1.96**	**-1.40**	**-3.49**	**11.0**	**196**	**913**
std dev			*538*	*9.0*	*10.8*	*2.89*	*8.28*	*1.82*	*13.0*	*241*	*758*
ND	**379**	**344**	**757**	**7.0**	**11.5**	**1.82**	**-3.00**	**-3.96**	**16.2**	**250**	**1093**
std dev			*934*	*14.3*	*15.8*	*2.43*	*11.06*	*1.43*	*30.1*	*416*	*1354*
2018-N	**37**	**34**	**473**	**2.4**	**6.0**	**2.11**	**1.78**	**-3.86**	**7.8**	**111**	**680**
std dev			*218*	*1.4*	*2.6*	*2.19*	*2.89*	*1.42*	*5.1*	*45*	*328*
2018-S	**17**	**15**	**444**	**1.9**	**5.1**	**2.83**	**2.49**	**-4.15**	**6.5**	**95**	**611**
std dev			*162*	*1.1*	*1.9*	*2.23*	*2.41*	*1.40*	*4.9*	*37*	*261*
2018-all	**54**	**49**	**464**	**2.2**	**5.8**	**2.33**	**2.00**	**-3.95**	**7.4**	**106**	**659**
std dev			*201*	*1.3*	*2.4*	*2.21*	*2.74*	*1.40*	*5.0*	*43*	*308*
2006-N	**15**	**14**	**367**	**2.4**	**5.0**	**2.08**	**0.40**	**-3.67**	**7.6**	**90**	**555**
std dev			*113*	*1.6*	*2.3*	*1.84*	*2.69*	*1.52*	*3.7*	*41*	*160*
2006-S	**30**	**27**	**355**	**1.1**	**3.9**	**3.15**	**2.37**	**-4.05**	**5.4**	**61**	**503**
std dev			*73*	*0.7*	*1.5*	*1.53*	*1.70*	*1.35*	*2.9*	*27*	*109*
2006-all	**45**	**41**	**359**	**1.5**	**4.3**	**2.78**	**1.70**	**-3.92**	**6.1**	**70**	**521**
std dev			*87*	*1.2*	*1.8*	*1.70*	*2.27*	*1.40*	*3.4*	*35*	*129*
DOS	**10**	**10**	**552**	**1.1**	**4.7**	**4.08**	**3.90**	**-4.96**	**4.9**	**85**	**807**
std dev			*252*	*0.9*	*2.9*	*1.22*	*1.37*	*0.91*	*2.4*	*66*	*397*
Overall	**576**	**521**	**673**	**5.8**	**10.1**	**2.01**	**-1.79**	**-3.90**	**13.6**	**211**	**975**
std dev			*801*	*12.3*	*13.7*	*2.44*	*9.78*	*1.50*	*25.3*	*357*	*1159*

Table S2 (continued).

Category	rPSA	N	O	Fsp3	Stereo	Stereo /HA	Rings	Rng Lg	Rng Sys	RRSys	Rng Ar
N	**0.20**	**4.2**	**10.0**	**0.71**	**8.4**	**0.19**	**3.2**	**8.8**	**1.9**	**1.7**	**0.7**
std dev	*0.12*	*11.5*	*12.0*	*0.20*	*7.7*	*0.13*	*2.5*	*9.5*	*1.6*	*1.6*	*1.3*
ND	**0.21**	**6.8**	**10.9**	**0.59**	**7.7**	**0.15**	**3.9**	**7.9**	**2.2**	**2.1**	**1.4**
std dev	*0.10*	*15.4*	*16.6*	*0.20*	*9.0*	*0.09*	*2.3*	*7.0*	*1.6*	*1.5*	*1.7*
2018-N	**0.18**	**3.3**	**4.9**	**0.50**	**3.7**	**0.10**	**3.6**	**6.9**	**2.5**	**1.5**	**1.9**
std dev	*0.08*	*2.7*	*3.2*	*0.17*	*4.0*	*0.09*	*2.0*	*5.4*	*1.6*	*1.0*	*1.5*
2018-S	**0.17**	**4.2**	**3.1**	**0.33**	**1.2**	**0.04**	**3.7**	**6.0**	**2.7**	**1.4**	**2.7**
std dev	*0.07*	*2.3*	*2.0*	*0.13*	*1.5*	*0.03*	*1.8*	*0.0*	*1.2*	*0.3*	*1.4*
2018-all	**0.18**	**3.6**	**4.4**	**0.45**	**3.0**	**0.08**	**3.7**	**6.7**	**2.6**	**1.5**	**2.2**
std dev	*0.08*	*2.6*	*3.0*	*0.18*	*3.6*	*0.08*	*2.0*	*4.5*	*1.5*	*0.9*	*1.5*
2006-N	**0.17**	**1.8**	**4.2**	**0.54**	**2.3**	**0.08**	**2.4**	**5.6**	**1.6**	**1.6**	**1.6**
std dev	*0.12*	*1.5*	*2.2*	*0.25*	*2.6*	*0.08*	*1.2*	*1.6*	*0.9*	*1.0*	*1.3*
2006-S	**0.13**	**2.4**	**2.4**	**0.33**	**0.8**	**0.03**	**3.0**	**6.1**	**2.3**	**1.3**	**2.3**
std dev	*0.07*	*1.3*	*1.3*	*0.13*	*0.7*	*0.03*	*0.9*	*0.3*	*0.7*	*0.4*	*0.8*
2006-all	**0.14**	**2.2**	**3.0**	**0.40**	**1.3**	**0.05**	**2.8**	**5.9**	**2.1**	**1.4**	**2.0**
std dev	*0.09*	*1.4*	*1.9*	*0.20*	*1.8*	*0.06*	*1.1*	*1.0*	*0.9*	*0.6*	*1.1*
DOS	**0.10**	**2.8**	**4.5**	**0.38**	**3.8**	**0.10**	**5.2**	**10.2**	**3.4**	**1.6**	**2.8**
std dev	*0.04*	*1.8*	*3.8*	*0.15*	*2.1*	*0.05*	*1.0*	*9.2*	*0.7*	*0.4*	*0.9*
Overall	**0.19**	**5.7**	**9.4**	**0.58**	**6.8**	**0.14**	**3.7**	**7.8**	**2.2**	**1.9**	**1.5**
std dev	*0.10*	*13.4*	*14.6*	*0.21*	*8.3*	*0.10*	*2.2*	*7.1*	*1.6*	*1.4*	*1.6*

Table S3. Parameter averages by category, excluding vs only ribosomal peptides. **(continues on following page)**

Category	Total Cmpds	Unique Structs	MW	HBD	HBA	ALOG Ps	LogD	ALOG pS	Rot	tPSA	VWSA
N											
excluding	**84**	**73**	**537**	**4.5**	**8.7**	**2.11**	**-0.24**	**-3.45**	**9.4**	**160**	**811**
std dev			*392*	*5.7*	*8.6*	*2.88*	*6.13*	*1.86*	*10.5*	*162*	*560*
only	**4**	**4**	**1953**	**30.5**	**34.5**	**-0.87**	**-22.59**	**-4.22**	**41.5**	**843**	**2771**
std dev			*1062*	*20.1*	*17.6*	*1.05*	*14.29*	*0.79*	*19.5*	*503*	*1480*
ND											
excluding	**340**	**305**	**519**	**3.2**	**7.8**	**2.10**	**-0.55**	**-3.89**	**8.5**	**145**	**747**
std dev			*284*	*3.5*	*6.4*	*2.36*	*5.91*	*1.49*	*7.0*	*120*	*397*
only	**39**	**39**	**2616**	**36.6**	**40.9**	**-0.40**	**-22.16**	**-4.50**	**76.6**	**1076**	**3798**
std dev			*1798*	*26.9*	*30.2*	*1.75*	*19.98*	*0.73*	*59.6*	*812*	*2612*
2018-N											
excluding	**30**	**28**	**425**	**2.4**	**5.8**	**1.83**	**1.38**	**-3.69**	**7.1**	**103**	**617**
std dev			*179*	*1.4*	*2.6*	*2.21*	*2.88*	*1.47*	*4.6*	*42*	*285*
only	**7**	**6**	**695**	**2.5**	**7.3**	**3.41**	**3.68**	**-4.62**	**10.7**	**145**	**976**
std dev			*263*	*1.2*	*2.3*	*1.72*	*2.23*	*0.79*	*6.5*	*45*	*381*
2006-N											
excluding	**14**	**13**	**363**	**2.4**	**5.0**	**2.16**	**0.43**	**-3.60**	**7.5**	**89**	**548**
std dev			*116*	*1.7*	*2.3*	*1.90*	*2.80*	*1.56*	*3.8*	*43*	*165*
only	**1**	**1**	**424**	**2.0**	**5.0**	**1.14**	**0.07**	**-4.61**	**10.0**	**95**	**638**
std dev			*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
Overall											
excluding	**525**	**471**	**501**	**3.1**	**7.4**	**2.21**	**0.00**	**-3.84**	**8.2**	**135**	**727**
std dev			*288*	*3.7*	*6.4*	*2.39*	*5.53*	*1.54*	*7.2*	*120*	*409*
only	**51**	**50**	**2288**	**31.4**	**35.6**	**0.05**	**-18.65**	**-4.49**	**64.6**	**926**	**3314**
std dev			*1749*	*27.0*	*29.4*	*2.09*	*20.05*	*0.73*	*58.0*	*796*	*2543*

Table S3 (continued).

Category	rPSA	N	O	Fsp3	Stereo	Stereo /HA	Rings	Rng Lg	Rng Sys	RRSys	Rng Ar
N											
excluding	**0.19**	**3.0**	**9.3**	**0.72**	**8.0**	**0.19**	**3.1**	**7.7**	**1.8**	**1.8**	**0.6**
std dev	*0.12*	*10.0*	*11.5*	*0.20*	*7.5*	*0.13*	*2.5*	*7.0*	*1.5*	*1.6*	*1.2*
only	**0.30**	**26.8**	**23.8**	**0.56**	**16.0**	**0.12**	**4.0**	**27.8**	**3.5**	**1.3**	**2.5**
std dev	*0.02*	*16.3*	*13.9*	*0.08*	*8.1*	*0.01*	*0.0*	*25.3*	*1.0*	*0.5*	*1.0*
ND											
excluding	**0.20**	**3.6**	**7.9**	**0.60**	**5.8**	**0.15**	**3.7**	**7.2**	**1.9**	**2.2**	**1.1**
std dev	*0.10*	*10.7*	*11.6*	*0.21*	*5.3*	*0.09*	*2.1*	*5.2*	*1.2*	*1.5*	*1.3*
only	**0.27**	**32.2**	**34.2**	**0.55**	**22.1**	**0.11**	**5.9**	**13.4**	**4.9**	**1.2**	**4.1**
std dev	*0.04*	*22.1*	*28.0*	*0.10*	*16.4*	*0.02*	*2.4*	*14.1*	*2.1*	*0.2*	*1.9*
2018-N											
excluding	**0.19**	**2.8**	**4.6**	**0.51**	**3.6**	**0.11**	**3.2**	**6.7**	**2.1**	**1.5**	**1.7**
std dev	*0.08*	*2.4*	*3.3*	*0.18*	*4.2*	*0.09*	*1.7*	*5.6*	*1.2*	*1.1*	*1.3*
only	**0.15**	**5.7**	**6.3**	**0.45**	**4.2**	**0.08**	**5.8**	**8.0**	**4.5**	**1.4**	**3.2**
std dev	*0.02*	*2.7*	*2.0*	*0.13*	*3.4*	*0.06*	*2.3*	*4.9*	*1.9*	*0.5*	*1.6*
2006-N											
excluding	**0.18**	**1.8**	**4.2**	**0.55**	**2.3**	**0.08**	**2.4**	**5.5**	**1.5**	**1.6**	**1.5**
std dev	*0.12*	*1.6*	*2.3*	*0.25*	*2.8*	*0.09*	*1.3*	*1.7*	*1.0*	*1.0*	*1.4*
only	**0.15**	**2.0**	**5.0**	**0.38**	**2.0**	**0.06**	**3.0**	**7.0**	**2.0**	**1.5**	**2.0**
std dev	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
Overall											
excluding	**0.19**	**3.3**	**7.3**	**0.58**	**5.5**	**0.14**	**3.5**	**7.2**	**2.0**	**2.0**	**1.2**
std dev	*0.11*	*9.5*	*10.6*	*0.22*	*5.6*	*0.10*	*2.1*	*5.4*	*1.2*	*1.4*	*1.4*
only	**0.26**	**27.9**	**29.5**	**0.54**	**19.1**	**0.11**	**5.7**	**13.8**	**4.7**	**1.2**	**3.8**
std dev	*0.05*	*22.0*	*26.8*	*0.11*	*16.0*	*0.03*	*2.3*	*14.7*	*2.0*	*0.3*	*1.9*

Table S4. Standard deviation and percent contribution for each principal component in the PCA. The first three principal components represent 77.4% of the variance in the entire 20-dimensional dataset. The first six principal components represent >90% of the variance in the entire dataset.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	3.147	1.714	1.560	1.158	0.901	0.832	0.738	0.497	0.441	0.427
Proportion of Variance	0.502	0.149	0.123	0.068	0.041	0.035	0.028	0.013	0.010	0.009
Cumulative Proportion	0.502	0.651	0.775	0.843	0.884	0.919	0.947	0.959	0.969	0.978

Table S5. Component loadings for PC1–PC3. The five parameters that contribute the greatest to the variance along each principal component axis are highlighted.

Parameter	PC1	PC2	PC3
MW	-0.307	-0.082	0.001
N	-0.262	0.075	-0.066
O	-0.280	0.042	0.062
HBD	-0.304	0.021	0.008
HBA	-0.311	0.021	0.033
Rot	-0.288	-0.022	-0.031
Stereo	-0.279	-0.100	0.238
tPSA	-0.310	0.016	0.009
Rings	-0.148	-0.334	-0.041
RngAr	-0.149	-0.148	-0.444
RngSys	-0.214	-0.124	-0.197
RngLg	-0.151	-0.189	0.114
RRSys	0.035	-0.295	0.239
Fsp3	-0.004	-0.018	0.558
LogD	0.279	-0.195	-0.034
VWSA	-0.302	-0.098	0.014
rPSA	-0.123	0.448	-0.017
Stereo/HA	-0.025	-0.106	0.542
ALOGPs	0.117	-0.462	-0.052
ALOGpS	0.009	0.478	0.122

Table S6. Parameter averages by biosynthetic origin. **(continues on following page)**

Category	Total Cmpds	Unique Structs	MW	HBD	HBA	ALOGPs	LogD	ALOGpS	Rot	tPSA	VWSA
Alkaloid	**54**	**52**	**402**	**2.0**	**4.6**	**2.07**	**0.34**	**-3.88**	**7.2**	**78**	**631**
std dev			*178*	*1.0*	*2.3*	*2.09*	*2.16*	*1.50*	*5.5*	*32*	*319*
Carbohyd	**71**	**59**	**840**	**8.6**	**17.6**	**0.83**	**-5.45**	**-2.81**	**14.0**	**308**	**1171**
std dev			*449*	*6.3*	*9.8*	*2.42*	*9.42*	*1.62*	*10.2*	*192*	*581*
(Aminoglyc)	**51**	**41**	**849**	**8.3**	**17.1**	**1.16**	**-5.10**	**-3.08**	**13.6**	**301**	**1201**
std dev			*426*	*6.1*	*8.8*	*2.42*	*9.50*	*1.44*	*7.8*	*181*	*533*
Fatty Acid	**35**	**34**	**485**	**2.9**	**5.5**	**4.59**	**3.05**	**-5.11**	**17.5**	**108**	**806**
std dev			*272*	*3.7*	*5.7*	*1.76*	*3.79*	*1.17*	*12.2*	*104*	*439*
(Eicosanoid)	**19**	**19**	**409**	**2.4**	**4.6**	**3.96**	**2.70**	**-4.58**	**12.2**	**88**	**662**
std dev			*50*	*0.7*	*1.3*	*0.98*	*1.72*	*0.51*	*3.1*	*22*	*70*
Nucleoside	**24**	**22**	**399**	**2.6**	**7.5**	**0.19**	**-0.73**	**-2.26**	**7.5**	**143**	**531**
std dev			*165*	*1.7*	*3.2*	*1.72*	*3.84*	*1.14*	*6.2*	*73*	*209*
(Nucleobase)	**2**	**2**	**286**	**0.0**	**4.5**	**0.07**	**0.31**	**-1.66**	**4.5**	**76**	**421**
std dev			*28*	*0.0*	*0.7*	*1.43*	*1.13*	*0.56*	*3.5*	*1*	*83*
Peptide	**149**	**140**	**1269**	**14.6**	**20.2**	**0.70**	**-8.43**	**-4.07**	**30.4**	**478**	**1796**
std dev			*1323*	*20.8*	*21.6*	*2.02*	*14.77*	*1.05*	*43.5*	*589*	*1944*
(β-Lactam)	**58**	**56**	**487**	**3.1**	**9.0**	**0.07**	**-4.49**	**-3.30**	**7.8**	**175**	**590**
std dev			*114*	*1.3*	*2.5*	*1.13*	*2.95*	*0.87*	*2.9*	*47*	*133*
(NRPS)	**40**	**34**	**1059**	**8.9**	**15.7**	**2.68**	**0.08**	**-4.72**	**17.4**	**320**	**1550**
std dev			*391*	*7.6*	*8.3*	*1.79*	*6.92*	*0.94*	*12.3*	*177*	*506*
(Ribosomal)	**51**	**50**	**2288**	**31.4**	**35.6**	**0.05**	**-18.65**	**-4.49**	**64.6**	**926**	**3314**
std dev			*1749*	*27.0*	*29.4*	*2.09*	*20.05*	*0.73*	*58.0*	*796*	*2543*
Polyketide	**63**	**61**	**748**	**4.0**	**11.5**	**3.12**	**2.08**	**-4.38**	**9.5**	**192**	**1118**
std dev			*260*	*3.0*	*5.2*	*1.61*	*4.56*	*1.01*	*6.4*	*98*	*381*
(Aromatic)	**13**	**13**	**596**	**5.1**	**10.9**	**2.06**	**0.28**	**-3.42**	**5.5**	**188**	**816**
std dev			*148*	*1.2*	*2.3*	*1.65*	*3.63*	*0.87*	*3.0*	*30*	*225*
(Flavonoid)	**4**	**4**	**659**	**4.5**	**14.0**	**1.77**	**-4.13**	**-4.12**	**12.0**	**236**	**834**
std dev			*422*	*4.7*	*12.6*	*3.14*	*11.33*	*1.41*	*8.3*	*247*	*420*
Terpenoid	**138**	**110**	**420**	**1.3**	**3.8**	**3.74**	**3.31**	**-4.68**	**4.8**	**70**	**647**
std dev			*132*	*1.2*	*2.0*	*1.57*	*2.31*	*0.95*	*3.9*	*39*	*193*
(Steroid)	**91**	**70**	**417**	**1.1**	**3.6**	**3.35**	**3.14**	**-4.78**	**3.8**	**69**	**633**
std dev			*90*	*1.0*	*1.5*	*1.41*	*2.18*	*0.87*	*3.0*	*27*	*137*
(Terpene)	**4**	**4**	**187**	**0.0**	**0.0**	**4.70**	**4.09**	**-4.25**	**0.0**	**0**	**348**
std dev			*34*	*0.0*	*0.0*	*1.21*	*0.91*	*0.92*	*0.0*	*0*	*62*

Category	Total Cmpds	Unique Structs	MW	HBD	HBA	ALOG Ps	LogD	ALOG pS	Rot	tPSA	VWSA
Other	**61**	**58**	**359**	**2.8**	**5.8**	**0.83**	**-1.81**	**-2.93**	**7.1**	**113**	**520**
std dev			*273*	*2.2*	*4.1*	*2.22*	*3.50*	*1.91*	*5.6*	*79*	*373*
(Amino Acid)	**18**	**17**	**241**	**3.0**	**4.5**	**-0.32**	**-3.13**	**-2.24**	**6.6**	**96**	**356**
std dev			*87*	*1.5*	*1.5*	*2.07*	*3.05*	*1.39*	*3.8*	*36*	*137*
(Porphyrin)	**5**	**5**	**796**	**4.8**	**11.0**	**2.59**	**-3.81**	**-4.74**	**13.0**	**246**	**1089**
std dev			*287*	*2.5*	*4.3*	*0.82*	*3.98*	*0.23*	*7.5*	*128*	*428*
(Folate)	**5**	**5**	**353**	**4.4**	**7.6**	**0.34**	**-1.91**	**-3.27**	**5.8**	**144**	**475**
std dev			*106*	*1.8*	*2.5*	*2.16*	*3.86*	*1.20*	*3.8*	*53*	*143*
(Shikimate)	**10**	**9**	**530**	**3.2**	**9.2**	**2.33**	**0.78**	**-3.61**	**9.9**	**149**	**725**
std dev			*451*	*2.8*	*7.0*	*1.59*	*3.13*	*1.08*	*9.3*	*110*	*566*

Table S6 (continued).

Category	rPSA	N	O	Fsp3	Stereo	Stereo /HA	Rings	Rng Lg	Rng Sys	RRSys	Rng Ar
Alkaloid	**0.14**	**2.1**	**4.0**	**0.52**	**2.9**	**0.10**	**3.9**	**5.5**	**1.7**	**2.3**	**1.8**
std dev	*0.07*	*1.2*	*2.7*	*0.17*	*2.5*	*0.07*	*2.7*	*1.9*	*1.0*	*1.5*	*1.2*
Carbohyd	**0.27**	**9.5**	**22.0**	**0.73**	**14.6**	**0.26**	**4.5**	**9.4**	**3.1**	**1.6**	**1.0**
std dev	*0.11*	*25.1*	*23.8*	*0.24*	*7.6*	*0.09*	*2.9*	*5.8*	*1.5*	*1.2*	*1.8*
(Aminoglyc)	**0.26**	**11.0**	**22.0**	**0.72**	**14.4**	**0.25**	**4.5**	**9.5**	**3.0**	**1.5**	**1.0**
std dev	*0.11*	*27.4*	*25.8*	*0.24*	*5.8*	*0.09*	*3.2*	*4.2*	*1.3*	*0.8*	*2.0*
Fatty Acid	**0.12**	**0.7**	**6.0**	**0.70**	**4.4**	**0.12**	**1.4**	**4.3**	**1.2**	**1.0**	**0.4**
std dev	*0.05*	*1.5*	*5.6*	*0.17*	*4.9*	*0.08*	*1.2*	*2.5*	*1.1*	*0.8*	*0.6*
(Eicosanoid)	**0.13**	**0.2**	**5.1**	**0.69**	**4.7**	**0.16**	**1.8**	**5.6**	**1.4**	**1.4**	**0.5**
std dev	*0.03*	*0.4*	*1.3*	*0.14*	*0.9*	*0.03*	*0.7*	*0.5*	*0.5*	*0.7*	*0.5*
Nucleoside	**0.28**	**4.5**	**6.0**	**0.54**	**3.0**	**0.11**	**2.9**	**6.0**	**2.2**	**1.4**	**1.9**
std dev	*0.10*	*1.3*	*5.0*	*0.12*	*2.7*	*0.08*	*0.9*	*0.3*	*0.9*	*0.4*	*0.8*
(Nucleobase)	**0.18**	**4.0**	**3.5**	**0.57**	**0.0**	**0.00**	**2.5**	**6.0**	**1.5**	**1.8**	**2.0**
std dev	*0.04*	*0.0*	*0.7*	*0.04*	*0.0*	*0.00*	*0.7*	*0.0*	*0.7*	*0.4*	*0.0*
Peptide	**0.26**	**13.7**	**16.9**	**0.52**	**11.0**	**0.12**	**4.8**	**11.3**	**3.4**	**1.5**	**2.4**
std dev	*0.07*	*17.1*	*19.1*	*0.15*	*12.1*	*0.05*	*2.5*	*10.8*	*1.8*	*0.8*	*2.0*
(β-Lactam)	**0.30**	**5.4**	**6.6**	**0.44**	**2.9**	**0.10**	**3.6**	**5.7**	**2.6**	**1.4**	**1.4**
std dev	*0.05*	*2.4*	*1.8*	*0.14*	*1.1*	*0.05*	*1.0*	*0.6*	*0.8*	*0.2*	*0.8*
(NRPS)	**0.20**	**6.3**	**15.2**	**0.62**	**12.5**	**0.17**	**5.4**	**16.7**	**2.9**	**2.0**	**2.1**
std dev	*0.06*	*5.0*	*6.8*	*0.14*	*4.8*	*0.05*	*3.4*	*8.6*	*1.6*	*1.5*	*2.3*
(Ribosomal)	**0.26**	**27.9**	**29.5**	**0.54**	**19.1**	**0.11**	**5.7**	**13.8**	**4.7**	**1.2**	**3.8**
std dev	*0.05*	*22.0*	*26.8*	*0.11*	*16.0*	*0.03*	*2.3*	*14.7*	*2.0*	*0.3*	*1.9*

Category	rPSA	N	O	Fsp3	Stereo	Stereo /HA	Rings	Rng Lg	Rng Sys	RRSys	Rng Ar
Polyketide	**0.17**	**2.2**	**11.5**	**0.64**	**10.8**	**0.20**	**4.4**	**14.0**	**2.5**	**2.2**	**1.1**
std dev	*0.06*	*2.8*	*5.6*	*0.20*	*5.8*	*0.09*	*1.6*	*8.0*	*1.3*	*1.5*	*1.3*
(Aromatic)	**0.24**	**2.3**	**9.7**	**0.46**	**5.9**	**0.13**	**5.0**	**8.8**	**1.8**	**3.3**	**2.1**
std dev	*0.04*	*1.4*	*2.9*	*0.10*	*3.2*	*0.05*	*1.1*	*6.8*	*0.9*	*1.4*	*1.0*
(Flavonoid)	**0.23**	**0.0**	**16.0**	**0.36**	**5.3**	**0.09**	**3.8**	**6.0**	**3.0**	**1.3**	**2.5**
std dev	*0.15*	*0.0*	*16.3*	*0.17*	*6.1*	*0.12*	*1.5*	*0.0*	*1.2*	*0.2*	*0.6*
Terpenoid	**0.11**	**0.4**	**4.3**	**0.71**	**6.3**	**0.21**	**4.0**	**6.2**	**1.4**	**3.2**	**0.4**
std dev	*0.04*	*1.0*	*2.6*	*0.14*	*3.1*	*0.10*	*1.5*	*1.1*	*0.7*	*1.3*	*0.7*
(Steroid)	**0.11**	**0.4**	**4.1**	**0.73**	**7.1**	**0.24**	**4.4**	**6.0**	**1.3**	**3.8**	**0.2**
std dev	*0.04*	*1.0*	*1.9*	*0.11*	*2.0*	*0.06*	*0.9*	*0.0*	*0.7*	*1.0*	*0.4*
(Terpene)	**0.00**	**0.0**	**0.0**	**0.75**	**2.3**	**0.17**	**2.0**	**8.3**	**1.0**	**2.0**	**0.0**
std dev	*0.00*	*0.0*	*0.0*	*0.11*	*2.1*	*0.14*	*0.8*	*2.2*	*0.0*	*0.8*	*0.0*
Other	**0.25**	**2.4**	**4.9**	**0.55**	**2.1**	**0.06**	**2.1**	**4.7**	**1.2**	**1.2**	**1.1**
std dev	*0.11*	*2.4*	*3.8*	*0.26*	*3.8*	*0.08*	*2.3*	*4.5*	*1.3*	*1.3*	*1.4*
(Amino Acid)	**0.29**	**1.9**	**3.4**	**0.59**	**0.9**	**0.05**	**0.5**	**2.1**	**0.5**	**0.4**	**0.4**
std dev	*0.10*	*1.6*	*0.9*	*0.23*	*0.7*	*0.04*	*0.8*	*2.7*	*0.6*	*0.6*	*0.8*
(Porphyrin)	**0.22**	**6.2**	**8.4**	**0.40**	**4.8**	**0.07**	**5.6**	**15.8**	**1.4**	**4.5**	**3.0**
std dev	*0.04*	*3.8*	*3.8*	*0.15*	*5.9*	*0.06*	*1.3*	*0.4*	*0.9*	*1.0*	*2.0*
(Folate)	**0.31**	**5.2**	**3.4**	**0.30**	**1.2**	**0.05**	**2.6**	**6.0**	**1.8**	**1.5**	**2.4**
std dev	*0.09*	*1.1*	*2.3*	*0.15*	*1.3*	*0.07*	*0.5*	*0.0*	*0.4*	*0.4*	*0.9*
(Shikimate)	**0.21**	**1.6**	**9.0**	**0.43**	**4.1**	**0.08**	**3.3**	**6.4**	**2.2**	**1.6**	**1.7**
std dev	*0.06*	*2.1*	*7.1*	*0.17*	*6.4*	*0.08*	*2.7*	*1.3*	*1.9*	*0.8*	*0.9*

Table S7. Parameter averages for macrocycles. **(continues below)**

Category	Total Cmpds	Unique Structs	MW	HBD	HBA	ALOGPs	LogD	ALOGpS	Rot	tPSA	VWSA
Cyc Peptide	**25**	**25**	**1948**	**25.2**	**30.9**	**1.17**	**-14.64**	**-4.43**	**35.4**	**753**	**2730**
std dev			*1659*	*26.3*	*29.0*	*1.94*	*22.88*	*0.62*	*43.3*	*786*	*2337*
Macrolide	**32**	**32**	**831**	**3.3**	**12.2**	**3.87**	**4.02**	**-4.74**	**8.6**	**188**	**1303**
std dev			*116*	*1.6*	*2.1*	*1.05*	*2.56*	*0.85*	*3.9*	*38*	*186*

Table S7 (continued).

Category	rPSA	N	O	Fsp3	Stereo	Stereo/HA	Rings	Rng Lg	Rng Sys	RRSys	Rng Ar
Cyc Peptide	**0.25**	**19.9**	**27.1**	**0.55**	**18.3**	**0.14**	**6.5**	**27.4**	**3.9**	**1.8**	**3.3**
std dev	*0.06*	*21.5*	*27.0*	*0.12*	*14.6*	*0.03*	*3.6*	*13.8*	*2.1*	*0.9*	*2.3*
Macrolide	**0.14**	**1.8**	**12.5**	**0.77**	**14.4**	**0.25**	**4.4**	**19.8**	**2.4**	**2.4**	**0.5**
std dev	*0.02*	*1.5*	*2.7*	*0.14*	*3.9*	*0.07*	*1.5*	*5.8*	*0.9*	*1.9*	*1.0*

Table S8. Parameter averages by therapeutic class and subcategory. **(continues on page S33 below)**

Category	Total Cmpds	Unique Structs	MW	HBD	HBA	ALOGPs	LogD	ALOGpS	Rot	tPSA	VWSA
Anticancer	**71**	**63**	**604**	**3.7**	**8.4**	**2.40**	**1.15**	**-4.07**	**9.8**	**160**	**867**
std dev			*326*	*3.6*	*5.5*	*2.04*	*4.16*	*1.41*	*10.3*	*111*	*474*
max			*1682*	*21.0*	*29.0*	*6.54*	*7.57*	*0.18*	*54.0*	*639*	*2173*
min			*131*	*0.0*	*1.0*	*-2.85*	*-16.61*	*-6.32*	*0.0*	*34*	*203*
Anticancer	70	62	610	3.7	8.5	2.44	1.18	-4.10	10.0	161	875
Myelodysplasia	1	1	259	2.0	4.0	-0.43	-0.71	-2.05	1.0	92	340
Antiinflammatory	**26**	**24**	**485**	**1.6**	**5.1**	**3.20**	**2.65**	**-4.53**	**6.6**	**108**	**695**
std dev			*181*	*1.7*	*2.7*	*1.16*	*2.20*	*0.81*	*4.9*	*78*	*260*
max			*1270*	*9.0*	*17.0*	*6.59*	*5.06*	*-2.38*	*26.0*	*460*	*1842*
min			*278*	*1.0*	*2.0*	*0.84*	*-4.00*	*-6.04*	*2.0*	*37*	*463*
Antiallergic	2	2	373	1.0	3.5	4.84	3.47	-5.44	10.0	68	581
Antiarthritic	8	6	559	3.0	7.2	2.09	0.14	-3.71	7.5	166	769
Antiinflammatory	14	14	475	1.1	4.5	3.43	3.59	-4.74	5.9	92	685
Macular Degeneration	1	1	386	1.0	4.0	3.33	2.62	-4.53	4.0	80	586
Ocular Inflammation	1	1	509	1.0	5.0	3.28	3.00	-4.72	8.0	106	729
Antimicrobial	**116**	**114**	**602**	**4.6**	**10.6**	**1.04**	**-2.90**	**-3.43**	**8.6**	**194**	**823**
std dev			*326*	*4.8*	*5.8*	*2.00*	*4.99*	*1.13*	*5.5*	*119*	*471*
max			*1880*	*24.0*	*41.0*	*5.59*	*7.10*	*-0.47*	*35.0*	*748*	*2404*
min			*138*	*0.0*	*3.0*	*-2.93*	*-22.12*	*-6.12*	*0.0*	*45*	*179*
Antibacterial	96	94	613	4.8	10.9	0.77	-3.60	-3.39	9.0	203	827
Antifungal	3	3	1168	15.3	20.7	1.07	-6.79	-4.06	18.3	439	1652
Antiparasitic	11	11	398	1.5	5.9	2.78	2.28	-3.40	4.0	85	590
β-Lactamase Inhibitor	2	2	439	2.0	8.0	-0.51	-4.35	-2.09	5.0	157	547
Genital Warts	1	1	414	1.0	7.0	2.37	1.62	-3.56	4.0	92	567
Insecticide	2	2	739	0.0	10.0	4.12	3.91	-5.51	9.0	111	1199
PCP/Toxoplasmosis	1	1	369	3.0	8.0	2.36	1.91	-4.08	6.0	117	545
Antiulcer	**18**	**18**	**389**	**1.5**	**4.4**	**3.42**	**2.66**	**-4.27**	**9.0**	**80**	**600**
std dev			*108*	*1.2*	*1.7*	*2.43*	*4.18*	*1.38*	*4.5*	*34*	*210*
max			*713*	*5.0*	*8.0*	*9.97*	*15.89*	*-1.50*	*19.0*	*169*	*1294*
min			*277*	*0.0*	*1.0*	*-0.64*	*-4.51*	*-7.53*	*3.0*	*17*	*317*
Antiviral	**23**	**19**	**725**	**6.2**	**10.7**	**1.24**	**-0.89**	**-3.28**	**17.4**	**238**	**1030**
std dev			*942*	*13.9*	*13.8*	*2.46*	*10.40*	*1.35*	*32.8*	*405*	*1375*
max			*4492*	*63.0*	*67.0*	*5.97*	*7.63*	*-0.70*	*151.0*	*1902*	*6529*
min			*229*	*1.0*	*4.0*	*-2.29*	*-42.00*	*-5.80*	*2.0*	*87*	*272*
Peptide	6	5	1554	15.2	20.2	3.02	-4.73	-4.75	41.0	515	2240
Aminoglycoside	3	3	372	5.3	8.0	-0.04	-2.75	-2.81	9.7	160	556

Category	Total Cmpds	Unique Structs	MW	HBD	HBA	ALOGPs	LogD	ALOGpS	Rot	tPSA	VWSA
Nucleoside	10	8	391	2.3	7.3	-0.11	0.69	-2.33	9.1	136	541
Synthetic	4	3	583	2.7	6.7	n/a	3.14	n/a	7.7	126	793
Cardiovascular	**73**	**60**	**861**	**8.7**	**14.1**	**2.49**	**-4.84**	**-4.08**	**18.9**	**309**	**1229**
std dev			*1297*	*20.0*	*23.1*	*2.46*	*17.46*	*1.43*	*31.9*	*603*	*1821*
max			*6979*	*100.0*	*116.0*	*8.20*	*8.40*	*0.23*	*166.0*	*3107*	*9806*
min			*162*	*0.0*	*1.0*	*-1.68*	*-79.00*	*-6.95*	*0.0*	*26*	*239*
Antiarrhythmic	1	1	267	4.0	8.0	-1.21	-2.09	-1.28	2.0	139	334
Anticoagulant	16	7	1876	23.1	41.1	-0.41	-28.54	-2.51	39.0	895	2444
Antihypertensive	11	9	455	3.2	6.2	1.83	-0.09	-4.21	10.1	125	674
Antithrombotic	7	7	1734	21.6	28.0	2.68	-13.48	-4.58	46.6	680	2506
Cardiotonic	3	3	429	1.7	6.0	2.79	0.27	-3.88	11.0	107	681
Cardiovascular Disease	2	2	518	4.5	9.5	2.10	1.53	-3.74	8.5	156	731
Congestive Heart Failure	2	2	3266	54.5	57.0	-1.32	-35.29	-4.37	79.0	1458	4711
Coronary Artery Disease	2	2	632	2.0	8.5	3.45	-0.11	-4.07	11.0	172	785
Dyslipidemia	1	1	361	0.0	3.0	4.86	5.28	-5.71	7.0	52	533
Hematopoiesis	0	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Hypertriglyceridemia	2	2	315	1.0	2.0	6.68	3.98	-6.14	13.5	37	522
Hypocholesterolemic	12	10	445	2.4	4.1	4.07	2.86	-4.81	9.1	87	691
Hypolipidemic	1	1	162	3.0	5.0	-0.88	-6.96	0.23	4.0	94	239
Lipoprotein Disorders	1	1	312	0.0	2.0	3.85	4.00	-4.46	0.0	34	500
Male Sexual Dysfunction	1	1	389	1.0	4.0	2.36	1.64	-3.19	1.0	74	510
Platelet Agg. Inhibitor	3	3	370	2.7	4.7	3.93	0.34	-4.24	9.3	84	573
Porphyria	1	1	563	4.0	6.0	3.44	2.25	-5.01	8.0	131	808
Restenosis	2	2	976	2.0	13.5	5.09	7.86	-5.72	9.0	206	1551
Subarachnoid Hemorrhage	1	1	625	0.0	8.0	5.45	6.55	-3.91	6.0	72	974
Vasodilator; Cerebral	2	2	370	0.5	3.5	2.30	2.60	-2.99	5.0	65	520
Vasodilator; Coronary	1	1	211	1.0	6.0	0.05	0.25	-2.71	5.0	100	288
Venous Thromboembolism	1	1	436	1.0	5.0	1.74	1.90	-4.64	5.0	88	545
CNS & PNS	**68**	**65**	**373**	**2.0**	**4.3**	**2.44**	**0.37**	**-3.72**	**6.2**	**76**	**602**
std dev			*329*	*5.1*	*5.4*	*2.38*	*5.24*	*1.78*	*6.5*	*137*	*499*
max			*2639*	*42.0*	*43.0*	*7.49*	*7.30*	*0.68*	*38.0*	*1133*	*3680*
min			*104*	*0.0*	*0.0*	*-2.56*	*-35.61*	*-7.63*	*0.0*	*0*	*170*
ADHD	3	3	211	1.7	2.0	1.44	-0.88	-2.83	4.7	48	358
Alzheimer's Disease	2	2	301	1.0	2.5	2.78	1.06	-3.98	2.5	30	464
Analgesic	7	7	312	0.9	2.3	4.08	2.89	-4.05	3.1	36	502
Anorexia	1	1	415	1.0	4.0	5.41	2.00	-5.71	9.0	72	675

Category	Total Cmpds	Unique Structs	MW	HBD	HBA	ALOG Ps	LogD	ALOG pS	Rot	tPSA	VWSA
Antiarthritic	1	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Anticonvulsant	1	1	256	2.0	5.0	1.87	1.91	-2.72	1.0	90	280
Antidepressant	6	6	294	0.8	2.2	3.94	2.33	-4.68	3.8	28	467
Antiepileptic	5	4	197	1.8	4.3	-0.99	-1.16	-1.03	4.3	76	322
Antihyperprolactinemia	4	4	397	1.8	4.0	3.07	1.39	-3.80	5.3	62	624
Antimigraine	1	1	295	2.0	3.0	1.17	-1.24	-3.37	5.0	65	468
Antinarcolepsy	1	1	104	2.0	3.0	-0.63	-3.37	0.68	3.0	57	170
Antinarcotic	2	2	340	2.0	4.5	2.16	-0.17	-2.33	2.0	61	1200
Antinauseant	1	1	314	1.0	2.0	7.29	5.94	-5.08	4.0	29	541
Antispasmodic	1	1	358	1.0	3.0	0.45	-2.49	-5.21	7.0	59	547
Constipation	2	2	373	2.0	4.5	1.68	-0.37	-4.46	6.5	75	575
Hypnotic	1	1	307	0.0	2.0	3.15	3.01	-3.99	3.0	37	481
Multiple Sclerosis	7	7	260	1.4	3.0	2.49	1.00	-3.21	6.9	62	404
Muscle Relaxant	4	4	655	0.3	5.3	1.70	-2.14	-7.32	11.0	75	1111
Neuroleptic	4	4	389	0.8	4.3	3.54	2.90	-3.93	4.0	46	561
Neuropathic Pain	4	3	1054	15.3	16.0	0.91	-10.19	-3.20	16.0	412	1516
Nootropic	4	4	293	1.3	5.0	0.33	-1.95	-2.81	7.8	90	451
Opioid induced constipation	2	2	611	3.0	10.0	2.44	0.73	-3.73	15.0	134	916
Osteoarthritis	1	1	305	2.0	3.0	3.80	3.75	-4.56	9.0	58	519
Parkinson's Disease	1	1	211	3.0	4.0	-0.20	0.47	-1.99	4.0	92	304
Schizophrenia	1	1	665	0.0	5.0	7.49	7.30	-5.00	20.0	88	1095
Uremic Pruritus	1	1	477	2.0	5.0	3.24	0.42	-3.47	5.0	86	678
Hormone Therapy	**83**	**62**	**930**	**10.6**	**12.4**	**2.54**	**-2.39**	**-4.71**	**22.9**	**313**	**1385**
std dev			*1106*	*18.0*	*18.7*	*2.34*	*11.12*	*0.96*	*40.6*	*509*	*1633*
max			*5136*	*83.0*	*87.0*	*6.74*	*8.06*	*-1.17*	*181.0*	*2339*	*7590*
min			*249*	*0.0*	*1.0*	*-2.21*	*-44.00*	*-6.81*	*0.0*	*12*	*315*
5 α-reductase inhibitor	1	1	373	2.0	4.0	4.98	2.45	-5.59	2.0	65	617
Abortifacient	3	3	430	2.0	5.0	4.00	3.67	-4.50	9.0	85	669
Anabolic Metabolism	1	1	312	1.0	2.0	2.69	3.10	-4.64	0.0	37	498
Antiprogrestogenic	1	1	308	1.0	2.0	3.39	2.85	-4.50	1.0	37	444
Antisecretory	1	1	1019	13.0	12.0	0.42	-4.28	-4.92	17.0	332	1444
Benign Prostatic Hypertrophy	1	1	529	2.0	2.0	5.45	5.79	-5.77	4.0	58	707
Calcium Metabolism	9	9	1107	14.6	16.4	2.19	-5.48	-3.90	30.6	420	1662
CNS Stimulant	1	1	405	4.0	6.0	-2.21	-3.74	-2.17	6.0	170	553
Contraception	21	13	354	0.8	2.4	3.60	3.98	-5.04	1.8	46	549
Cushing's Syndrome	1	1	1047	9.0	10.0	3.03	-1.00	-5.71	18.0	281	1500
Dyspareunia	1	1	288	1.0	2.0	3.53	3.36	-3.82	0.0	37	478
Endometriosis	1	1	311	1.0	3.0	3.37	2.31	-3.61	1.0	61	469
Hormone	10	10	1484	19.9	21.1	1.07	-7.66	-4.80	40.0	554	2152
Hormone Repl. Therapy	19	6	332	1.0	3.0	2.45	2.89	-4.51	1.5	60	504

Category	Total Cmpds	Unique Structs	MW	HBD	HBA	ALOGPs	LogD	ALOGpS	Rot	tPSA	VWSA
Hyperparathyroidism	1	1	357	1.0	1.0	5.57	3.75	-6.81	7.0	12	527
Hypogonadism	1	1	457	0.0	2.0	6.74	8.06	-6.79	11.0	43	818
Lipodystrophy Syndrome	1	1	5136	83.0	87.0	-1.39	-44.00	-4.46	181.0	2339	7590
Menopause	1	1	326	0.0	2.0	4.44	4.66	-4.63	2.0	34	532
Osteoporosis	3	3	2856	41.7	45.0	0.75	-19.78	-4.53	98.7	1193	4314
Ovulation	1	1	1570	16.0	20.0	3.42	-2.27	-5.47	44.0	451	2382
Phototoxicity in Adults	1	1	1647	23.0	24.0	-1.41	-9.81	-4.86	50.0	642	2400
Premature Birth	1	1	994	13.0	22.0	1.28	-17.81	-4.18	18.0	392	1392
Secondary Hyperthyroidism	1	1	525	3.0	3.0	5.51	5.45	-5.43	8.0	60	751
Treatment of GH Deficiency	1	1	818	9.0	8.0	2.47	-1.51	-5.72	21.0	256	1204
Immunomodulatory	**22**	**22**	**678**	**6.4**	**11.6**	**1.69**	**-0.86**	**-3.95**	**11.8**	**214**	**1036**
std dev			*328*	*5.9*	*8.0*	*3.16*	*7.22*	*1.79*	*8.6*	*145*	*505*
max			*1305*	*23.0*	*36.0*	*5.36*	*7.45*	*-0.28*	*35.0*	*585*	*2032*
min			*223*	*1.0*	*3.0*	*-3.31*	*-13.56*	*-5.83*	*3.0*	*46*	*301*
Antiallergic	1	1	810	2.0	10.0	4.36	6.81	-5.73	6.0	158	1273
Antipsoriatic	3	3	385	2.3	3.0	5.06	3.79	-4.97	5.7	55	626
Hereditary Angioedema	1	1	1305	18.0	23.0	-2.17	-12.57	-4.56	30.0	516	1853
Immunomodulator	2	2	452	8.0	9.0	-1.12	-5.44	-3.36	13.0	197	655
Immunostimulant	5	5	742	12.6	19.4	-0.97	-8.84	-2.01	18.4	342	1048
Immunosuppressant	10	10	704	3.5	9.8	2.69	3.06	-4.49	8.9	176	1123
Metabolic Disorders	**27**	**25**	**1284**	**17.2**	**21.3**	**-0.24**	**-10.72**	**-2.65**	**38.8**	**516**	**1877**
std dev			*1747*	*24.6*	*27.2*	*2.49*	*15.93*	*2.06*	*60.8*	*724*	*2572*
max			*4906*	*70.0*	*78.0*	*7.61*	*8.11*	*0.50*	*169.0*	*2060*	*7194*
min			*77*	*0.0*	*1.0*	*-2.70*	*-46.00*	*-6.73*	*0.0*	*26*	*126*
Antidiabetic	17	15	1937	26.4	31.4	-0.33	-15.76	-3.31	60.3	778	2826
Antiobesity	1	1	496	1.0	3.0	7.61	8.11	-6.73	23.0	81	925
Cystinosis	1	1	77	2.0	1.0	0.01	-2.20	-0.52	1.0	26	126
Fabry's Disease	1	1	163	5.0	5.0	-2.22	-3.63	0.50	1.0	92	236
Gaucher's Disease	1	1	219	4.0	5.0	-1.14	-2.30	0.18	4.0	84	364
Homocystinuria	1	1	117	0.0	2.0	-2.70	-3.72	-1.96	2.0	40	220
Hyperammonemia	1	1	190	4.0	5.0	-1.06	-7.92	-1.00	5.0	129	257
Hyperphenylalaninemia	1	1	241	6.0	7.0	-2.40	-2.44	-1.34	2.0	132	310
Hypocalciuric	1	1	766	6.0	22.0	-1.12	-16.16	-1.00	15.0	403	940
Inborn Errors of Bile Synthesis	1	1	409	4.0	5.0	2.26	-0.35	-3.74	4.0	97	678
Orotic Aciduria	1	1	370	1.0	6.0	-0.31	-1.09	-1.03	8.0	137	486

Category	Total Cmpds	Unique Structs	MW	HBD	HBA	ALOGPs	LogD	ALOGpS	Rot	tPSA	VWSA
Respiratory	**19**	**19**	**450**	**1.9**	**4.6**	**2.80**	**2.22**	**-4.68**	**11.1**	**85**	**685**
std dev			*145*	*1.4*	*1.2*	*2.67*	*3.62*	*1.88*	*11.2*	*23*	*288*
max			*749*	*4.0*	*6.0*	*8.16*	*10.14*	*-0.73*	*41.0*	*148*	*1406*
min			*179*	*0.0*	*2.0*	*-2.35*	*-3.15*	*-8.35*	*2.0*	*46*	*267*
Antiallergic	2	2	539	1.0	4.0	3.73	4.13	-4.09	6.0	93	698
Antiasthmatic	5	5	457	1.6	5.4	3.81	3.46	-5.03	5.6	92	669
Antihistamine	1	1	389	1.0	5.0	2.98	0.65	-3.77	8.0	53	577
Antitussive	1	1	398	1.0	2.0	1.81	-0.70	-6.23	7.0	46	606
Bronchodilator	5	5	355	2.6	4.6	1.75	0.11	-4.09	9.6	77	563
COPD	2	2	479	2.5	4.5	1.02	0.09	-5.02	10.5	75	620
Expectorant	1	1	179	3.0	4.0	-2.35	-3.15	-0.73	6.0	83	267
Respiratory Distress Syndrome	2	2	742	1.5	5.0	6.73	9.57	-7.19	40.5	130	1383
Other	**20**	**20**	**827**	**9.1**	**14.4**	**1.21**	**-7.46**	**-3.41**	**19.3**	**315**	**1157**
std dev			*865*	*12.5*	*16.4*	*2.91*	*12.79*	*1.43*	*26.4*	*389*	*1196*
max			*3752*	*55.0*	*61.0*	*5.66*	*4.83*	*-1.22*	*125.0*	*1624*	*5446*
min			*163*	*1.0*	*2.0*	*-3.15*	*-40.00*	*-5.20*	*3.0*	*37*	*217*
Antiacne	2	2	244	1.5	3.0	3.52	-0.79	-3.36	6.5	56	403
Anticirrhotic	1	1	421	3.0	4.0	3.50	1.79	-4.58	5.0	77	730
Antidote	1	1	1993	16.0	47.0	0.08	-28.17	-1.79	40.0	769	2559
Antiglaucoma	6	6	456	2.8	4.7	4.20	3.89	-4.77	14.8	95	732
Dermatological Disorders	1	1	409	4.0	5.0	2.26	-0.35	-3.74	4.0	97	678
Dry Eye Syndrome	1	1	786	6.0	18.0	-0.35	-14.82	-1.51	14.0	404	839
Gastroprotectant	1	1	1249	8.0	31.0	-1.35	-20.72	-3.31	23.0	581	1396
Irritable Bowel Syndrome	2	2	1604	21.0	24.5	-1.44	-18.41	-3.59	20.5	646	2129
Mucositis	1	1	217	4.0	5.0	-3.15	-4.58	-1.23	6.0	135	322
Short Bowel Syndrome	1	1	3752	55.0	61.0	-1.39	-40.00	-4.46	125.0	1624	5446
Urolithiasis	1	1	163	3.0	4.0	0.08	-5.35	-1.78	3.0	69	217
Vulnerary	2	2	559	10.0	15.0	-2.45	-11.01	-1.64	11.5	288	748

Table S8 (continued).

Category	rPSA	N	O	Fsp3	Stereo	Stereo/HA	Rings	Rng Lg	Rng Sys	RRSys	Rng Ar
Anticancer	**0.19**	**3.5**	**7.6**	**0.52**	**5.7**	**0.12**	**4.5**	**8.0**	**2.3**	**2.4**	**1.9**
std dev	*0.08*	*3.5*	*5.5*	*0.19*	*5.1*	*0.09*	*2.3*	*5.4*	*1.5*	*1.9*	*1.5*
max	*0.39*	*18.0*	*25.0*	*1.00*	*20.0*	*0.37*	*9.0*	*29.0*	*7.0*	*9.0*	*6.0*
min	*0.06*	*0.0*	*1.0*	*0.11*	*0.0*	*0.00*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
Anticancer	0.19	3.5	7.7	0.52	5.8	0.12	4.5	8.0	2.3	2.4	1.9
Myelodysplasia	0.27	3.0	3.0	0.31	1.0	0.05	3.0	6.0	2.0	1.5	1.0

Category	rPSA	N	O	Fsp3	Stereo	Stereo /HA	Rings	Rng Lg	Rng Sys	RRSys	Rng Ar
Antiinflammatory	**0.15**	**1.0**	**5.8**	**0.67**	**7.3**	**0.21**	**4.0**	**6.1**	**1.2**	**3.5**	**0.3**
std dev	*0.04*	*2.9*	*2.6*	*0.13*	*3.1*	*0.08*	*1.3*	*2.3*	*0.7*	*1.2*	*0.8*
max	*0.25*	*13.0*	*14.0*	*0.79*	*15.0*	*0.30*	*8.0*	*15.0*	*3.0*	*5.0*	*3.0*
min	*0.07*	*0.0*	*1.0*	*0.12*	*0.0*	*0.00*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
Antiallergic	0.11	0.0	4.5	0.66	3.5	0.11	2.0	3.0	0.5	2.0	0.0
Antiarthritic	0.20	3.7	6.3	0.56	6.7	0.17	4.3	7.5	1.8	2.9	1.2
Antiinflammatory	0.13	0.1	5.8	0.72	8.1	0.25	4.1	6.0	1.1	4.0	0.1
Macular Degeneration	0.14	0.0	5.0	0.70	5.0	0.18	4.0	6.0	1.0	4.0	0.0
Ocular Inflammation	0.15	0.0	7.0	0.70	8.0	0.22	4.0	6.0	1.0	4.0	0.0
Antimicrobial	**0.25**	**4.5**	**8.7**	**0.57**	**6.8**	**0.15**	**4.0**	**8.8**	**2.5**	**1.8**	**1.4**
std dev	*0.08*	*3.1*	*5.6*	*0.23*	*6.1*	*0.11*	*2.1*	*5.8*	*1.0*	*1.1*	*1.5*
max	*0.49*	*17.0*	*33.0*	*1.00*	*23.0*	*0.38*	*14.0*	*31.0*	*6.0*	*6.0*	*7.0*
min	*0.08*	*0.0*	*0.0*	*0.17*	*0.0*	*0.00*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
Antibacterial	0.26	4.8	9.0	0.55	6.6	0.15	4.0	8.6	2.6	1.7	1.4
Antifungal	0.26	8.7	18.3	0.57	15.3	0.19	6.0	21.0	4.0	1.6	3.0
Antiparasitic	0.16	1.5	5.0	0.67	5.9	0.19	3.5	7.6	1.5	2.6	0.8
β-Lactamase Inhibitor	0.29	5.0	6.5	0.58	2.5	0.10	3.5	5.5	2.5	1.6	1.0
Genital Warts	0.16	0.0	8.0	0.41	4.0	0.13	5.0	6.0	2.0	2.5	2.0
Insecticide	0.09	1.0	10.0	0.86	17.0	0.32	6.0	12.0	3.0	2.0	0.0
PCP/Toxoplasmosis	0.21	5.0	3.0	0.26	0.0	0.00	3.0	6.0	2.0	1.5	3.0
Antiulcer	**0.15**	**1.4**	**3.8**	**0.47**	**1.7**	**0.06**	**2.2**	**5.7**	**1.4**	**1.5**	**1.7**
std dev	*0.08*	*1.6*	*1.6*	*0.24*	*1.6*	*0.06*	*1.3*	*3.2*	*0.7*	*1.1*	*1.4*
max	*0.38*	*4.0*	*6.0*	*0.94*	*5.0*	*0.17*	*5.0*	*16.0*	*2.0*	*5.0*	*4.0*
min	*0.03*	*0.0*	*1.0*	*0.11*	*0.0*	*0.00*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
Antiviral	0.23	7.2	9.2	0.53	5.4	0.10	3.9	6.6	2.7	1.5	2.6
std dev	*0.09*	*10.8*	*13.5*	*0.12*	*8.4*	*0.05*	*2.8*	*2.8*	*2.0*	*0.4*	*2.3*
max	*0.46*	*51.0*	*64.0*	*0.75*	*39.0*	*0.22*	*10.0*	*18.0*	*8.0*	*2.3*	*9.0*
min	*0.12*	*1.0*	*1.0*	*0.30*	*1.0*	*0.03*	*1.0*	*6.0*	*1.0*	*1.0*	*0.0*
Peptide	0.18	15.4	18.6	0.53	12.6	0.11	7.0	8.4	5.0	1.5	4.6
Aminoglycoside	0.30	3.3	6.3	0.68	4.3	0.17	1.0	6.0	1.0	1.0	0.0
Nucleoside	0.27	4.6	5.6	0.53	2.3	0.09	2.5	6.0	1.9	1.5	1.9
Synthetic	0.16	4.0	6.0	0.35	3.0	0.06	5.3	6.0	3.0	1.8	3.7
Cardiovascular	**0.20**	**11.6**	**17.0**	**0.53**	**8.6**	**0.12**	**3.4**	**9.7**	**2.5**	**1.6**	**1.4**
std dev	*0.12*	*27.4*	*29.7*	*0.21*	*12.7*	*0.10*	*2.1*	*12.9*	*1.9*	*1.2*	*1.4*
max	*0.49*	*119.0*	*119.0*	*0.96*	*63.0*	*0.36*	*10.0*	*56.0*	*9.0*	*6.0*	*5.0*
min	*0.04*	*0.0*	*1.0*	*0.14*	*0.0*	*0.00*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*

Category	rPSA	N	O	Fsp3	Stereo	Stereo /HA	Rings	Rng Lg	Rng Sys	RRSys	Rng Ar
Antiarrhythmic	0.42	5.0	4.0	0.50	4.0	0.21	3.0	6.0	2.0	1.5	2.0
Anticoagulant	0.40	58.4	74.1	0.48	23.9	0.18	5.1	13.1	4.9	1.1	1.0
Antihypertensive	0.18	3.1	5.2	0.46	3.1	0.09	3.1	6.1	1.9	2.1	1.8
Antithrombotic	0.17	15.6	27.0	0.59	16.1	0.10	3.9	12.1	3.4	1.1	1.6
Cardiotonic	0.16	1.3	6.7	0.62	3.0	0.08	1.7	6.0	1.0	1.7	0.7
Cardiovascular Disease	0.20	0.5	10.0	0.48	7.0	0.17	4.0	6.0	3.5	1.1	2.0
Congestive Heart Failure	0.31	47.5	40.5	0.62	26.0	0.11	3.5	53.0	3.5	1.0	2.5
Coronary Artery Disease	0.21	3.5	8.0	0.59	6.5	0.17	4.0	6.0	2.5	1.6	2.0
Dyslipidemia	0.10	0.0	4.0	0.30	0.0	0.00	2.0	6.0	2.0	1.0	2.0
Hematopoiesis	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Hypertriglyceridemia	0.07	0.0	2.0	0.43	0.0	0.00	0.0	0.0	0.0	0.0	0.0
Hypocholesterolemic	0.13	0.9	4.6	0.57	5.0	0.17	3.0	6.0	2.3	1.6	1.4
Hypolipidemic	0.39	0.0	5.0	0.67	0.0	0.00	0.0	0.0	0.0	0.0	0.0
Lipoprotein Disorders	0.07	0.0	2.0	0.71	5.0	0.22	4.0	6.0	1.0	4.0	0.0
Male Sexual Dysfunction	0.15	3.0	4.0	0.27	2.0	0.07	6.0	6.0	2.0	3.0	3.0
Platelet Agg. Inhibitor	0.15	0.0	4.7	0.66	5.7	0.21	2.3	5.3	1.0	2.3	0.3
Porphyria	0.16	4.0	4.0	0.24	0.0	0.00	5.0	16.0	1.0	5.0	5.0
Restenosis	0.13	3.0	13.0	0.75	15.0	0.22	4.5	29.0	2.5	1.8	0.5
Subarachnoid Hemorrhage	0.07	6.0	2.0	0.68	6.0	0.13	8.0	6.0	5.0	1.6	1.0
Vasodilator; Cerebral	0.13	3.0	3.0	0.59	1.5	0.06	3.5	6.0	1.0	3.5	2.0
Vasodilator; Coronary	0.35	3.0	4.0	0.25	0.0	0.00	1.0	6.0	1.0	1.0	1.0
Venous Thromboembolism	0.16	3.0	5.0	0.32	1.0	0.03	4.0	6.0	4.0	1.0	2.0
CNS & PNS	**0.13**	**2.1**	**3.5**	**0.57**	**2.7**	**0.09**	**3.1**	**6.0**	**1.5**	**1.8**	**1.2**
std dev	*0.09*	*4.5*	*4.4*	*0.23*	*3.6*	*0.09*	*2.4*	*5.5*	*1.1*	*1.2*	*1.0*
max	*0.37*	*36.0*	*32.0*	*1.00*	*22.0*	*0.35*	*12.0*	*46.0*	*4.0*	*5.0*	*4.0*
min	*0.00*	*0.0*	*0.0*	*0.00*	*0.0*	*0.00*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
ADHD	0.13	1.7	1.0	0.46	1.7	0.11	1.3	6.0	1.3	1.0	1.0
Alzheimer's Disease	0.07	1.5	1.5	0.55	3.0	0.14	4.0	6.5	1.0	4.0	1.5
Analgesic	0.06	0.6	2.3	0.61	3.3	0.14	3.9	7.3	1.6	2.1	0.7
Anorexia	0.11	0.0	5.0	0.60	2.0	0.07	3.0	6.0	1.0	3.0	1.0
Antiarthritic	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Anticonvulsant	0.32	5.0	0.0	0.00	0.0	0.00	2.0	6.0	2.0	1.0	2.0
Antidepressant	0.06	1.0	1.2	0.49	2.3	0.11	2.7	6.0	1.7	1.8	1.5
Antiepileptic	0.24	1.0	3.5	0.81	1.8	0.12	0.8	1.5	0.3	0.8	0.0
Antihyperprolactinemia	0.10	3.8	2.0	0.52	2.8	0.10	4.0	6.0	1.3	3.4	2.0
Antimigraine	0.14	3.0	2.0	0.43	0.0	0.00	2.0	6.0	1.0	2.0	2.0
Antinarcolepsy	0.34	0.0	3.0	0.75	0.0	0.00	0.0	0.0	0.0	0.0	0.0
Antinarcotic	0.09	1.0	3.5	0.63	4.0	0.16	6.0	6.0	2.0	3.0	1.0

Category	rPSA	N	O	Fsp3	Stereo	Stereo /HA	Rings	Rng Lg	Rng Sys	RRSys	Rng Ar
Antinauseant	0.05	0.0	2.0	0.62	2.0	0.09	3.0	6.0	1.0	3.0	1.0
Antispasmodic	0.11	1.0	4.0	0.67	7.0	0.27	5.0	6.0	3.0	1.7	1.0
Constipation	0.13	0.5	4.5	0.78	4.5	0.17	4.0	6.0	1.5	2.5	0.5
Hypnotic	0.08	3.0	1.0	0.26	0.0	0.00	3.0	6.0	2.0	1.5	3.0
Multiple Sclerosis	0.20	0.9	3.0	0.42	0.0	0.00	1.0	3.4	1.0	0.6	0.7
Muscle Relaxant	0.07	2.5	6.0	0.82	8.5	0.20	6.0	6.0	3.3	1.9	1.0
Neuroleptic	0.08	3.5	1.5	0.42	0.0	0.00	4.3	6.5	2.5	1.8	2.3
Neuropathic Pain	0.19	12.3	12.0	0.65	9.0	0.14	2.7	19.0	1.7	1.7	0.7
Nootropic	0.20	0.8	5.5	0.68	0.8	0.05	1.0	3.0	0.5	1.0	0.8
Opioid induced constipation	0.15	2.5	8.5	0.62	4.5	0.10	6.5	6.0	2.5	3.5	2.0
Osteoarthritis	0.11	1.0	3.0	0.50	0.0	0.00	1.0	6.0	1.0	1.0	1.0
Parkinson's Disease	0.30	1.0	4.0	0.30	1.0	0.07	1.0	6.0	1.0	1.0	1.0
Schizophrenia	0.08	4.0	4.0	0.69	1.0	0.02	5.0	6.0	3.0	1.7	3.0
Uremic Pruritus	0.13	2.0	5.0	0.54	5.0	0.14	7.0	6.0	3.0	2.3	2.0
Hormone Therapy	**0.16**	**9.2**	**10.0**	**0.63**	**9.0**	**0.17**	**4.4**	**6.9**	**2.5**	**2.5**	**1.8**
std dev	*0.11*	*16.5*	*14.4*	*0.17*	*9.0*	*0.08*	*1.8*	*4.2*	*1.8*	*1.6*	*2.2*
max	*0.51*	*72.0*	*67.0*	*1.00*	*44.0*	*0.37*	*8.0*	*24.0*	*7.0*	*7.0*	*6.0*
min	*0.02*	*0.0*	*0.0*	*0.17*	*0.0*	*0.00*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
5 α-reductase inhibitor	0.11	2.0	2.0	0.83	7.0	0.26	4.0	6.0	1.0	4.0	0.0
Abortifacient	0.13	0.7	4.7	0.59	4.3	0.14	2.7	5.7	1.7	1.5	0.7
Anabolic Metabolism	0.07	0.0	2.0	0.76	6.0	0.26	4.0	6.0	1.0	4.0	0.0
Antiprogrestogenic	0.08	0.0	2.0	0.57	4.0	0.17	4.0	6.0	1.0	4.0	0.0
Antisecretory	0.23	10.0	10.0	0.45	10.0	0.14	5.0	20.0	4.0	1.3	4.0
Benign Prostatic Hypertrophy	0.08	2.0	2.0	0.63	7.0	0.19	5.0	6.0	2.0	2.5	1.0
Calcium Metabolism	0.22	11.0	14.1	0.72	10.2	0.12	2.9	6.7	2.2	1.1	1.2
CNS Stimulant	0.31	7.0	5.0	0.53	3.0	0.10	3.0	6.0	3.0	1.0	1.0
Contraception	0.08	0.2	2.7	0.71	6.4	0.25	4.6	6.0	1.2	4.2	0.2
Cushing's Syndrome	0.19	10.0	9.0	0.33	7.0	0.09	8.0	18.0	6.0	1.3	6.0
Dyspareunia	0.08	0.0	2.0	0.84	6.0	0.29	4.0	6.0	1.0	4.0	0.0
Endometriosis	0.13	1.0	2.0	0.70	4.0	0.17	4.0	6.0	1.0	4.0	0.0
Hormone	0.25	19.2	15.5	0.47	11.3	0.10	6.5	7.4	5.2	1.2	4.9
Hormone Repl. Therapy	0.12	0.0	3.5	0.66	4.8	0.21	4.0	6.0	1.0	4.0	0.7
Hyperparathyroidism	0.02	1.0	0.0	0.27	1.0	0.04	3.0	6.0	2.0	1.5	3.0
Hypogonadism	0.05	0.0	3.0	0.87	6.0	0.18	4.0	6.0	1.0	4.0	0.0
Lipodystrophy Syndrome	0.31	72.0	67.0	0.63	44.0	0.12	3.0	6.0	3.0	1.0	3.0
Menopause	0.06	0.0	2.0	0.73	4.0	0.17	4.0	6.0	1.0	4.0	0.0
Osteoporosis	0.23	37.0	35.0	0.71	25.0	0.15	4.0	5.7	3.3	1.2	3.0
Ovulation	0.19	18.0	13.0	0.50	10.0	0.09	6.0	6.0	5.0	1.2	5.0

Category	rPSA	N	O	Fsp3	Stereo	Stereo /HA	Rings	Rng Lg	Rng Sys	RRSys	Rng Ar
Phototoxicity in Adults	0.27	21.0	19.0	0.50	12.0	0.10	6.0	6.0	5.0	1.2	5.0
Premature Birth	0.28	11.0	12.0	0.63	9.0	0.13	3.0	20.0	3.0	1.0	1.0
Secondary Hyperthyroidism	0.08	0.0	3.0	0.78	6.0	0.17	3.0	6.0	2.0	1.5	0.0
Treatment of GH Deficiency	0.21	9.0	6.0	0.33	6.0	0.10	5.0	6.0	3.0	1.7	5.0
Immunomodulatory	**0.21**	**3.2**	**10.6**	**0.70**	**9.8**	**0.18**	**2.8**	**11.8**	**2.0**	**1.3**	**0.4**
std dev	*0.11*	*4.7*	*7.6*	*0.17*	*8.3*	*0.12*	*2.0*	*10.5*	*1.6*	*0.6*	*0.6*
max	*0.46*	*19.0*	*36.0*	*1.00*	*35.0*	*0.45*	*7.0*	*33.0*	*7.0*	*2.0*	*2.0*
min	*0.08*	*0.0*	*3.0*	*0.29*	*0.0*	*0.00*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
Antiallergic	0.12	1.0	11.0	0.81	14.0	0.25	4.0	21.0	2.0	2.0	0.0
Antipsoriatic	0.09	0.0	3.0	0.59	4.7	0.16	2.7	6.0	2.0	1.3	0.3
Hereditary Angioedema	0.28	19.0	13.0	0.61	12.0	0.13	7.0	6.0	5.0	1.4	2.0
Immunomodulator	0.28	5.0	6.0	0.64	3.0	0.09	0.5	3.0	0.5	0.5	0.5
Immunostimulant	0.32	2.6	18.0	0.82	15.8	0.28	2.8	6.0	2.8	1.0	0.2
Immunosuppressant	0.19	2.7	9.9	0.69	9.0	0.16	2.8	17.9	1.6	1.6	0.3
Metabolic Disorders	**0.27**	**17.1**	**20.8**	**0.66**	**12.2**	**0.15**	**3.4**	**5.5**	**2.9**	**1.1**	**1.8**
std dev	*0.11*	*25.2*	*27.3*	*0.25*	*15.0*	*0.12*	*3.0*	*3.8*	*2.7*	*0.8*	*2.2*
max	*0.50*	*81.0*	*81.0*	*1.00*	*42.0*	*0.43*	*10.0*	*20.0*	*9.0*	*4.0*	*6.0*
min	*0.09*	*0.0*	*0.0*	*0.25*	*0.0*	*0.00*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
Antidiabetic	0.27	22.1	26.9	0.60	17.3	0.14	4.7	6.5	4.2	1.1	2.9
Antiobesity	0.09	1.0	5.0	0.90	4.0	0.11	1.0	4.0	1.0	1.0	0.0
Cystinosis	0.21	1.0	0.0	1.00	0.0	0.00	0.0	0.0	0.0	0.0	0.0
Fabry's Disease	0.39	1.0	4.0	1.00	4.0	0.36	1.0	6.0	1.0	1.0	0.0
Gaucher's Disease	0.23	1.0	4.0	1.00	4.0	0.27	1.0	6.0	1.0	1.0	0.0
Homocystinuria	0.18	1.0	2.0	0.80	0.0	0.00	0.0	0.0	0.0	0.0	0.0
Hyperammonemia	0.50	2.0	5.0	0.50	1.0	0.08	0.0	0.0	0.0	0.0	0.0
Hyperphenylalaninemia	0.43	5.0	3.0	0.56	3.0	0.18	2.0	6.0	1.0	2.0	1.0
Hypocalciuric	0.43	81.0	81.0	0.25	15.0	0.19	3.0	6.0	3.0	1.0	0.0
Inborn Errors of Bile Synthesis	0.14	0.0	5.0	0.96	11.0	0.38	4.0	6.0	1.0	4.0	0.0
Orotic Aciduria	0.28	2.0	9.0	0.53	4.0	0.15	2.0	6.0	2.0	1.0	1.0
Respiratory	**0.14**	**1.0**	**4.9**	**0.58**	**3.9**	**0.12**	**3.2**	**5.1**	**1.8**	**1.8**	**1.4**
std dev	*0.06*	*1.0*	*1.9*	*0.21*	*3.6*	*0.11*	*1.9*	*2.2*	*1.1*	*1.5*	*1.2*
max	*0.31*	*4.0*	*10.0*	*0.95*	*9.0*	*0.29*	*6.0*	*6.0*	*4.0*	*5.0*	*4.0*
min	*0.08*	*0.0*	*3.0*	*0.19*	*0.0*	*0.00*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
Antiallergic	0.13	0.0	6.0	0.59	9.0	0.24	5.0	6.0	2.0	2.5	1.0
Antiasthmatic	0.15	0.6	5.2	0.56	5.6	0.17	4.8	6.0	1.8	3.5	1.4

Category	rPSA	N	O	Fsp3	Stereo	Stereo /HA	Rings	Rng Lg	Rng Sys	RRSys	Rng Ar
Antihistamine	0.09	2.0	3.0	0.38	1.0	0.04	3.0	6.0	3.0	1.0	2.0
Antitussive	0.08	1.0	3.0	0.46	4.0	0.14	4.0	6.0	3.0	1.3	2.0
Bronchodilator	0.15	1.8	4.0	0.51	2.2	0.09	2.6	6.0	2.0	1.3	1.8
COPD	0.12	1.0	4.5	0.51	3.0	0.11	3.5	6.0	2.5	1.3	2.0
Expectorant	0.31	1.0	3.0	0.83	1.0	0.09	0.0	0.0	0.0	0.0	0.0
Respiratory Distress Syndrome	0.09	0.5	9.0	0.93	2.5	0.05	0.0	0.0	0.0	0.0	0.0
Other	**0.25**	**4.8**	**14.7**	**0.67**	**10.3**	**0.17**	**2.8**	**9.0**	**1.8**	**1.7**	**0.9**
std dev	*0.13*	*10.5*	*16.1*	*0.16*	*11.3*	*0.11*	*2.3*	*10.5*	*1.4*	*2.0*	*1.3*
max	*0.48*	*44.0*	*55.0*	*0.96*	*40.0*	*0.38*	*9.0*	*40.0*	*4.0*	*9.0*	*5.0*
min	*0.08*	*0.0*	*2.0*	*0.45*	*0.0*	*0.00*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
Antiacne	0.15	0.0	3.0	0.61	0.0	0.00	0.5	3.0	0.5	0.5	0.0
Anticirrhotic	0.11	0.0	4.0	0.96	11.0	0.37	4.0	6.0	1.0	4.0	0.0
Antidote	0.30	0.0	47.0	0.89	40.0	0.31	9.0	40.0	1.0	9.0	0.0
Antiglaucoma	0.13	0.3	5.5	0.64	4.7	0.14	1.8	5.8	1.8	1.0	0.8
Dermatological Disorders	0.14	0.0	5.0	0.96	11.0	0.38	4.0	6.0	1.0	4.0	0.0
Dry Eye Syndrome	0.48	4.0	23.0	0.56	12.0	0.24	4.0	6.0	4.0	1.0	2.0
Gastroprotectant	0.42	0.0	39.0	0.46	10.0	0.13	5.0	6.0	4.0	1.3	3.0
Irritable Bowel Syndrome	0.30	16.5	23.5	0.60	15.0	0.14	4.0	28.5	2.0	2.0	1.0
Mucositis	0.42	3.0	4.0	0.63	2.0	0.13	0.0	0.0	0.0	0.0	0.0
Short Bowel Syndrome	0.30	44.0	55.0	0.59	38.0	0.14	5.0	6.0	4.0	1.3	5.0
Urolithiasis	0.32	1.0	3.0	0.60	1.0	0.10	0.0	0.0	0.0	0.0	0.0
Vulnerary	0.38	4.0	13.5	0.71	11.0	0.23	2.5	5.5	2.5	1.0	0.5

Table S9. PCA coordinates and key parameters for selected antimicrobial drugs.

Brand Name	Generic Name	PC1	PC2	PC3	MW	ALOGPs	Fsp3	RngAr
Glycopeptides								
Dalvance	dalbavancin	-4.10	-3.12	-1.25	1,817	3.58	0.42	7
Orbactiv	oritavancin	-4.70	-2.78	-1.69	1,793	1.92	0.42	7
Targocid	teicoplanin	-5.02	-2.66	-1.12	1,880	1.74	0.42	7
Vibativ	telavancin hydrochloride	-4.17	-2.20	-0.54	1,756	2.32	0.53	5
Macrocycles								
Erasis	erythromycin acistrate	-0.40	-0.48	2.64	776	2.85	0.92	0
Ketek	telithromycin	-0.42	-1.58	0.55	812	4.00	0.72	2
Klaricid	clarithromycin	-0.33	-0.61	2.76	748	3.18	0.95	0
Nortron	dirithromycin	-0.73	-0.82	2.94	835	2.90	0.98	0
Ricamycin	rokitamycin	-0.42	-0.77	1.94	828	3.40	0.81	0
Ritro	flurithromycin ethylsuccinate	-0.61	-0.53	2.38	880	2.97	0.91	0
Rulid	roxithromycin	-0.66	-0.51	2.58	837	2.90	0.95	0
Sunamed	azithromycin	-0.45	-0.45	2.88	749	3.03	0.97	0
Zalig	RV11 / propionyl erythromycin mercaptosuccinate	-0.39	-0.62	2.59	790	3.18	0.93	0
Aminoglycosides								
Fortimicin	astromycin sulfate	-0.04	2.62	2.84	405	-1.80	0.94	0
Habekacin	arbekacin sulfate	-1.26	2.99	2.97	553	-2.93	0.95	0
Isepacin	isepanicin	-1.30	2.92	3.11	570	-2.76	0.95	0
Netromicine	netilimicin sulfate	-0.44	2.06	2.47	476	-1.42	0.90	0
Sagamicin	micronomicin sulfate	-0.52	2.19	3.01	464	-1.84	1.00	0
Zemdri	plazomicin	-1.06	2.44	2.25	593	-2.16	0.88	0

Table S10. PCA coordinates and key parameters for selected anticancer drugs.

Brand Name	Generic Name	PC1	PC2	PC3	MW	ALOGPs	Fsp3	RngAr
Anthracyclines								
Aclacin	aclarubicin	-0.67	-0.99	0.34	812	2.79	0.62	2
Calsed	amrubicin hydrochloride	0.41	0.59	-0.43	483	1.40	0.40	2
Farmorubicin	epirubicin hydrochloride	0.11	0.60	-0.21	544	1.41	0.44	2
Pinorubicin	pirarubicin	-0.12	-0.14	-0.23	628	2.06	0.53	2
Valstar	valrubicin	0.12	-0.52	-0.42	724	2.67	0.50	2
Zavedos	idarubicin hydrochloride	0.36	0.39	-0.22	498	1.69	0.42	2
Etoposides								
Etopophos	etoposide phosphate	-0.56	-0.01	0.38	669	0.84	0.55	2
Microtubule Stabilizers								
Ixempra	ixabepilone	0.80	-1.02	0.88	507	3.28	0.74	1
Taxol	paclitaxel	-0.65	-1.66	-0.83	854	3.20	0.45	3
Taxotere	docetaxel	-0.40	-1.12	0.04	808	2.59	0.56	2

Table S11. PCA coordinates and key parameters for selected antiviral drugs.

Brand Name	Generic Name	PC1	PC2	PC3	MW	ALOGPs	Fsp3	Rng Ar
Nucleosides								
Epclusa	sofosbuvir	0.39	0.80	-0.36	529	1.41	0.50	2
Genvoya	emtricitabine	0.97	2.53	-0.21	247	-1.41	0.50	1
Genvoya	tenofovir alafenamide	0.63	0.51	-1.21	476	1.49	0.43	3
Triumeq	abacavir sulfate	0.76	1.17	-0.85	286	0.61	0.50	2
Triumeq	lamivudine	1.04	2.98	-0.05	229	-1.79	0.50	1
Truvada[a]	tenofovir disoproxil fumarate	0.68	1.39	-0.48	519	-0.02	0.63	2
Valtrex	valaciclovir hydrochloride	0.84	2.26	-0.58	324	-1.11	0.54	2
Viread[b]	tenofovir disoproxil fumarate	0.68	1.39	-0.48	519	-0.02	0.63	2
Aminoglycosides								
Inavir	laninamivir octanoate	0.60	1.61	0.87	473	0.86	0.71	0
Relenza	zanamivir	0.30	3.43	1.03	332	-2.29	0.58	0
Tamiflu	oseltamivir	1.47	1.36	0.87	312	1.30	0.75	0
Peptidomimetics								
Epclusa	velpatasvir	-0.75	-2.71	-2.72	883	5.40	0.39	6
Genvoya	cobicistat	-0.07	-1.83	-2.33	776	4.39	0.48	4
Mavyret	glecaprevir	-0.30	-1.76	0.04	839	4.26	0.63	2
Mavyret	pibrentasvir	-1.49	-3.36	-2.92	1,113	5.97	0.47	6
Symtuza	darunavir	0.47	-0.12	-0.62	548	1.89	0.52	2
Ribosomal Peptide (outlier)								
Fuzeon	enfuvirtide	-14.65	-0.67	-1.68	4,492	-1.39[c]	0.56	9
Synthetics								
Genvoya	elvitegravir	1.17	-0.42	-2.01	448	3.43	0.30	3
Triumeq	dolutegravir	0.96	0.94	-1.08	419	0.55	0.35	2

[a] ND category. [b] 2018-N category. [c] Placeholder value, see Supplementary Data Footnotes.

Table S12. PCA coordinates and key parameters for selected antiulcer drugs.

Brand Name	Generic Name	PC1	PC2	PC3	MW	ALOGPs	Fsp3	Rng Ar
Polyisoprenoids and Eicosanoids								
Alloca	ornoprostil	1.42	0.34	1.04	411	2.55	0.78	0
Cytotec	misoprostol	1.65	-0.02	0.93	383	3.88	0.82	0
Gardrin	enprostil	1.20	0.09	-0.50	400	2.58	0.43	1
Kelnac	plaunotol	2.48	0.33	-0.71	306	4.93	0.60	0
Olcenon	tretinoin tocoferil	1.90	-3.43	-0.79	713	9.97	0.65	1
Rosal	rosaprostol sodium	1.92	-0.76	1.07	297	5.58	0.94	0
Selbex	teprenone	2.77	-0.60	-0.81	331	6.99	0.61	0
Solon	sofalcone	1.38	-0.85	-2.26	451	4.94	0.26	2
Synthetics								
Aciphex	rabeprazole sodium	1.18	0.32	-1.77	359	2.04	0.33	3
Nexium	esomeprazole magnesium	1.21	0.44	-1.85	345	1.66	0.29	3
Protonix	pantoprazole sodium	1.11	0.49	-1.99	383	2.11	0.25	3

Table S13. PCA coordinates and key parameters for selected cardiovascular drugs.

Brand Name	Generic Name	PC1	PC2	PC3	MW	ALOGPs	Fsp3	RngAr
Anticoagulants (selected, heparins)								
Arixtra	fondaparinux sodium	-4.96	2.38	2.59	1,498	0.07	0.94	0
Fragmin[a]	dalteparin sodium	-3.72	2.87	2.48	1,135	-1.68	0.88	0
Orgaran[b]	chondroitin sulfate	-5.98	3.53	0.06	909	-0.50	0.22	0
Orgaran	dermatan sulfate	-5.99	3.39	0.06	933	0.05	0.24	0
Orgaran	heparan sulfate	-7.77	3.72	-0.21	1,233	-1.62	0.20	0
Antithrombotics (selected, hirudins)								
Angiomax	bivalrudin	-6.39	0.49	-0.99	2,180	-0.76	0.54	3
Refludan	lepirudin	-24.16	-1.36	1.33	6,979	-1.39[c]	0.62	4
Revasc	desirudin	-24.00	-1.39	1.28	6,963	-1.39[c]	0.62	4
Congestive Heart Failure (selected, natriuretic peptide analogues)								
Hanp	carperitide	-10.19	0.01	0.47	3,080	-1.20	0.57	3
Natrecor	nesiritide	-11.09	0.31	1.31	3,452	-1.43	0.67	2
Hypocholesterolemic (selected, statins)								
Crestor	rosuvastatin	0.76	0.83	-1.24	482	1.47	0.41	2
Lipitor	atorvastatin calcium	0.48	-1.43	-2.76	559	4.24	0.27	4
Mevacor	lovastatin	1.37	-0.76	1.52	405	4.11	0.75	0
Mevalotin	pravastatin	0.98	0.50	1.81	425	2.23	0.74	0
Zocor	simvastatin	1.44	-0.92	1.25	419	4.51	0.76	0

[a] Analyzed as representative tetrameric fragment; also corresponds to other low-molecular-weight heparins: Alphaparin (certoparin sodium), Clivarin (reviparin sodium), Fluxum (parnaparin sodium), Fraxiparine (nadroparin calcium), Hibor (bemiparin sodium), Logiparin (tinzaparin sodium), Lovenox (enoxaparin sodium), Normiflo (ardeparin sodium), and Sandoparin (certoparin sodium). [b] Generic name, danaparoid sodium. [c] Placeholder value, see Supplementary Data Footnotes.

Table S14. PCA coordinates and key parameters for selected immunomodulatory drugs.

Brand Name	Generic Name	PC1	PC2	PC3	MW	ALOGPs	Fsp3	RngAr
Macrolides								
Afinitor	everolimus	-0.34	-2.38	1.78	958	5.01	0.75	0
Elidel	pimecrolimus	0.17	-2.05	1.93	810	4.36	0.81	0
Prograf	tacrolimus	-0.03	-1.58	1.87	804	3.19	0.77	0
Rapamune	sirolimus	-0.23	-2.32	1.81	914	4.85	0.75	0
Cyclic Peptide								
Sandimmun	cyclosporine	-0.68	-1.16	1.45	1,203	4.12	0.79	0
Nucleoside								
Bredinin	mizoribine	0.48	3.56	0.62	259	-2.31	0.56	1
Spermidine								
Spanidin	gusperimus trihydrochloride	0.91	2.82	0.31	388	-0.73	0.82	0
Terpenoids								
CellCept	mycophenolate mofetil	1.29	0.24	-0.88	434	2.17	0.57	1
Myfortic	mycophenolate sodium	1.53	0.59	-1.08	320	2.36	0.41	1

Table S15. PCA coordinates and key parameters for selected metabolic disorder drugs.

Brand Name	Generic Name	PC1	PC2	PC3	MW	ALOGPs	Fsp3	Rng Ar
Antidiabetic Peptides								
Byetta	extenatide	-13.74	-0.30	-0.78	4,187	-1.39[a]	0.61	5
Fulaimei	PEG-loxenatide / PEX168	-15.57	-0.85	-0.81	4,906	-1.39[a]	0.64	5
Lyxumia	lixisenatide	-15.86	-0.16	-0.43	4,859	-1.39[a]	0.65	5
Normylin	triproamylin acetate	-12.91	-0.58	-0.11	3,949	-1.39[a]	0.61	4
Ozempic	semaglutide	-12.54	0.01	-0.98	4,114	-1.39[a]	0.61	6
Victoza	liraglutide	-11.43	-0.02	-0.97	3,751	-1.39[a]	0.59	6
Metabolite Analogues								
Basen	voglibose	0.57	3.73	2.65	267	-2.34	1.00	0
Biopten	sapropterin hydrochloride	0.72	3.18	0.60	241	-2.40	0.56	1
Carbaglu	carglumic acid	1.20	4.65	-0.12	190	-1.06	0.50	0
Cystadane	betaine anhydrous	2.05	3.26	0.19	117	-2.70	0.80	0
Cystaran	cysteamine hydrochloride	2.27	3.26	0.76	77	0.01	1.00	0
Diastabol	miglitol	0.99	3.63	2.70	207	-2.30	1.00	0
Galafold	migalastat hydrochloride	1.02	3.80	3.11	163	-2.22	1.00	0
Janumet	metformin hydrochloride	1.52	4.25	-0.66	129	-1.83	0.50	0
Zavesca	miglustat	1.26	2.86	2.55	219	-1.14	1.00	0
Synthetic Antidiabetics								
Actos	pioglitazone hydrochloride	1.26	-0.44	-1.93	356	3.17	0.32	2
Avandia	rosiglitazone maleate	1.19	-0.04	-1.95	357	2.95	0.28	2
Januvia	sitagliptin	1.23	0.18	-1.34	407	1.95	0.44	2
Jardiance	empagliflozin	0.44	-0.04	-0.49	451	1.79	0.48	2
Jentadueto	linagliptin	0.56	-0.45	-2.18	473	2.62	0.40	4

[a] Placeholder value, see Supplementary Data Footnotes.

Table S16. PCA coordinates and key parameters for selected respiratory drugs.

Brand Name	Generic Name	PC1	PC2	PC3	MW	ALOGPs	Fsp3	Rng Ar
Advair	fluticasone propionate	1.38	-1.39	1.92	501	3.69	0.72	0
Alvesco	ciclesonide	0.99	-1.92	1.49	541	4.08	0.78	0
Pulmicort	budesonide	1.17	-1.17	2.41	431	2.42	0.76	0
Singulair	montelukast sodium	0.93	-3.27	-3.06	585	7.39	0.31	4
Solfa	amlexanox	1.34	0.46	-2.05	298	2.76	0.19	3

C. CALCULATION OF STRUCTURAL AND PHYSICOCHEMICAL PROPERTIES

1. CALCULATION OF PARAMETERS USING INSTANTJCHEM

Physicochemical parameters for approved small-molecule drugs were calculated using the InstantJChem cheminformatic platform by ChemAxon.

The analysis was performed according to the following protocol:

1) SMILES codes for drugs are saved as a comma-separated values file (.csv) file in MS Excel with two columns:

   Column A contains compound names

   Column B containsthe SMILES codes.

   A header row is not included.

2) SMILES codes are imported into InstantJChem as follows:

   a) Click “File” > “New Project” > “Next.”

   b) Specify name and location for the project folder, then click “Finish.”

   c) Click “File” > “Import File” > “Next.”

   d) Click on the folder icon next to the “File to Import” field and locat the file containing the SMILES codes. Under “File Format” choose “Delineated text files (*.csv, *.tab, *.txt).” InstantJChem scans the file and indicates the number of fields found. Once the file has been scanned, click “Next.”

   e) The “Field Details” panel gives a summary of the fields to be imported from the text file. The structure, molecular weight, molecular formula, and compound name of each entry are displayed by default. Undesired term fields are removed using the “< Remove” button. When finished, click “Next.”

   f) The “Monitor Import” window gives a summary of the imported data. Once fully processed, click “Finish.

3) New chemical terms are added by clicking “Data” > “New Chemical Terms Field.” Some chemical terms are manually defined in the expression window. User-defined chemical terms can be saved to the “Expression” menu by clicking on the yellow star to the right of the menu.

   Molecular weight (MW): mass()
   N atom count (N): atomCount(“7”)
   O atom count (O): atomCount(“8”)
   C atom count (C): atomCount("6")
   Heavy atom Count (HA): atomCount()-atomCount(“1”)
   H-bond donor count (HBD): donorCount()
   H-bond acceptor (HBA): acceptorCount()
   Rotatable bond count (Rot): rotatableBondCount()
   Stereocenter count (nStereo): chiralCenterCount()
   Topological polar surface area (tPSA): PSA()
   Number of rings (Rings): ringCount()

Aromatic ring count (RngAr): aromaticRingCount()
Ring system count (RngSys): ringSystemCount()
Size of largest ring (RngLg): largestRingSize()
Number of sp3-hybridized carbons (sp3): count(filter('atno()==6 &&connections()==4')
n-Octanol/water partition coefficient at pH=7.4 (LogD): logd('7.4')
van der Waals surface area (VWSA): vanDerWaalsSurfaceArea()
Relative polar surface area (relPSA): PSA()/vanDerWaals SurfaceArea()

4) Once all desired chemical terms are added to the grid, use "File" > "Export to File" to export the data as an Excel file (.xls).

5) Once the export process completed, click "Finish."

6) In Excel, descriptors consisting of one or more parameters are then calculated:

a) The fraction of sp3 hybridized carbons (Fsp3) is calculated by dividing number of sp3-hybridized carbons (sp3) by the number of carbon atoms (C).

b) Stereochemical density (Stereo/HA, Stereo/MW) is calculated by dividing the number of stereocenters (Stereo) by the number of heavy atoms (HA) or by the molecular weight (MW), respectively.

2. Batch Calculation of ALOGPs and ALOGpS Using VCC Laboratory

Calculation of the ALOGPs and ALOGpS parameters was carried out using the Virtual Computational Chemistry (VCC) Laboratory according to the following protocol:

1) SMILES codes for drugs are saved as a single-column tab delimited text (.txt) file in MS Excel.

Compound names and a header row are not included.

2) The VCC Lab website is accessed at: http://www.vcclab.org/lab/alogps/start.html

3) Click "upload file."

4) Select "smiles -- SMILES file – default" from the dropdown menu and click "Proceed with file uploading."

5) A new browser window opens. Follow "Step 1: Click Browse and select a file."

6) Follow "Step 2: Click Upload File."

7) The window displays "Your file "yourfile.txt" was uploaded as "smiles" successfully. Please, close this page in your browser now."

8) A new browser window opens entitled "results.txt."

9) Save the results page as a text (.txt) file.

10) Open the text file containing the results using MS Excel. Here, "logP" = "ALOGPs" and "logS" = "ALOGpS." The output for SMILES codes with errors is shown as "error in..."

D. PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) on the physicochemical features of approved drugs was carried out using the "R" open source statistical computing software package. Our procedure for PCA using R has been reported previously for both Mac2 and Windows6 operating systems. The current analysis was carried out using R and MS Excel for Mac OS (10.13.6). The protocol is as follows:

1) In MS Excel, a "Raw" worksheet was created with compounds in rows and descriptors in columns.

   Drugs were sorted by compound source (i.e., natural product drugs (NP), natural product-derived drugs (ND), natural product-inspired drugs (S*, S*/NM), and completely synthetic drugs (S, S/NM)).

   Note that compound names must not have spaces or other punctuation.

2) Mean values and standard deviation values were calculated for each column.

3) A "Norm" worksheet was created and mean-centered, standardized values were generated for each column using the equation:

   normval = (val – Column Mean) / Column Standard Deviation

4) With the upper left cell blank (R requires this to recognize a header row), the Number format was designated for all data columns to 4 decimal places.

5) The Excel workbook was saved.

6) The "Norm" worksheet was saved as "Data.txt" (Text–Tab Delimited) on the Desktop (Mac).

7) The Excel workbook was closed and the changes discarded.

8) The "R" open source computing package was opened and the following commands were entered:

9) R> read.table("~/Desktop/Data.txt") -> a # read data into dataframe a

10) R> prcomp(a) -> b # PCA of dataframe a, results to b

11) R> summary(b) # prints summary of %contributions

12) R> b # prints the rotation (loading) matrix

13) R> b$x # prints the rotated data (scores)

14) R > b$x - > c # defines the scores as c

15) R > write.table(c, "~/Desktop/scores.txt", sep = "\t") # saves scores as tab delimited txt file to the Desktop.

16) The scores were imported into a new Excel worksheet "Scores".

17) The first three columns (compound names, PC1, and PC2) where copied into a new worksheet "PCA", and the Number format was designated to 3 decimal places.

18) PCA plots were generated by plotting PC1 and PC2 on XY scatter plots in MS Excel or GraphPad Prism.

19) The loading plot was generated by plotting the loadings for PC1 and PC2 in GraphPad Prism.