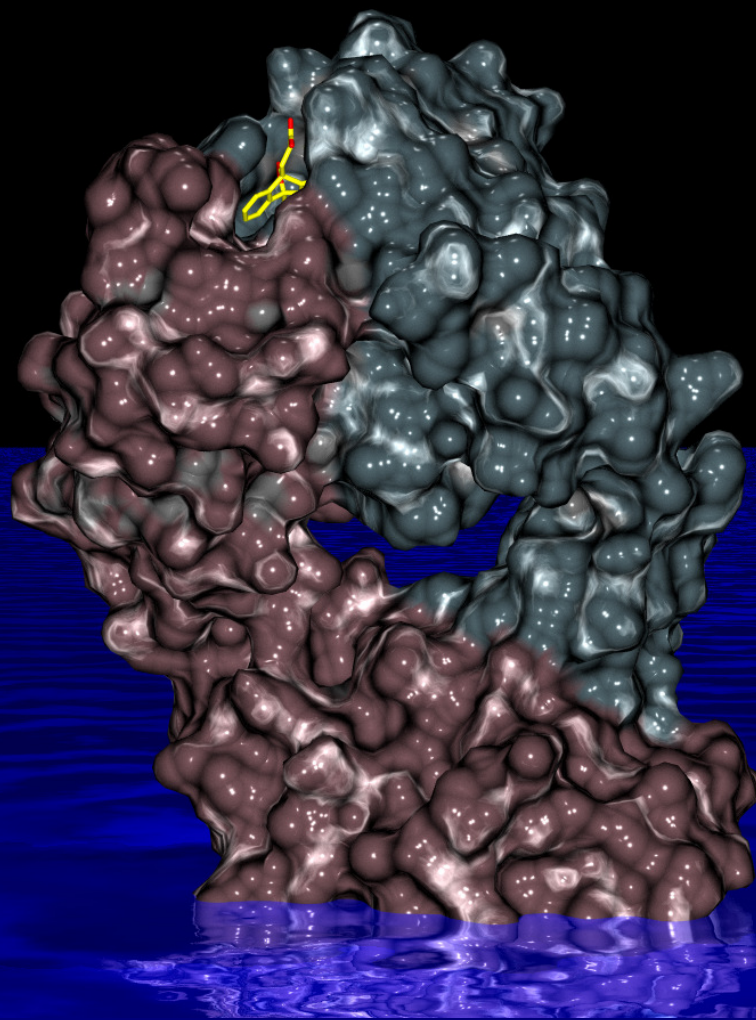


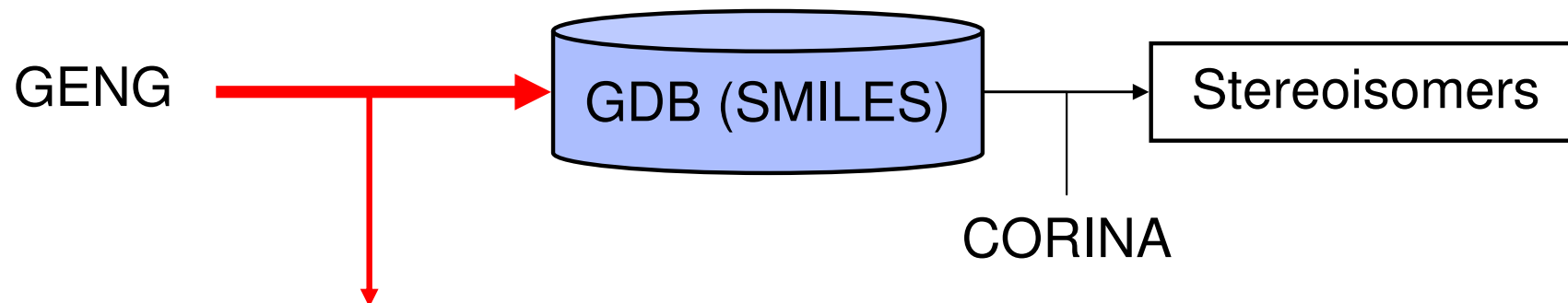
# Enumerating, Mapping and Scoring of Chemical Space for Drug Discovery



1. GDB, CST
2. MQN
3. Scoring

Jean-Louis Reymond  
IPAM/UCLA  
14 April 2011

# GDB Assembly



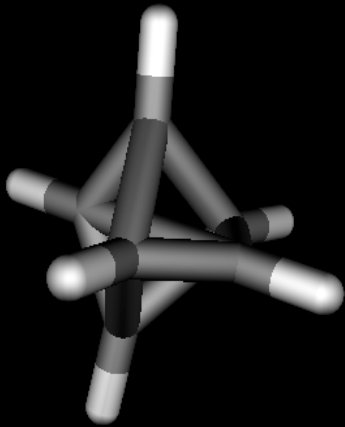
- I. Select hydrocarbon graphs (ring strain, topology)
- II. C-C to C=C, C#C following valency rules, no allenes, no DB in bridgehead or 3- and 4-rings, no TB in <9-rings, no DB torsion
- III. C to N or O following valency rules
- IV. Filter bad functional groups, select tautomers
- V. Post-processing: Halogens, S, etc.

T. Fink et al. *Angew. Chem. Int. Ed.* **2005**, *44*, 1504-1508, *J. Chem. Inf. Model.* **2007**, *47*, 342-353 (GDB-11)

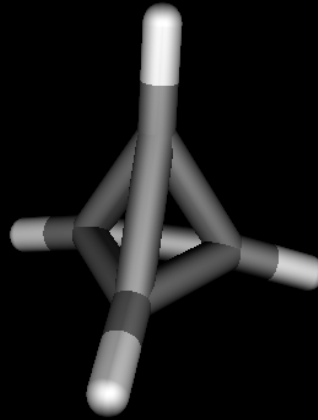
L. C. Blum, J.-L. Reymond, *J. Am. Chem. Soc.* **2009**, *131*, 8732-3 (GDB-13)

Lars Ruddigkeit, Ruud van Deursen, Lorenz Blum (GDB-17)

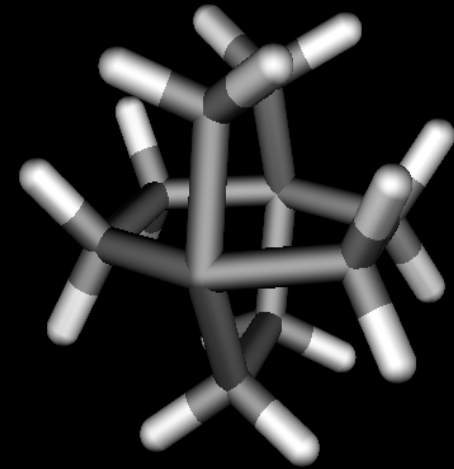
# I. Graph Selection



Claus Benzol  
 $K_{3,3}$  Graph



Tetrahedrane



Tricyclo[2.2.2.2]decane

# IV. Filter Bad Functional Groups

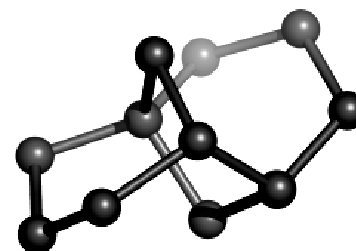
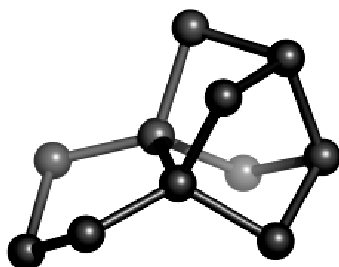
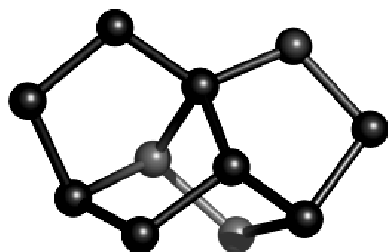
1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56
57	58	59	60	61	62	63	64
65	66	67	68	69	70	71	72

# GDB-11 (CNOF)

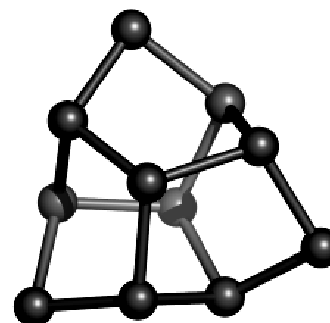
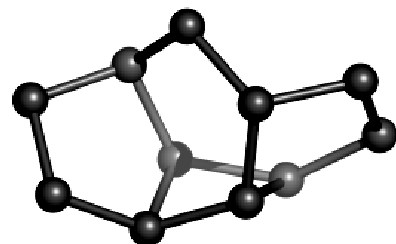
**Table 2.** Overview of the structure generation process.

Nodes	Graphs <sup>a</sup>	Generated <sup>b</sup>	Accepted <sup>c</sup>	Unique Tautomers (GDB) <sup>d</sup>	All Tautomers	Stereoisomers <sup>e</sup>
1	1	4	4	4	4	4
2	1	10	9	9	9	9
3	2	52	20	20	21	20
4	4	332	80	80	88	87
5	8	2'294	357	352	397	469
6	20	18'066	1'906	1'850	2'135	2'911
7	57	154'542	10'953	10'568	12'438	19'904
8	194	1'445'073	69'563	66'706	79'899	153'601
9	705	14'213'741	464'402	444'313	540'002	1'258'963
10	2'822	146'004'340	3'259'036	3'114'041	3'827'907	10'898'065
11	11'912	1'558'491'448	23'875'101	22'796'628	28'240'425	98'645'474
<b>Total</b>	<b>15'726</b>	<b>1'720'329'902</b>	<b>27'681'431</b>	<b>26'434'571</b>	<b>32'703'325</b>	<b>110'979'507</b>

**99.8 % are unknown**

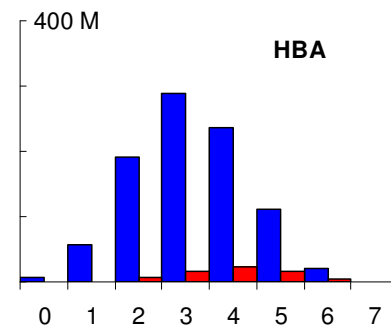
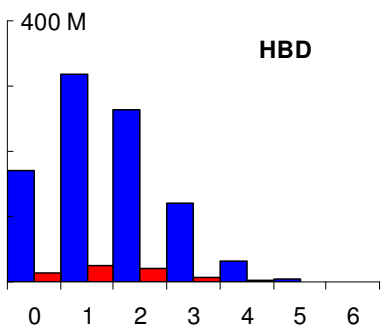
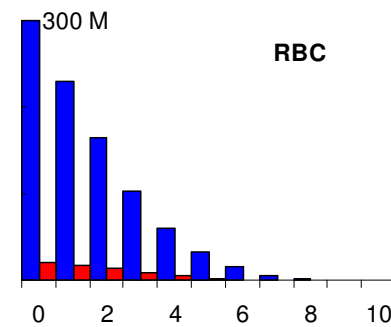
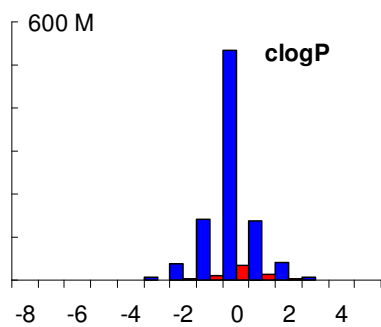
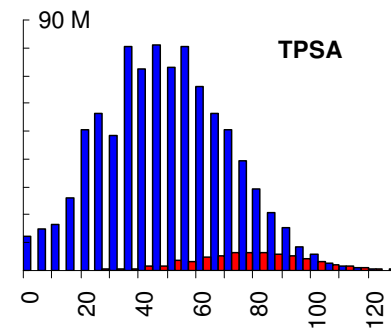
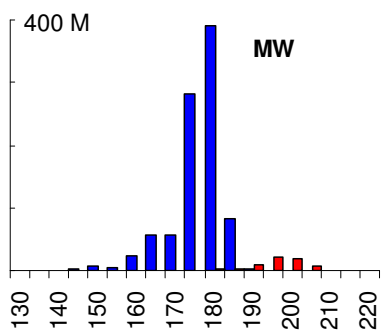
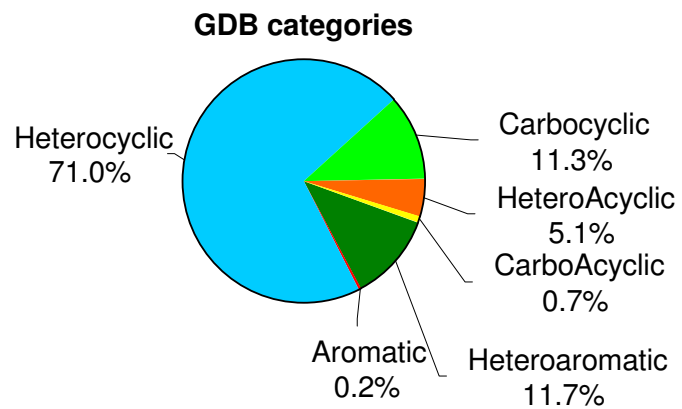
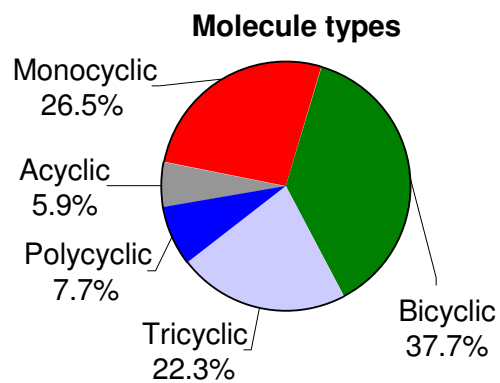
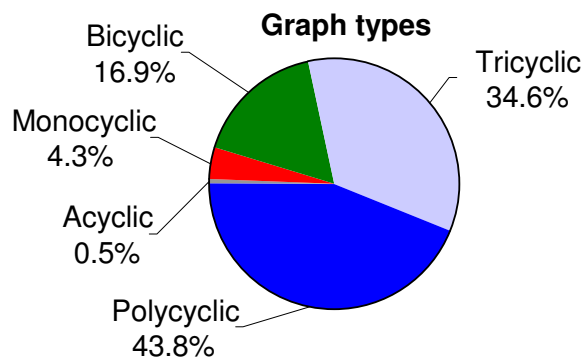


Number of 3-membered rings	Number of 4-membered rings			Total
	0	1	2	
0	124 [3]	189 [60]	103 [67]	416 [130]
1	225 [50]	238 [177]	20 [19]	483 [246]
2	201 [88]	55 [48]	-	256 [136]
3	53 [26]	-	-	53 [26]
<b>Total</b>	<b>603 [167]</b>	<b>482 [285]</b>	<b>123 [86]</b>	<b>1'208 [538]</b>



# GDB-13 (CNOSCl, max. heteroatom ratio)

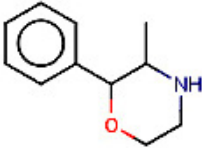
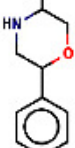
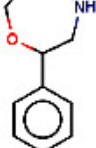
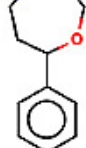
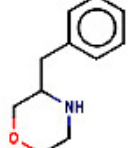
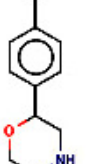
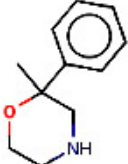
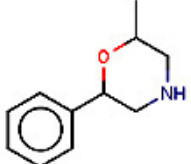
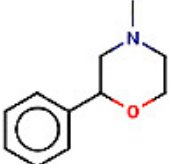
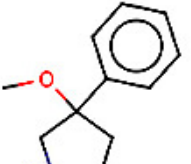
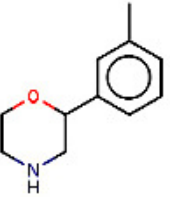
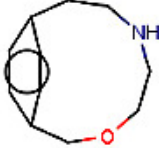
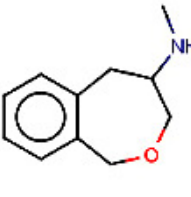
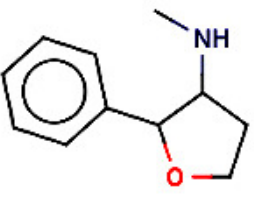
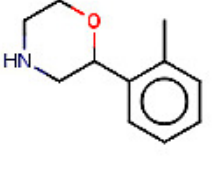
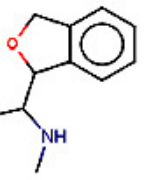
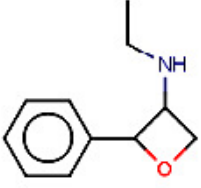
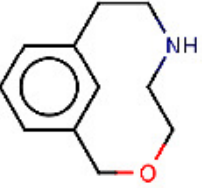
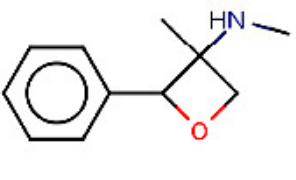
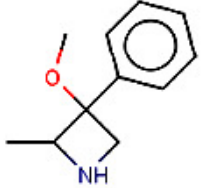
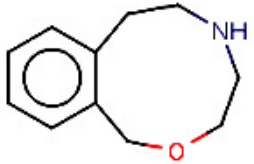
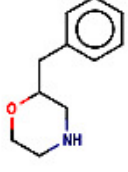
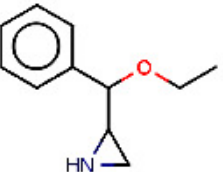
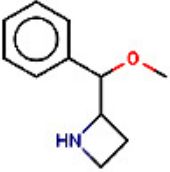
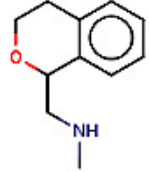
nodes <sup>a</sup>	graphs <sup>b</sup>	GDB <sup>c</sup>	C/S <sup>d</sup>
1	1	1	0
2	1	3	0
3	2	12	0
4	4	43	0
5	8	155	3
6	20	934	19
7	57	5 726	315
8	194	37 151	2 438
9	706	255 542	17 056
10	2 831	1 784 626	130 465
11	12 011	12 961 686	938 704
12	53 789	99 821 343	7 240 108
13	250 268	795 244 451	59 027 533
<b>Total</b>	<b>319 892</b>	<b>910 111 673</b>	<b>67 356 641</b>





# Drugs and isomers in GDB-13

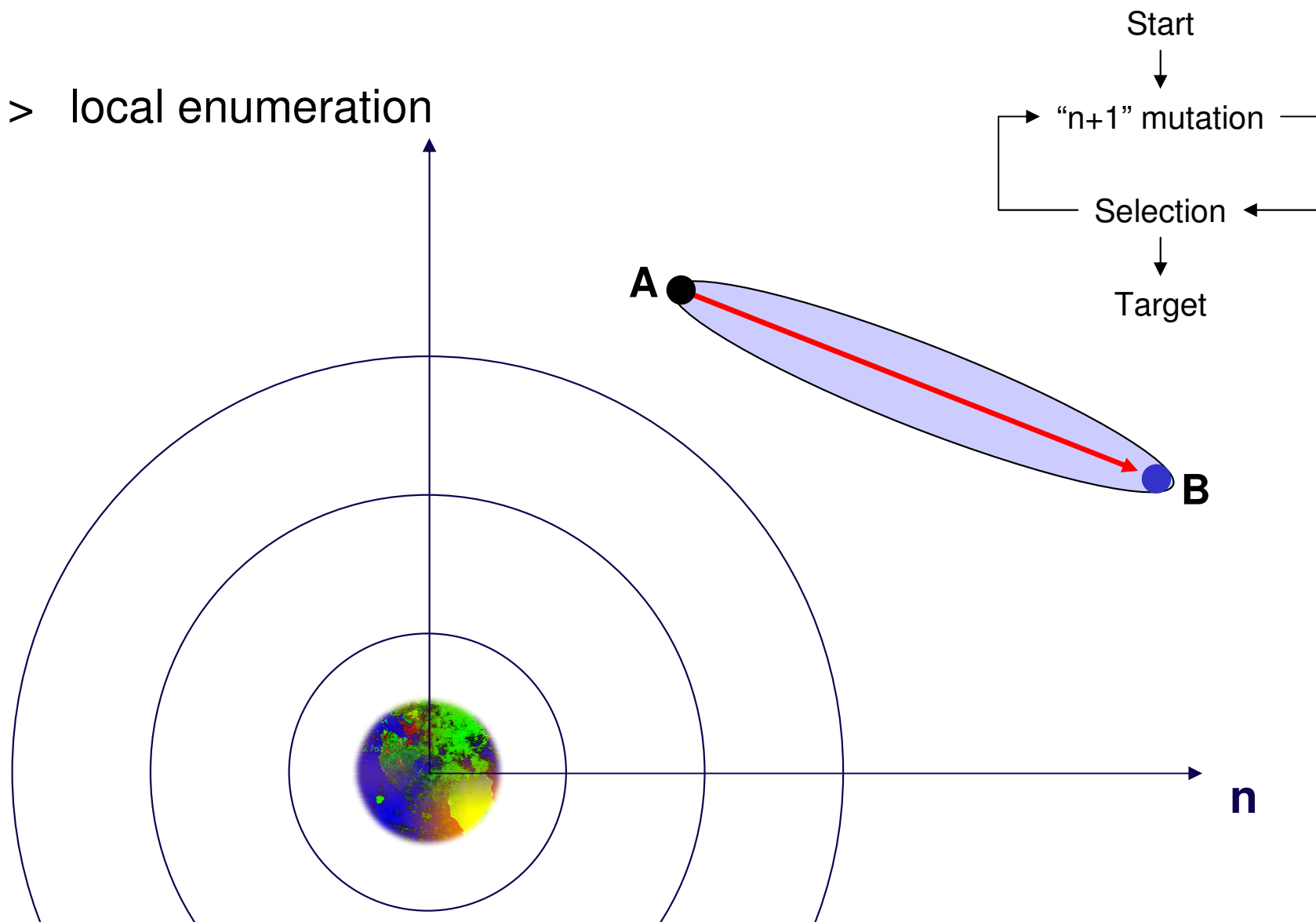
Name	Formula	Same Formula	T <sub>SF</sub>	
			AVG	> 0.7
Aspirin	C <sub>9</sub> H <sub>8</sub> O <sub>4</sub>	804 153	0.23	178
Benzocaine	C <sub>9</sub> H <sub>11</sub> NO <sub>2</sub>	1 846 579	0.24	74
L-Tyrosine	C <sub>9</sub> H <sub>11</sub> NO <sub>3</sub>	9 276 529	0.46	24 952
Levetiracetam	C <sub>8</sub> H <sub>14</sub> N <sub>2</sub> O <sub>2</sub>	2 154 955	0.28	35
Memantine	C <sub>12</sub> H <sub>21</sub> N	2 872 586	0.31	10 912
Menadione	C <sub>11</sub> H <sub>8</sub> O <sub>2</sub>	233 715	0.44	112 186
Metaraminol	C <sub>9</sub> H <sub>13</sub> NO <sub>2</sub>	2 920 516	0.26	30
Mexiletine	C <sub>11</sub> H <sub>17</sub> NO	18 371 393	0.25	119
Propofol	C <sub>12</sub> H <sub>18</sub> O	5 263 227	0.25	240
Rasagiline	C <sub>12</sub> H <sub>13</sub> N	1 323 525	0.13	411
Rimantadine	C <sub>12</sub> H <sub>21</sub> N	2 872 586	0.26	173

1 	2 	3 	4 	5 
0.929204	0.888889	0.884298	0.87069	
6 	7 	8 	9 	10 
0.867257	0.867257	0.867257	0.867257	0.861789
11 	12 	13 	14 	15 
0.853448	0.850877	0.849206	0.84252	0.838983
16 	17 	18 	19 	20 
0.837209	0.835938	0.830508	0.829457	0.828125
21 	22 	23 	24 	25 
0.825	0.809917	0.804511	0.804511	0.801587

Phenmetrazine isomers

# Chemical Space Travel

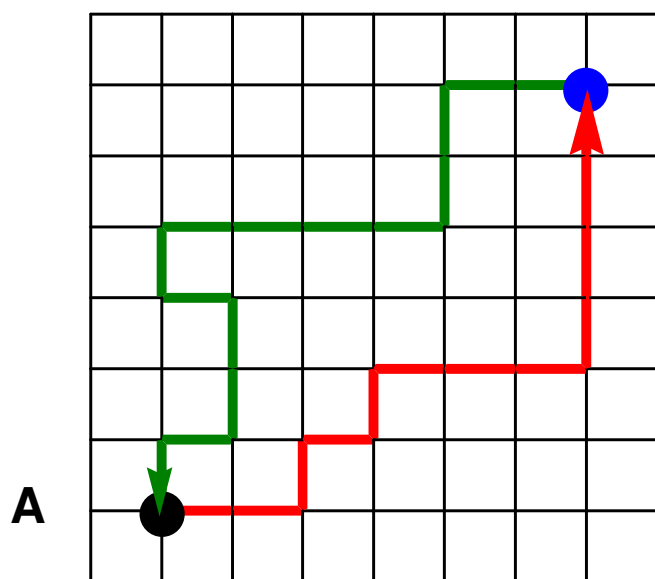
> local enumeration



R. Van Deursen, J.-L. Reymond *ChemMedChem* **2007**, 2, 636-640

**Julian Schwartz**

# Chemical Space as a Graph



**B**

## Nearest neighbour mutations

Atom type exchange

Atom inversion

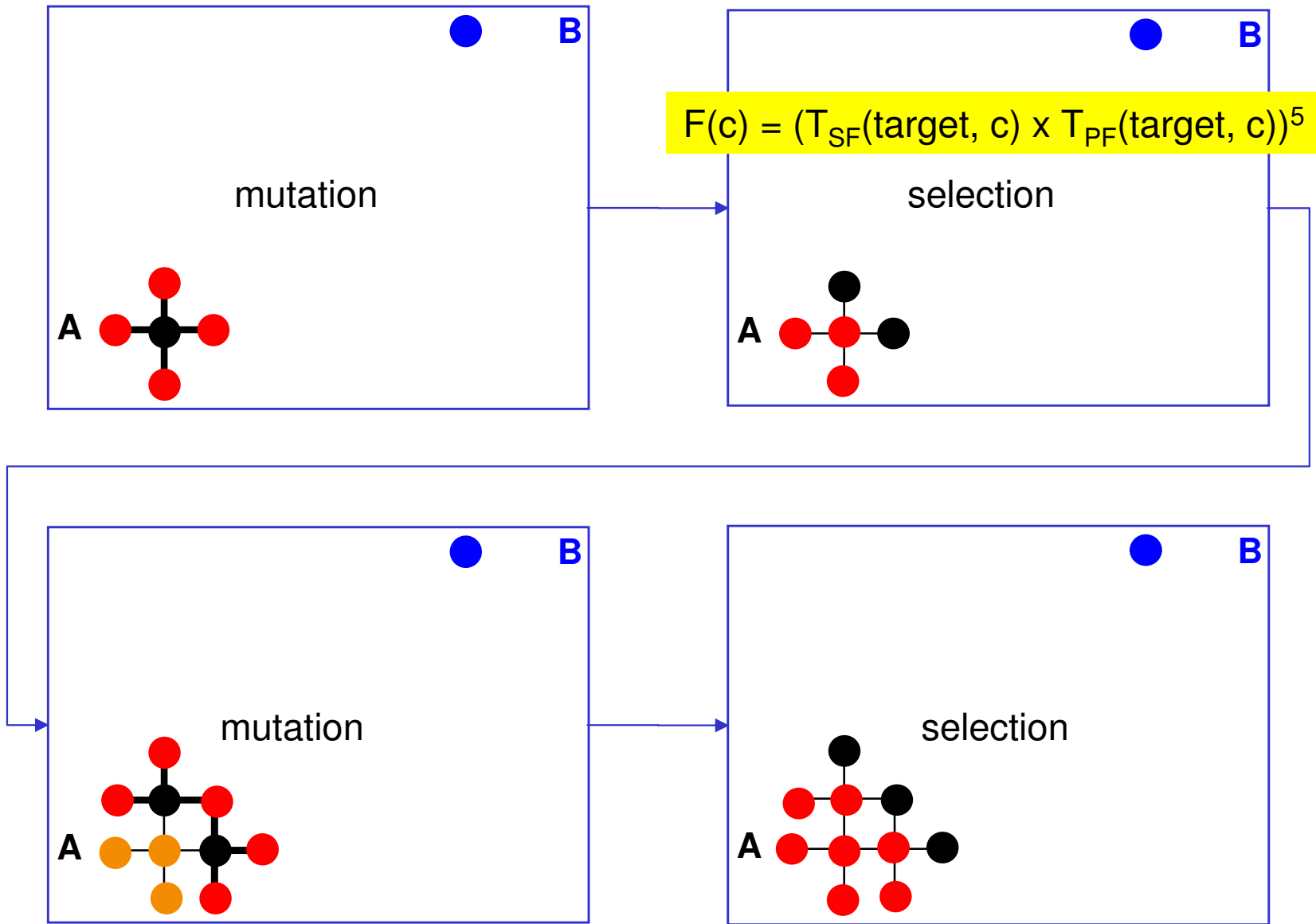
Atom removal

Atom addition

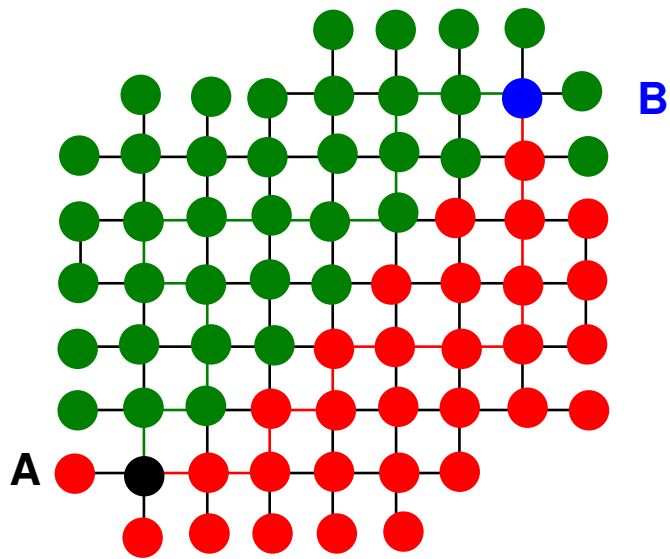
Bond saturation

Bond unsaturation

Bond rearrangement

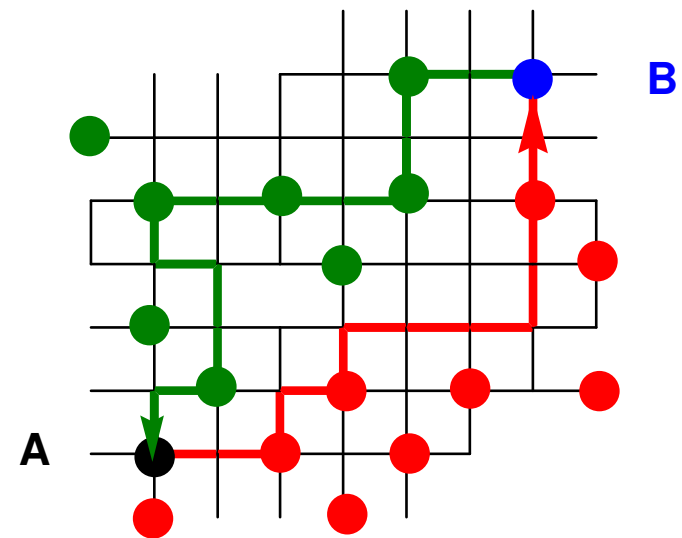


# Trajectory Libraries



"crude"

Filters  
→



possible

# Trajectory Examples (1)

Compound	Formula	mass	$n^{[a]}$	Steps from $\text{CH}_4^{[b]}$		$N^{[d]}$	Steps to $\text{MeOH}^{[b]}$	$N^{[d]}$
				Nearest neighbours	With aromatic <sup>[c]</sup>			
Cubane	$\text{C}_8\text{H}_8$	104	8	12	-	6'638	7	994
Fluorouracil	$\text{C}_4\text{H}_3\text{FN}_2\text{O}_2$	130	9	16	9*	2'456	7*	560
Metheneamine	$\text{C}_6\text{H}_{12}\text{N}_4$	140	10	12	-	6'157	9	1'768
3-Tetrazene-2-carboximidamide	$\text{C}_2\text{H}_6\text{N}_{10}$	170	12	12	-	4'685	11*	2'007
Aspirine	$\text{C}_9\text{H}_8\text{O}_4$	180	13	15	8(1)	2'567	12	2'582
9-Ethyl-carbazole	$\text{C}_{14}\text{H}_{13}\text{N}$	197	15	n.f. <sup>[e]</sup>	20(2)	20'501	16	5'357
Vitamin H	$\text{C}_{10}\text{H}_{16}\text{N}_2\text{O}_3\text{S}$	244	16	18*	-	27'161	14*	6'304
VX (van)	$\text{C}_{11}\text{H}_{26}\text{NO}_2\text{PS}$	267	16	21	-	29'460	14*	3'954
Adenosine	$\text{C}_{10}\text{H}_{13}\text{N}_5\text{O}_4$	267	19	n.f. <sup>[e]</sup>	25(2)	23'680	19*	13'639
$\beta$ -estradiol	$\text{C}_{18}\text{H}_{24}\text{O}_2$	272	20	23	15(2)	43'089	20	19'067
Retinal	$\text{C}_{20}\text{H}_{28}\text{O}$	284	21	23	-	45'176	19*	15'100
Morphine	$\text{C}_{17}\text{H}_{19}\text{NO}_3$	285	21	26	18(2)	69'113	20*	16'247
Aspartame	$\text{C}_{14}\text{H}_{18}\text{N}_2\text{O}_5$	294	21	303	16(2)	34'172	20*	11'430
Cocaine	$\text{C}_{17}\text{H}_{21}\text{NO}_4$	303	22	n.f. <sup>[e]</sup>	20(2)*	70'807	22	17'993

# Trajectory Examples (2)

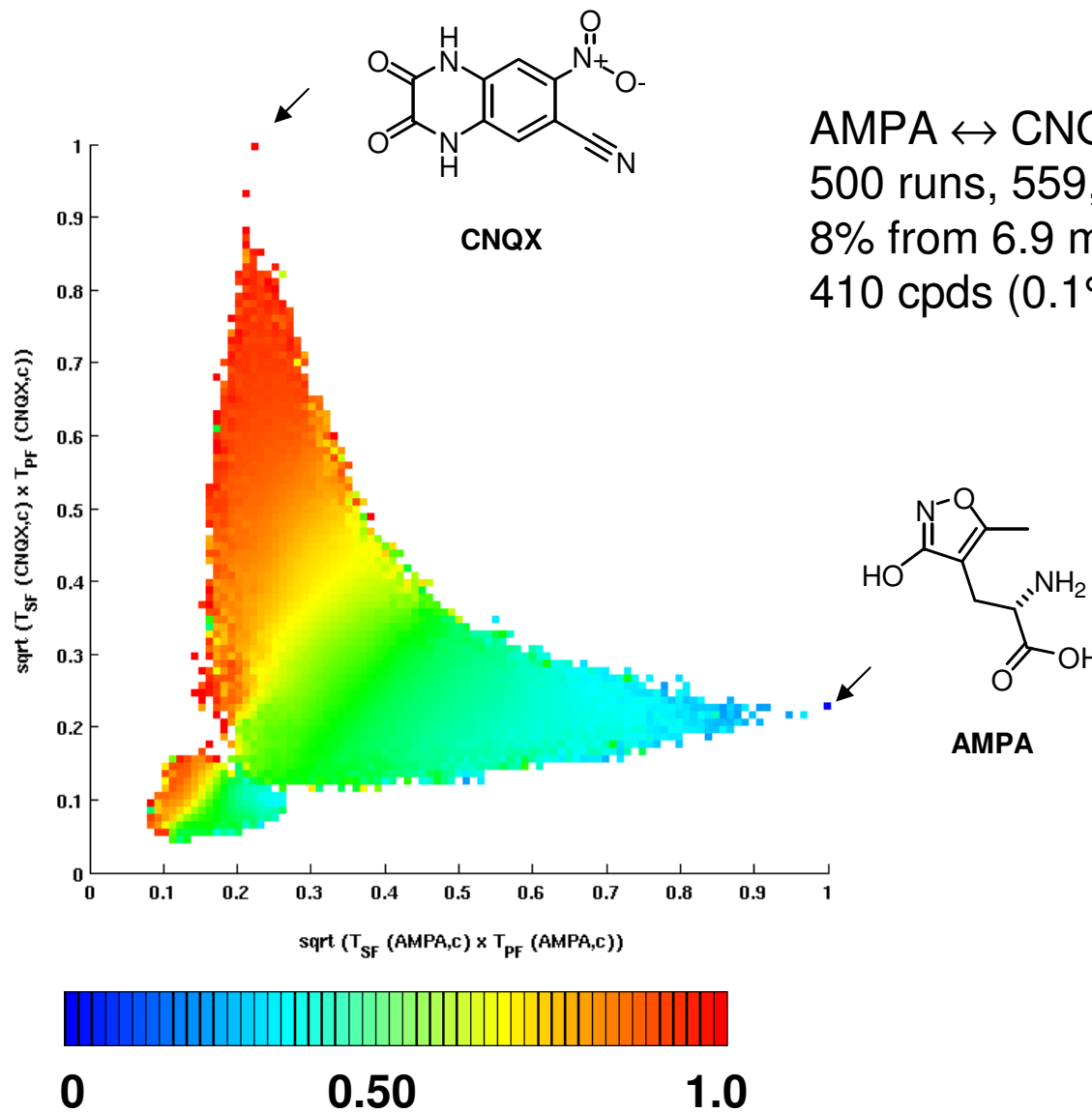
Compound	Formula	mass	$n^{[a]}$	Steps from CH <sub>4</sub> <sup>[b]</sup>		$N^{[d]}$	Steps to MeOH <sup>[b]</sup>	$N^{[d]}$
				Nearest neighbours	With aromatic <sup>[c]</sup>			
Tetrodotoxine	C <sub>11</sub> H <sub>17</sub> N <sub>3</sub> O <sub>8</sub>	319	22	28	-	106'158	20*	16'757
Sucrose	C <sub>12</sub> H <sub>22</sub> O <sub>11</sub>	342	23	25*	-	67'052	21	19'552
Penicillin G	C <sub>16</sub> H <sub>18</sub> N <sub>2</sub> O <sub>4</sub> S	334	23	n.f. <sup>[e]</sup>	20(2)	70'497	23*	15'748
Strychnine	C <sub>21</sub> H <sub>22</sub> N <sub>2</sub> O <sub>2</sub>	332	25	n.f. <sup>[e]</sup>	26(2)	176'721	25	32'479
Papaverin	C <sub>20</sub> H <sub>21</sub> NO <sub>4</sub>	339	25	n.f. <sup>[e]</sup>	25(3)	53'099	25	28'449
Colchicine	C <sub>22</sub> H <sub>25</sub> NO <sub>6</sub>	399	29	37	32(3)	136'519	28	33'592
Calcitriol	C <sub>27</sub> H <sub>44</sub> O <sub>3</sub>	417	30	37*	-	298'327	28*	65'595
Dipicrylamine	C <sub>12</sub> H <sub>5</sub> N <sub>7</sub> O <sub>12</sub>	439	31	n.f. <sup>[e]</sup>	21(2)	21'015	26	13'950
Tetracycline	C <sub>22</sub> H <sub>24</sub> N <sub>2</sub> O <sub>8</sub>	428	31	36	30(1)	173'734	30	34'883
Vitamin K	C <sub>31</sub> H <sub>46</sub> O <sub>2</sub>	451	33	55	42(3)	411'107	32*	77'337
Epothilone	C <sub>27</sub> H <sub>41</sub> NO <sub>6</sub> S	508	35	n.f. <sup>[e]</sup>	62(4)	709'250	34*	75'219
Vitamin E	C <sub>29</sub> H <sub>50</sub> O <sub>2</sub>	531	38	71	40(2)	443'477	37*	140'017
Reserpine	C <sub>33</sub> H <sub>40</sub> N <sub>2</sub> O <sub>9</sub>	609	44	n.f. <sup>[e]</sup>	68(5)	286'342	62	230'646
Taxotere	C <sub>45</sub> H <sub>55</sub> NO <sub>15</sub>	808	58	n.f. <sup>[e]</sup>	74(4)	1'128'960	57*	304'172



# Cross-Trajectories

<b>From:</b>	<b>To:</b>	<b>Cubane</b>	<b>Aspirine</b>	<b>VX</b>	<b>Adenosine</b>	<b>Sucrose</b>	<b>Penicillin G</b>	<b>Strychnine</b>	<b>Colchicine</b>	<b>Tetracycline</b>	<b>Vitamin K</b>
<b>Cubane</b>	-	10	18	23	19	18	18	22	24	26	
<b>Aspirine</b>	10*	-	14	21	15	16	24	22	22	33	
<b>VX</b>	13	17 (1)	-	31 (1)	18	15 (1)	21 (1)	20 (2)	24*	25*	
<b>Adenosine</b>	17*	27	18*	-	14	15	24	23	27*	29	
<b>Sucrose</b>	18*	22 (1)	22*	29 (1)	-	25	26 (1)	31 (1)	25 (1)	25 (1)	
<b>Penicillin G</b>	19*	13*	14*	23	19*	-	20	19*	21*	29	
<b>Strychnine</b>	21*	17*	20	26	22	16*	-	30*	17*	22*	
<b>Colchicine</b>	27	22*	21	26	18	22	23	-	22*	21*	
<b>Tetracycline</b>	28*	20	25*	49	19	19*	16	28	-	17	
<b>Vitamin K</b>	30*	24*	30*	34*	28*	27*	19*	30*	22*	-	

# AMPA $\leftrightarrow$ CNQX

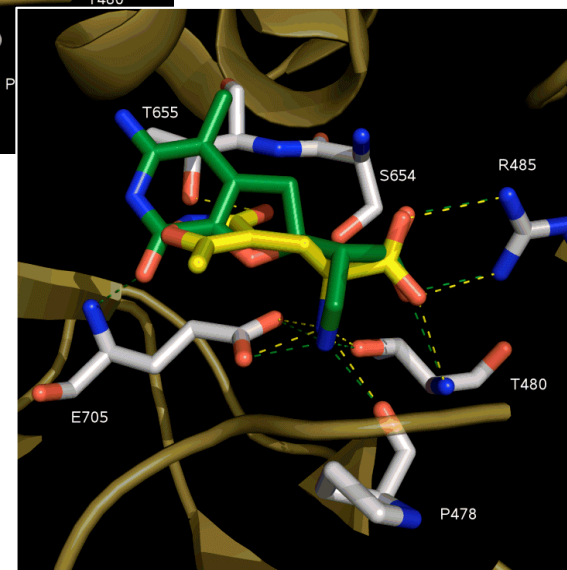
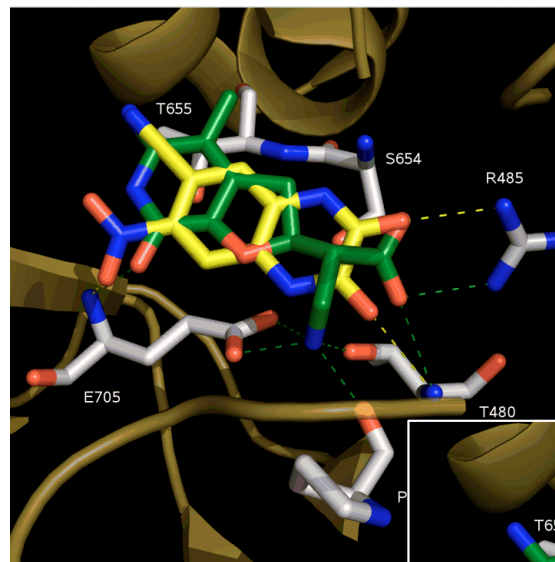
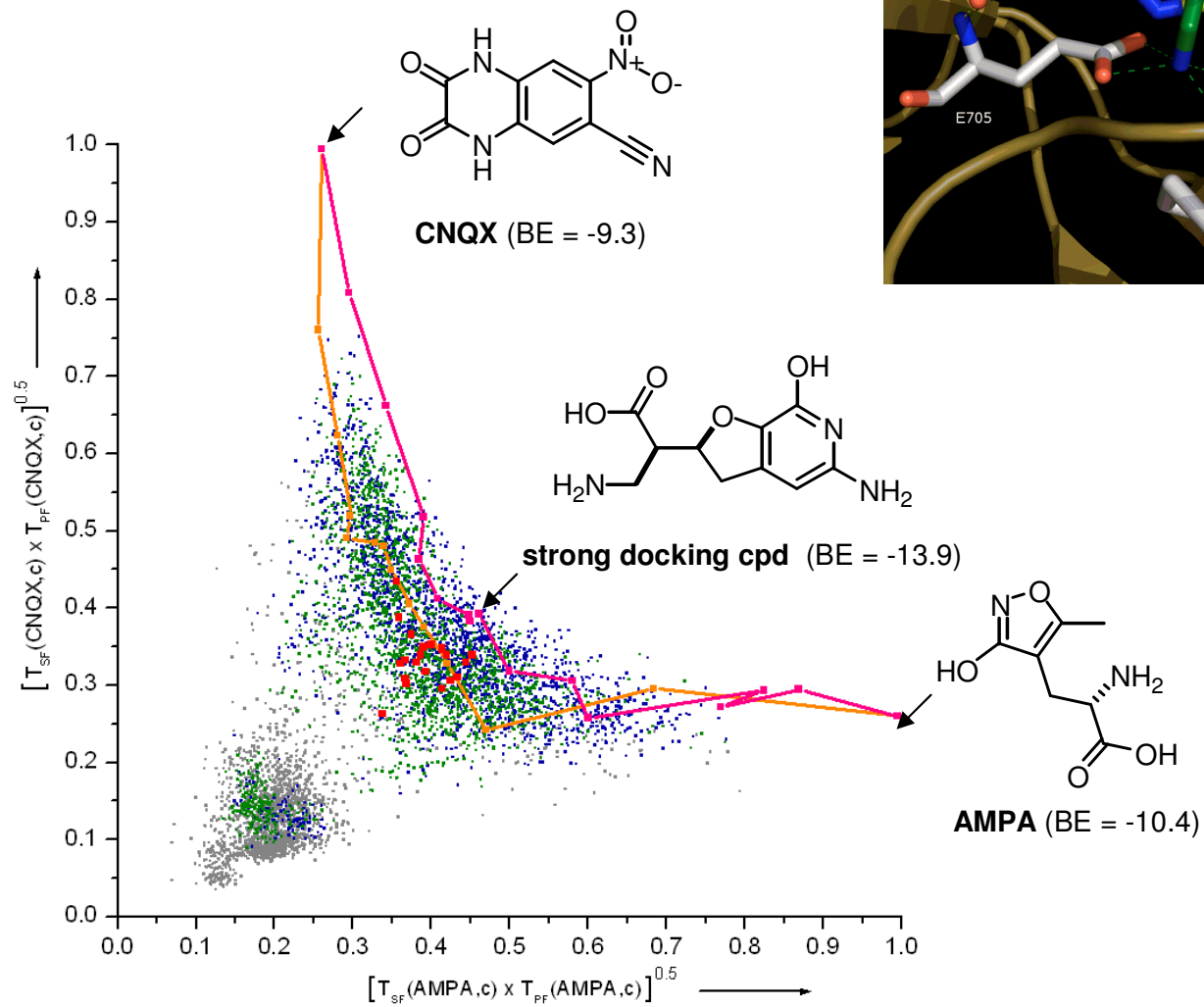


AMPA  $\leftrightarrow$  CNQX

500 runs, 559,658 cpds

8% from 6.9 million generated

410 cpds (0.1%) also in ZINC



1FTK.pdb

# Mapping Chemical Space



1. GDB, CST
2. **MQN**
3. Scoring

K. T. Nguyen, L. C. Blum, R. van Deursen, J.-L. Reymond *ChemMedChem* **2009**, *4*, 1803-5  
R. van Deursen, L. C. Blum, J.-L. Reymond, *J. Chem. Inf. Model.* **2010**, *50*, 1924-1934

# The Periodic System of the Elements

	1																	18						
1	<b>1</b> H Wasserstoff 1,0079 u 1																	<b>2</b> He Helium 4,0026 u 2						
2	<b>3</b> Li Lithium 6,941 u 2/1	<b>4</b> Be Beryllium 9,0122 u 2/2																	<b>5</b> B Bor 10,81 u 2/3	<b>6</b> C Kohlenstoff 12,011 u 2/4	<b>7</b> N Stickstoff 14,007 u 2/5	<b>8</b> O Sauerstoff 15,999 u 2/6	<b>9</b> F Fluor 18,998 u 2/7	<b>10</b> Ne Neon 20,179 u 2/8
3	<b>11</b> Na Natrium 22,99 u 2/8/1	<b>12</b> Mg Magnesium 24,305 u 2/8/2																	<b>13</b> Al Aluminium 26,982 u 2/8/3	<b>14</b> Si Silicium 28,086 u 2/8/4	<b>15</b> P Phosphor 30,974 u 2/8/5	<b>16</b> S Schwefel 32,06 u 2/8/6	<b>17</b> Cl Chlor 35,453 u 2/8/7	<b>18</b> Ar Argon 39,948 u 2/8/8
4	<b>19</b> K Kalium 39,098 u 2/8/8/1	<b>20</b> Ca Calcium 40,08 u 2/8/8/2	<b>21</b> Sc Scandium 44,956 u 2/8/9/2	<b>22</b> Ti Titan 47,90 u 2/8/10/2	<b>23</b> V Vanadium 50,942 u 2/8/11/2	<b>24</b> Cr Chrom 51,996 u 2/8/13/1	<b>25</b> Mn Mangan 54,938 u 2/8/13/2	<b>26</b> Fe Eisen 55,845 u 2/8/14/2	<b>27</b> Co Cobalt 58,933 u 2/8/15/2	<b>28</b> Ni Nickel 58,693 u 2/8/16/2	<b>29</b> Cu Kupfer 63,546 u 2/8/18/1	<b>30</b> Zn Zink 65,38 u 2/8/18/2	<b>31</b> Ga Gallium 69,723 u 2/8/18/3	<b>32</b> Ge Germanium 72,59 u 2/8/18/4	<b>33</b> As Arsen 74,922 u 2/8/18/5	<b>34</b> Se Selen 78,96 u 2/8/18/6	<b>35</b> Br Brom 79,904 u 2/8/18/7	<b>36</b> Kr Krypton 83,80 u 2/8/18/8						
5	<b>37</b> Rb Rubidium 85,468 u 2/8/18/8/1	<b>38</b> Sr Strontium 87,62 u 2/8/18/8/2	<b>39</b> Y Yttrium 88,906 u 2/8/18/10/2	<b>40</b> Zr Zirkonium 91,224 u 2/8/18/12/2	<b>41</b> Nb Niob 92,906 u 2/8/18/13/1	<b>42</b> Mo Molybdän 95,94 u 2/8/18/13/2	<b>43</b> Tc Technetium 98,906 u 2/8/18/15/1	<b>44</b> Ru Ruthenium 101,07 u 2/8/18/16/1	<b>45</b> Rh Rhodium 102,91 u 2/8/18/16/2	<b>46</b> Pd Palladium 106,4 u 2/8/18/18/0	<b>47</b> Ag Silber 107,87 u 2/8/18/18/1	<b>48</b> Cd Cadmium 112,41 u 2/8/18/18/2	<b>49</b> In Indium 114,82 u 2/8/18/18/3	<b>50</b> Sn Zinn 118,69 u 2/8/18/18/4	<b>51</b> Sb Antimon 121,75 u 2/8/18/18/5	<b>52</b> Te Tellur 127,60 u 2/8/18/18/6	<b>53</b> I Iod 126,90 u 2/8/18/18/7	<b>54</b> Xe Xenon 131,30 u 2/8/18/18/8						
6	<b>55</b> Cs Cäsium 132,91 u 2/8/18/18/8/1	<b>56</b> Ba Barium 137,33 u 2/8/18/18/8/2	<b>57-71</b> siehe unten	<b>72</b> Hf Hafnium 178,49 u 2/8/18/32/10/2	<b>73</b> Ta Tantal 180,95 u 2/8/18/32/11/2	<b>74</b> W Wolfram 183,85 u 2/8/18/32/12/2	<b>75</b> Re Rhenium 186,21 u 2/8/18/32/13/2	<b>76</b> Os Osmium 190,2 u 2/8/18/32/14/2	<b>77</b> Ir Iridium 192,22 u 2/8/18/32/15/2	<b>78</b> Pt Platin 195,09 u 2/8/18/32/17/1	<b>79</b> Au Gold 196,97 u 2/8/18/32/18/1	<b>80</b> Hg Quecksilber 200,59 u 2/8/18/32/18/2	<b>81</b> Tl Thallium 204,37 u 2/8/18/32/18/3	<b>82</b> Pb Blei 207,19 u 2/8/18/32/18/4	<b>83</b> Bi Wismut 208,98 u 2/8/18/32/18/5	<b>84</b> Po Polonium 208,98 u 2/8/18/32/18/6	<b>85</b> At Astatin (210 u) 2/8/18/32/18/7	<b>86</b> Rn Radon (222 u) 2/8/18/32/18/8						
7	<b>87</b> Fr Francium (223 u) 2/8/18/32/18/8/1	<b>88</b> Ra Radium 226,03 u 2/8/18/32/18/8/2	<b>89-103</b> siehe unten	<b>104</b> Rf Rutherford. (261 u) 2/8/18/32/32/10/2	<b>105</b> Ha Hahnium (262 u) 2/8/18/32/32/11/2	<b>106</b> Sg Seaborgium (263 u) 2/8/18/32/32/12/2	<b>107</b> Bh Bohrium (262 u) 2/8/18/32/32/13/2	<b>108</b> Hs Hassium (265 u) 2/8/18/32/32/14/2	<b>109</b> Mt Meitnerium (266 u) 2/8/18/32/32/15/2	<b>110</b> Ds Darmstadt. (269 u) 2/8/18/32/32/17/1	<b>111</b> Rg Röntgenium (272 u) 2/8/18/32/32/18/1	<b>112</b> Uub Ununbium (277 u) 2/8/18/32/32/18/2	<b>113</b> Uut Ununtrium (284 u) 2/8/18/32/32/18/3	<b>114</b> Uuq Ununquad. (289 u) 2/8/18/32/32/18/4	<b>115</b> Uup Ununpent. (288 u) 2/8/18/32/32/18/5	<b>116</b> Uuh Ununhex. (289 u) 2/8/18/32/32/18/6								
	<div style="border: 1px solid black; padding: 5px; display: inline-block;"> <p><b>Atomic Number</b></p> <p>Ordnungszahl    Symbole    Ordnungszahl    Serie</p> <p>schwarz = nicht radioaktiv    <input type="checkbox"/> Alkalimetalle    <input type="checkbox"/> Metalle</p> <p>gelb = radioaktiv    <input type="checkbox"/> Erdalkalimetalle    <input type="checkbox"/> Halbmetalle</p> <p>Name    Platin    Serie</p> <p>Atomgewicht    195,084    <input type="checkbox"/> Übergangsmetalle    <input type="checkbox"/> Nichtmetalle</p> <p>Elektronenhüllen    2/8/18/32/17/1    <input type="checkbox"/> Lanthanoide    <input type="checkbox"/> Halogene</p> <p><input type="checkbox"/> Actinoide    <input type="checkbox"/> Edelgase</p> <p>Symbol</p> <p>schwarz = fest    durchgehend = natürliche Elemente</p> <p>blau = flüssig    schraffiert = künstliche Elemente</p> <p>rot = gasförmig</p> </div>																							
	<div style="border: 1px solid black; padding: 5px; display: inline-block;"> <p><b>Principal Quantum Number</b></p> </div>																							
	<b>57</b> La Lanthan 138,91 u 2/8/18/18/9/2	<b>58</b> Ce Cer 140,12 u 2/8/18/19/9/2	<b>59</b> Pr Praseodym 140,91 u 2/8/18/21/8/2	<b>60</b> Nd Neodym 144,24 u 2/8/18/22/8/2	<b>61</b> Pm Promethium 146,9 u 2/8/18/23/8/2	<b>62</b> Sm Samarium 150,35 u 2/8/18/24/8/2	<b>63</b> Eu Europium 151,96 u 2/8/18/25/8/2	<b>64</b> Gd Gadolinium 157,25 u 2/8/18/25/9/2	<b>65</b> Tb Terbium 158,93 u 2/8/18/27/8/2	<b>66</b> Dy Dysprosium 162,50 u 2/8/18/28/8/2	<b>67</b> Ho Holmium 164,93 u 2/8/18/29/8/2	<b>68</b> Er Erbium 167,26 u 30/8/2	<b>69</b> Tm Thulium 168,93 u 31/8/2	<b>70</b> Yb Ytterbium 173,04 u 2/8/18/32/8/2	<b>71</b> Lu Lutetium 174,97 u 32/9/2									
	<b>89</b> Ac Actinium (227 u) 2/8/18/32/18/9/2	<b>90</b> Th Thorium 232,04 u 2/8/18/32/18/10/2	<b>91</b> Pa Protaktin. 231,04 u 2/8/18/32/20/9/2	<b>92</b> U Uran 238,03 u 2/8/18/32/21/9/2	<b>93</b> Np Neptunium 237,05 u 2/8/18/32/22/9/2	<b>94</b> Pu Plutonium (244,1 u) 2/8/18/32/24/8/2	<b>95</b> Am Americium (243,1 u) 2/8/18/32/25/8/2	<b>96</b> Cm Curium (247,1 u) 2/8/18/32/25/9/2	<b>97</b> Bk Berkelium (247,1 u) 2/8/18/32/27/8/2	<b>98</b> Cf Californium (251,1 u) 2/8/18/32/28/8/2	<b>99</b> Es Einsteinium (254,1 u) 2/8/18/32/29/8/2	<b>100</b> Fm Fermium (257,1 u) 30/8/2	<b>101</b> Md Mendelev. (258 u) 31/8/2	<b>102</b> No Nobelium (259 u) 32/8/2	<b>103</b> Lr Lawrencium (260 u) 32/9/2									

# Molecular Quantum Numbers

## Atom counts

Carbon	17	14
Fluorine	0	0
Chlorine	0	0
Bromine	0	0
Iodine	0	0
Sulphur	0	1
Phosphorus	0	0
Acyclic nitrogen	0	1
Cyclic nitrogen	1	1
Acyclic oxygen	2	4
Cyclic oxygen	1	1
Heavy atom count	21	21

## Polarity counts

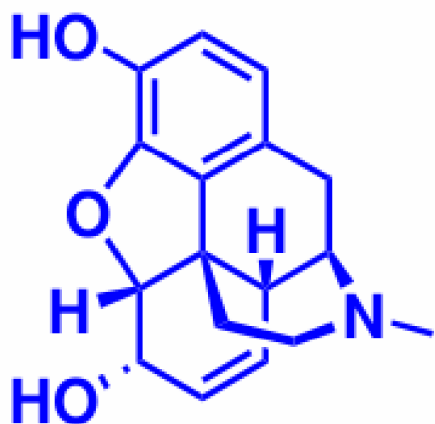
H-Bond donor atoms	3	1
H-Bond donor sites	3	1
H-Bond acc. atoms	3	4
H-Bond acc. sites	3	7
Positive charges	1	0
Negative charges	0	1

## Bond counts

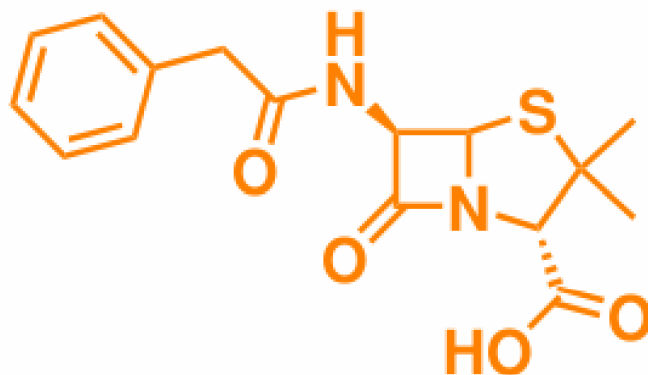
Acyclic single bonds	3	6
Acyclic double bonds	0	3
Acyclic triple bonds	0	0
Cyclic single bonds	18	11
Cyclic double bonds	4	3
Cyclic triple bonds	0	0
Rotatable bonds	0	4

## Topology counts

Acyclic monovalent nodes	3	4
Acyclic divalent nodes	0	2
Acyclic trivalent nodes	0	2
Acyclic tetravalent nodes	0	0
Cyclic divalent nodes	8	7
Cyclic trivalent nodes	9	6
Cyclic tetravalent nodes	1	0
3-Membered rings	0	0
4-Membered rings	0	1
5-Membered rings	1	1
6-Membered rings	4	1
7-Membered rings	0	0
8-Membered rings	0	0
9-Membered rings	0	0
≥10-Membered rings	0	0
Atoms shared by fused rings	7	2
Bonds shared by fused rings	6	1

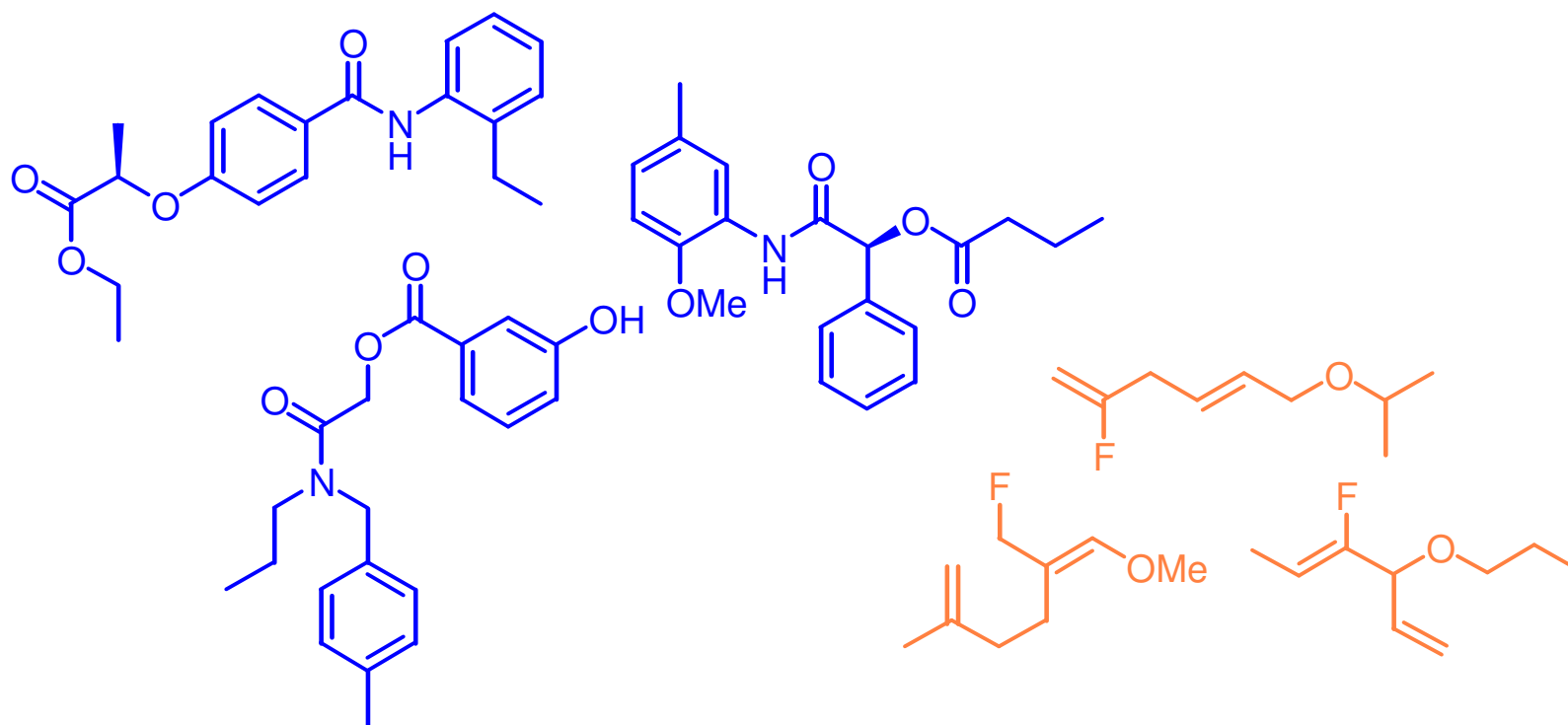


*Morphine*

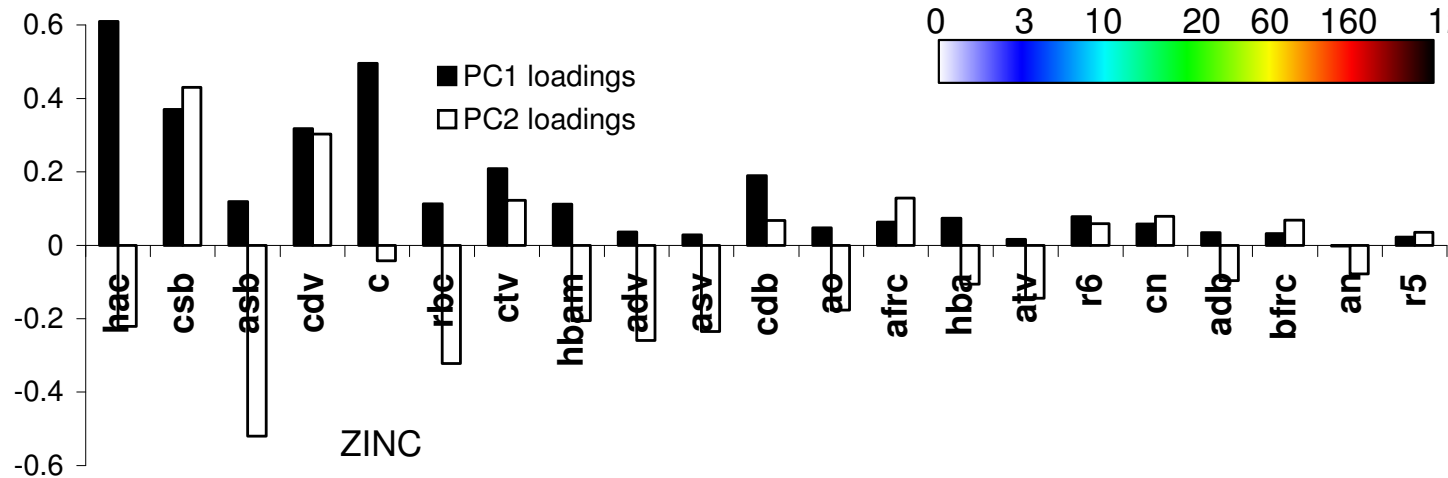
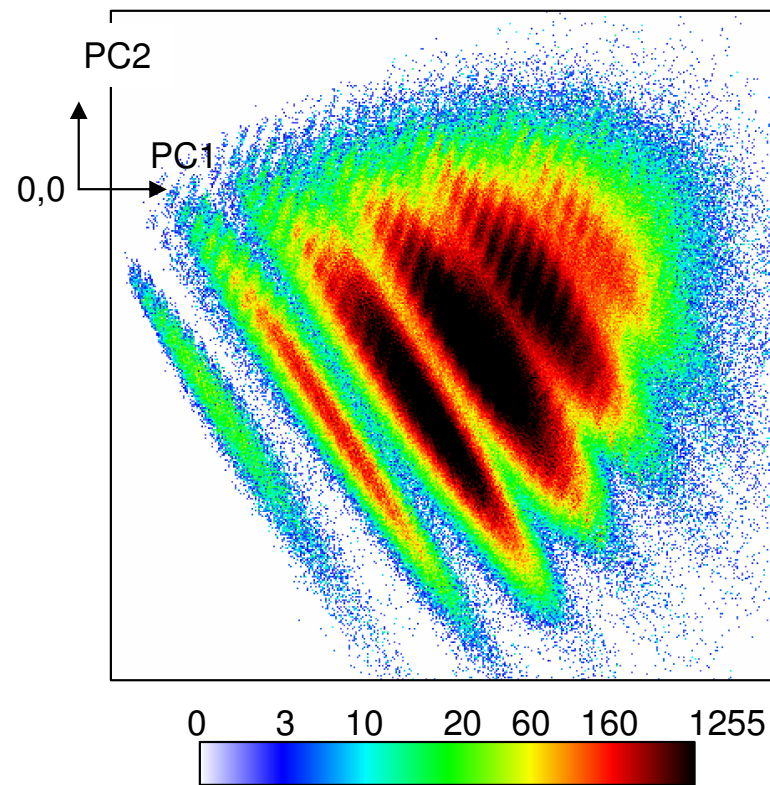
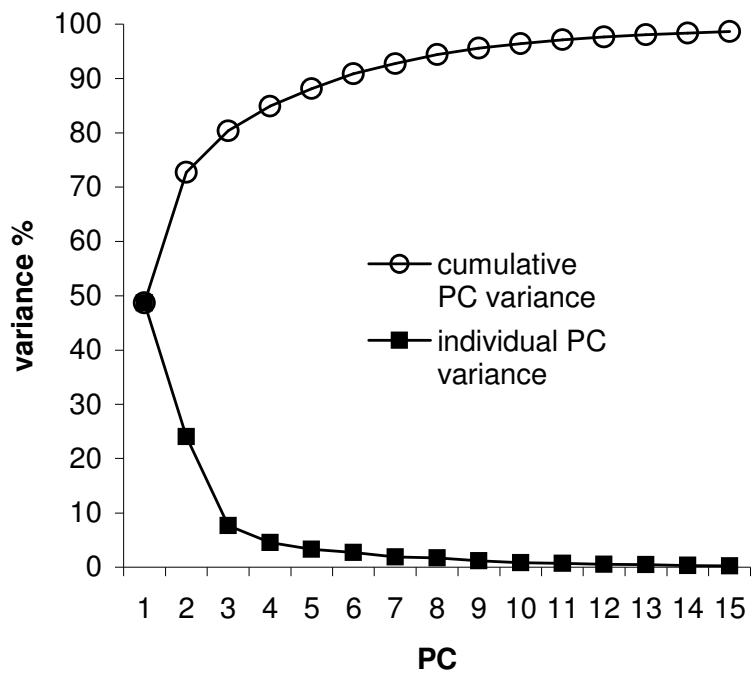


*Penicillin G*

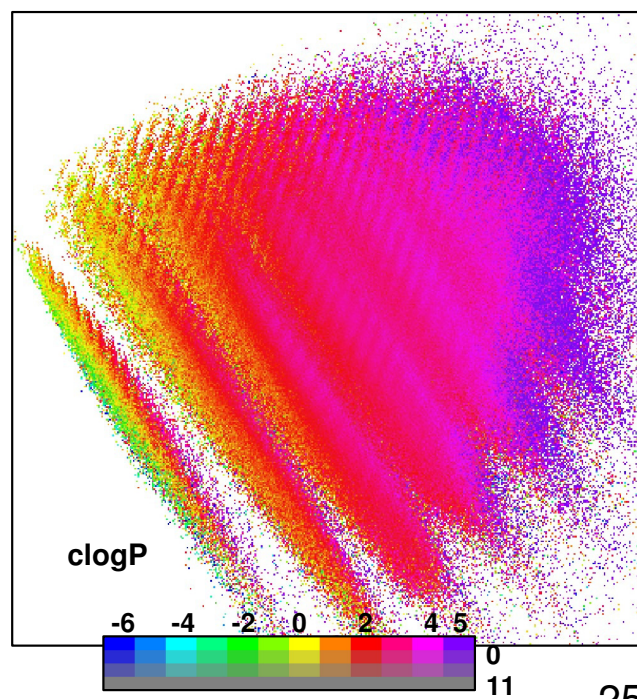
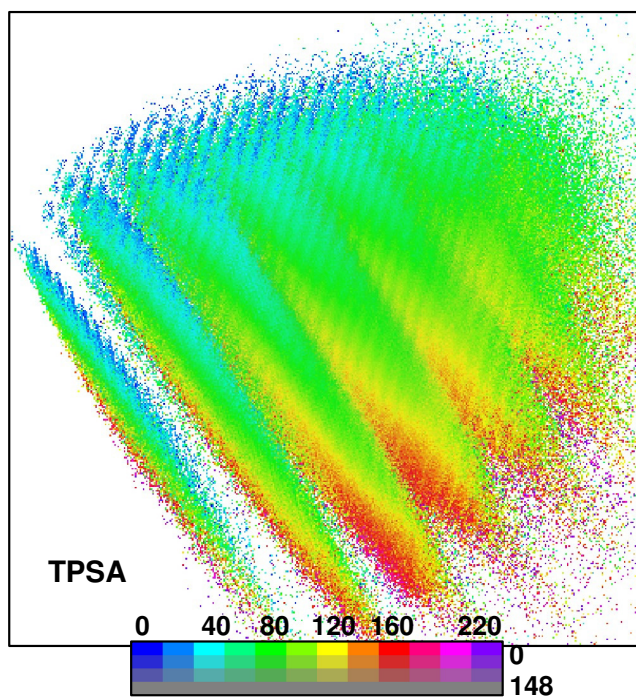
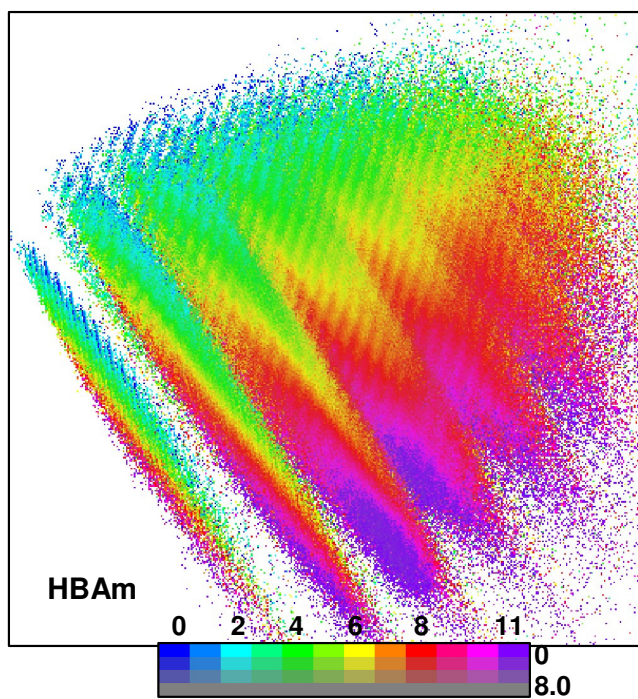
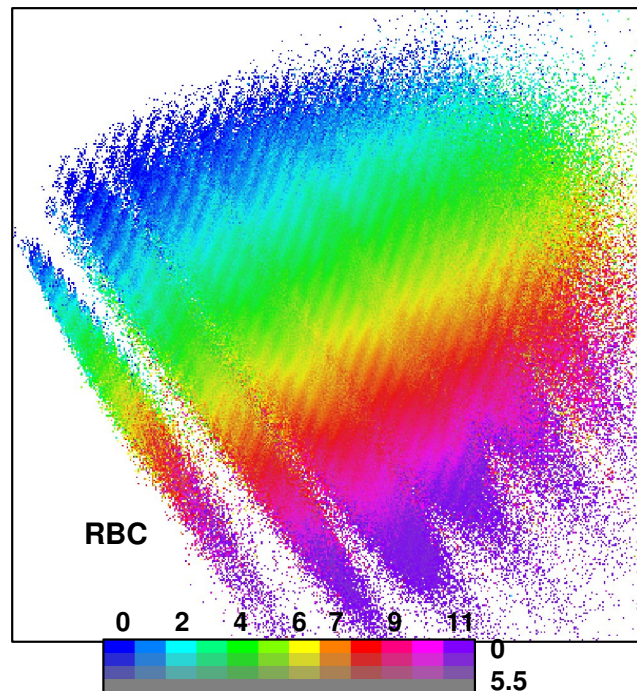
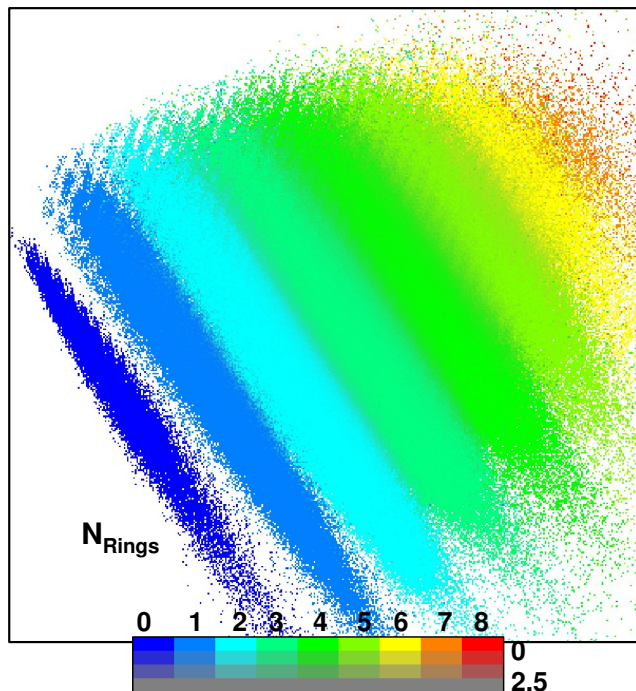
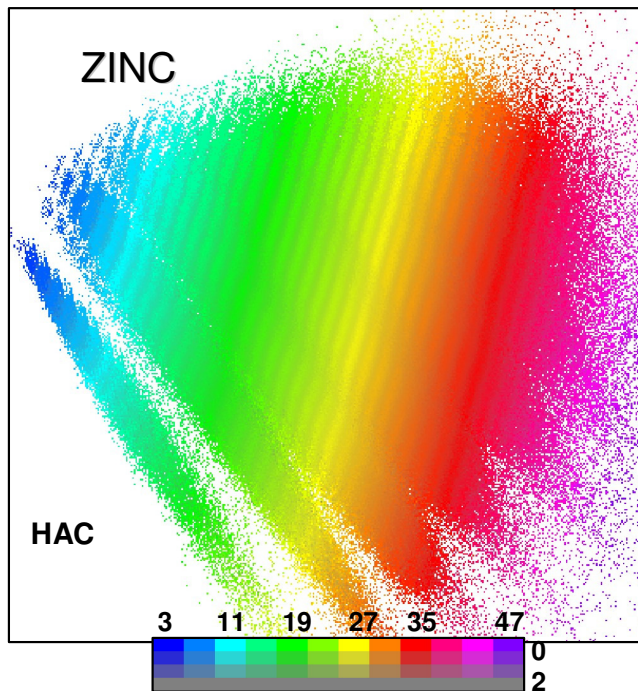
	ZINC	GDB-11
no. of cpds	8 436 272	26 434 567
no. of MQN-bins	3 654 836	2 859 938
no. of single occupied MQN-bins	1 832 566	660 851
no. of cpds in most occupied MQN-bin	300	1 982
no. of shared MQN-bins	13 769	13 769
no. of cpds in shared MQN-bins	30 779	254 604



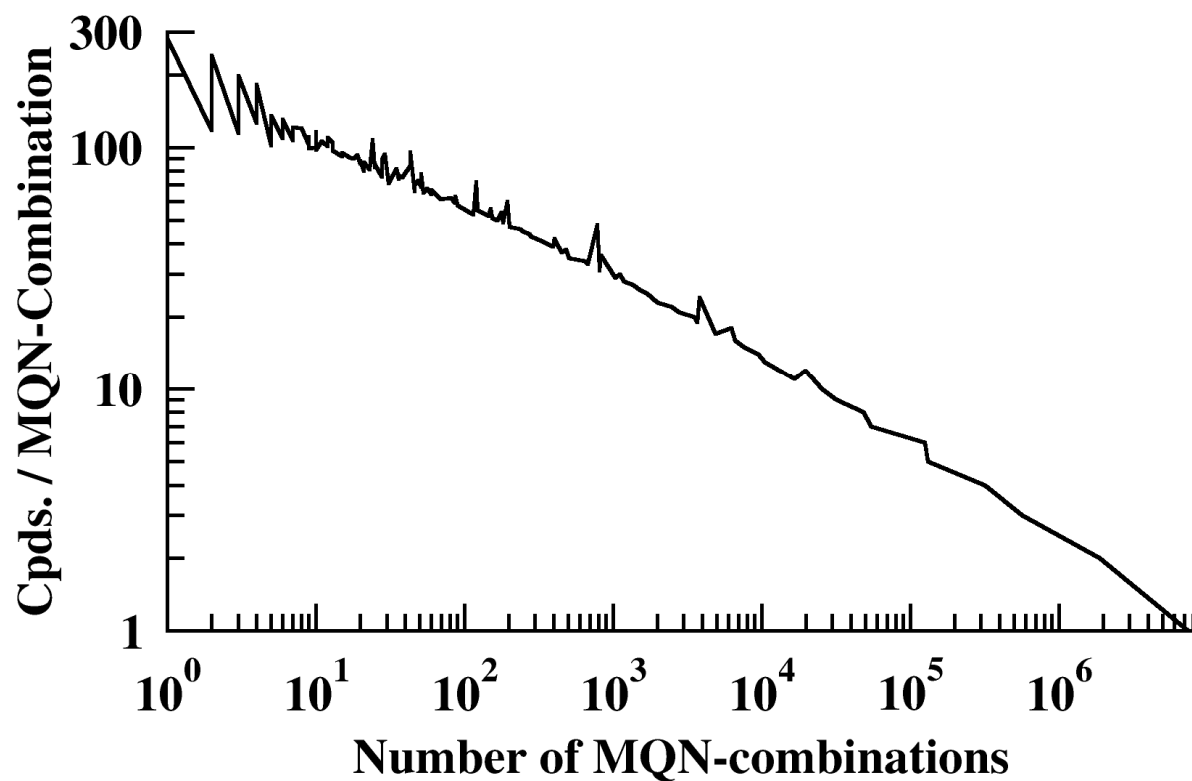
# ZINC (8.4 million SMILES)





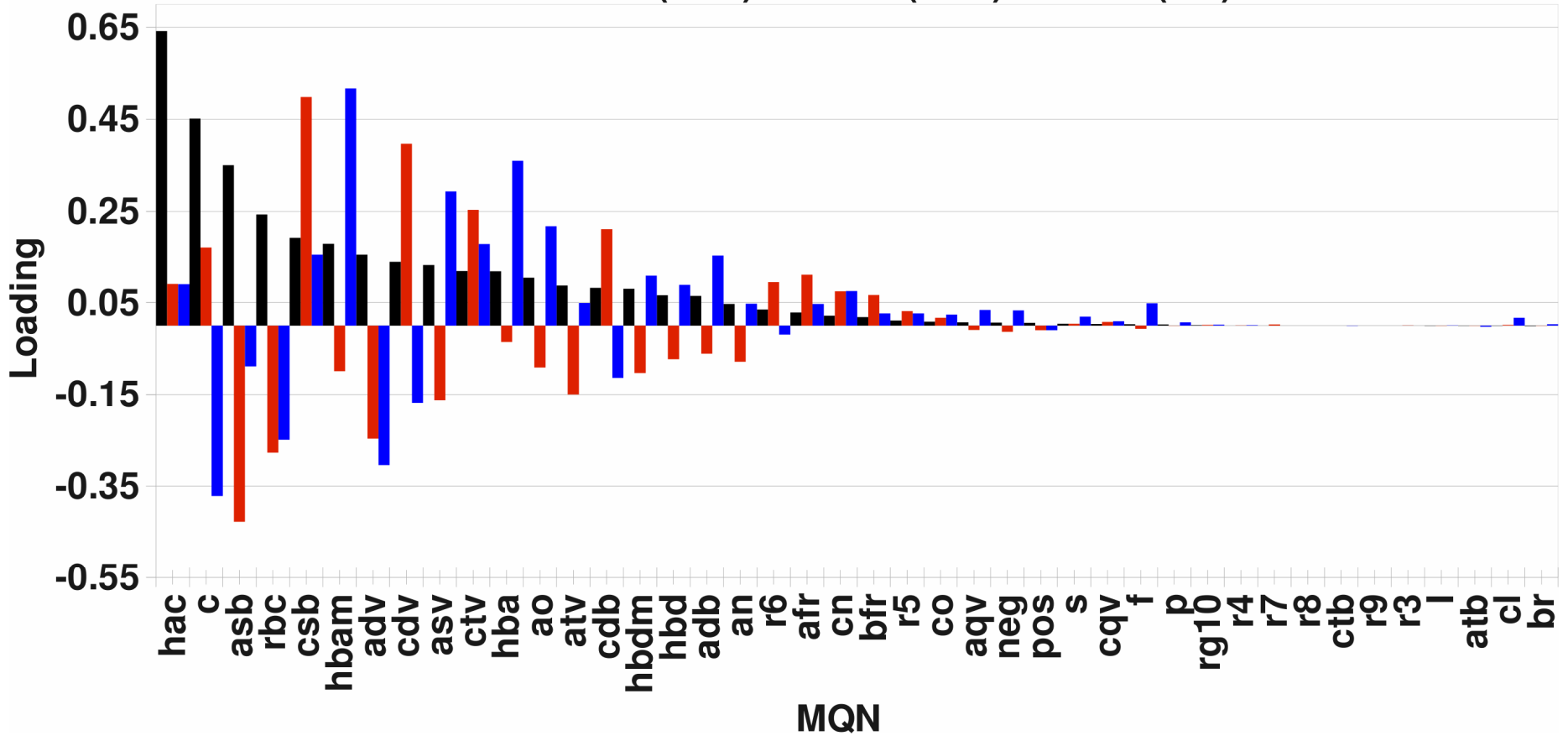


# PubChem (19.2 million SMILES)



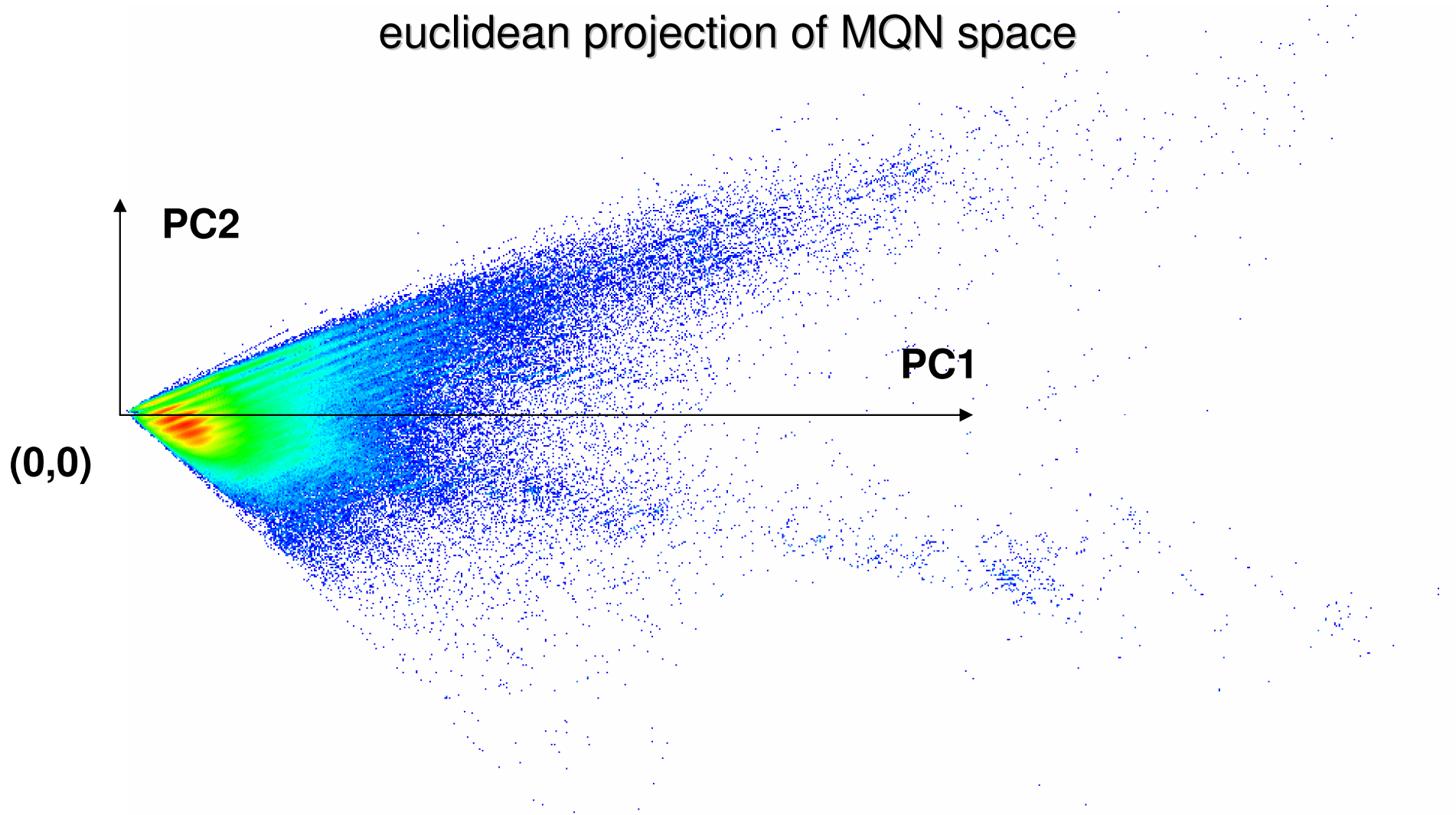
# PC-Loadings for MQNs

■ PC1 (65%) ■ PC2 (18%) ■ PC3 (7%)



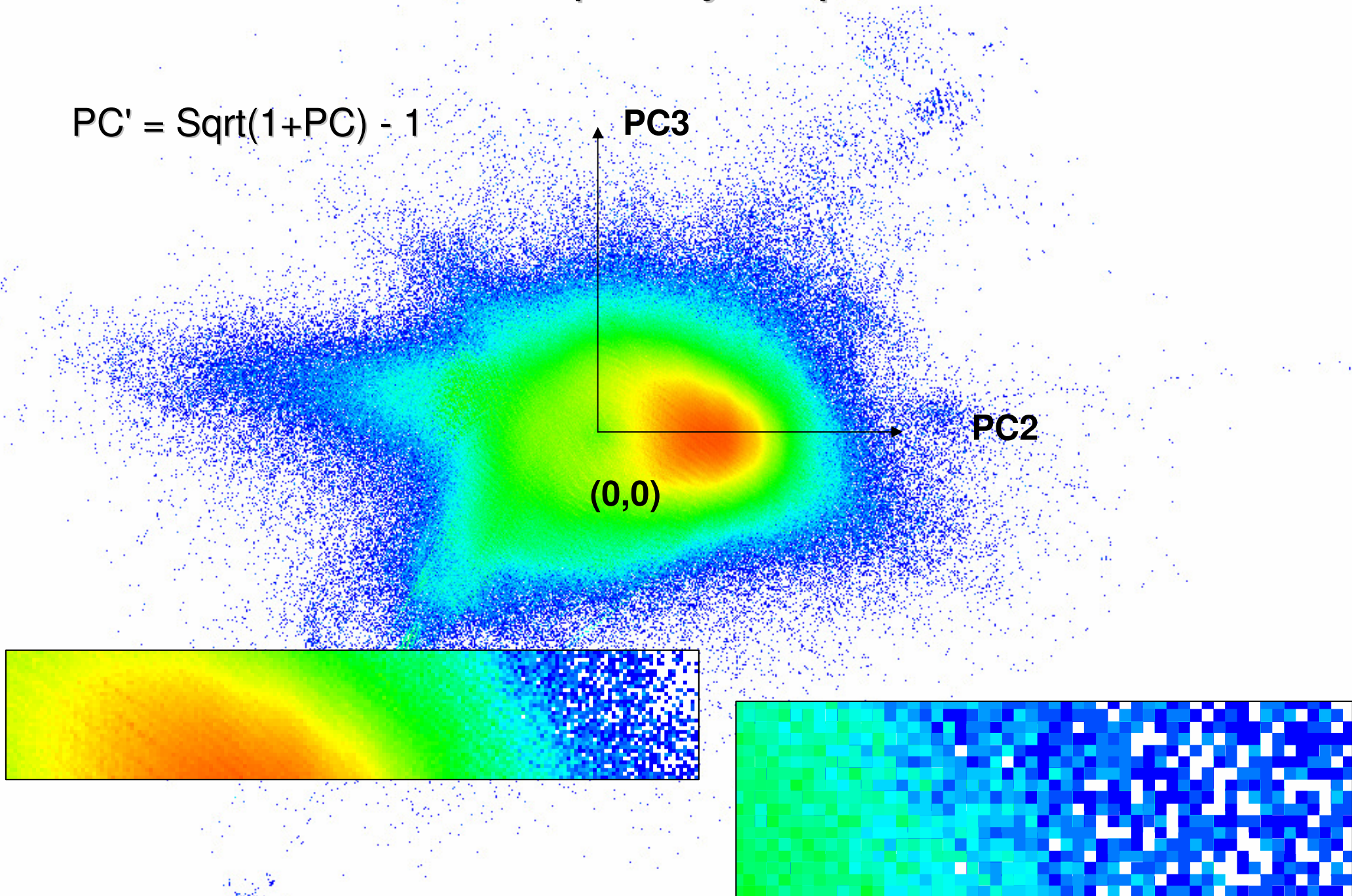
# Frequency map

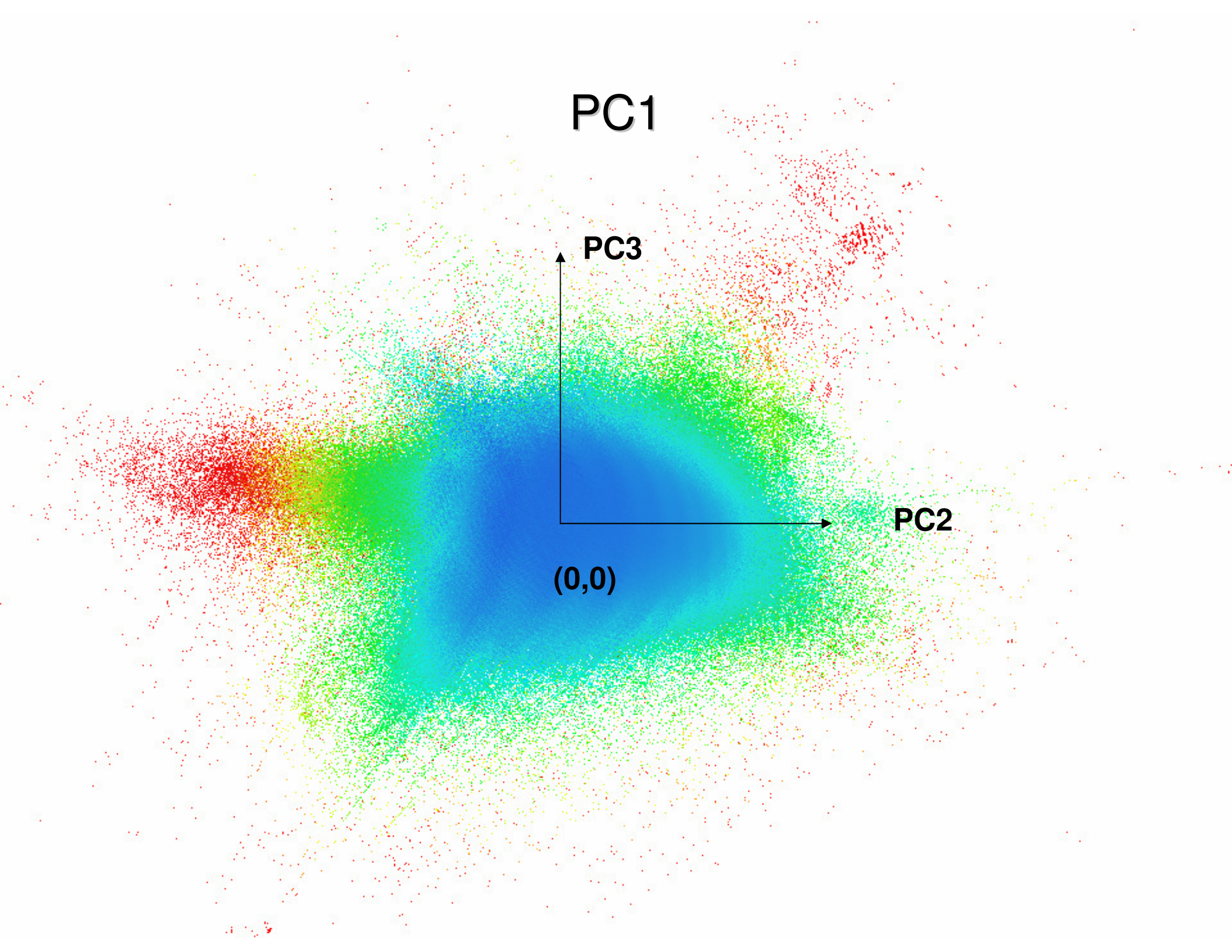
euclidean projection of MQN space



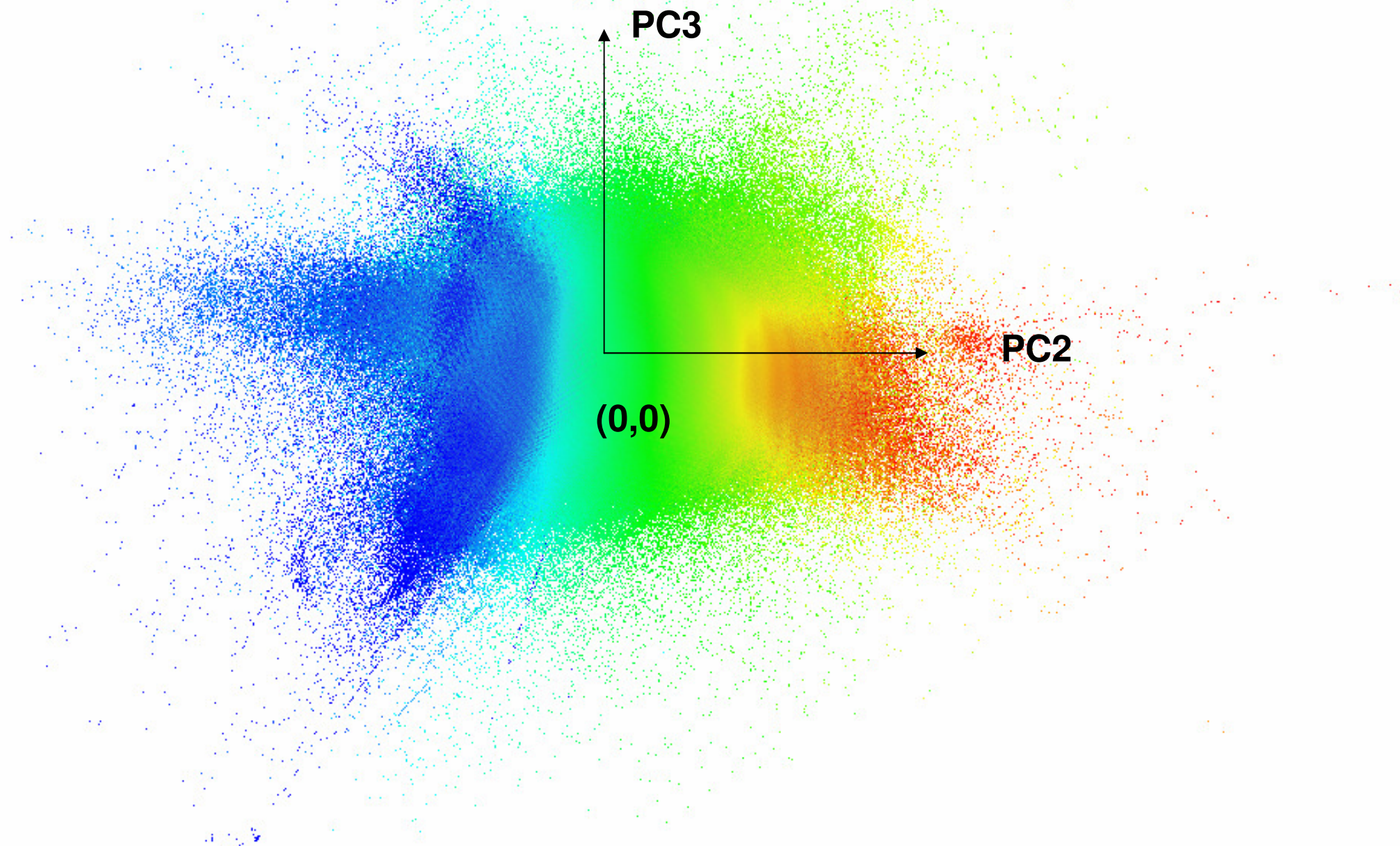
# Frequency map

$$PC' = \text{Sqrt}(1+PC) - 1$$

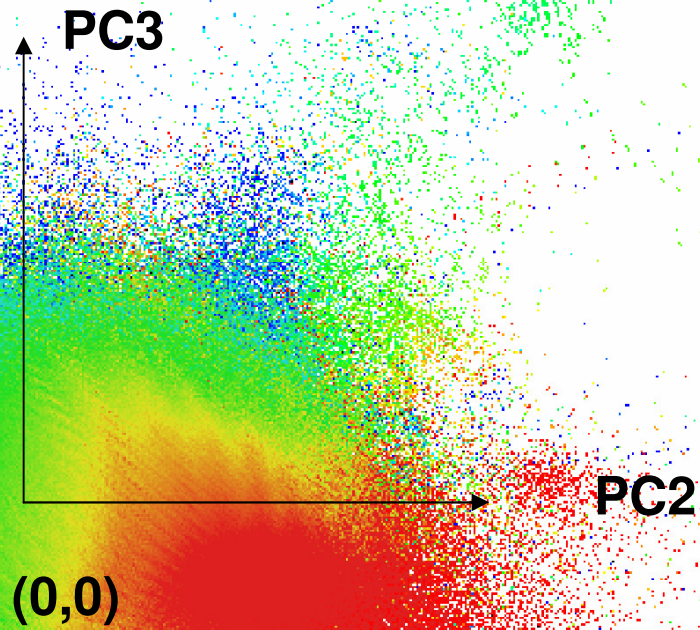




# Ring atom ratio

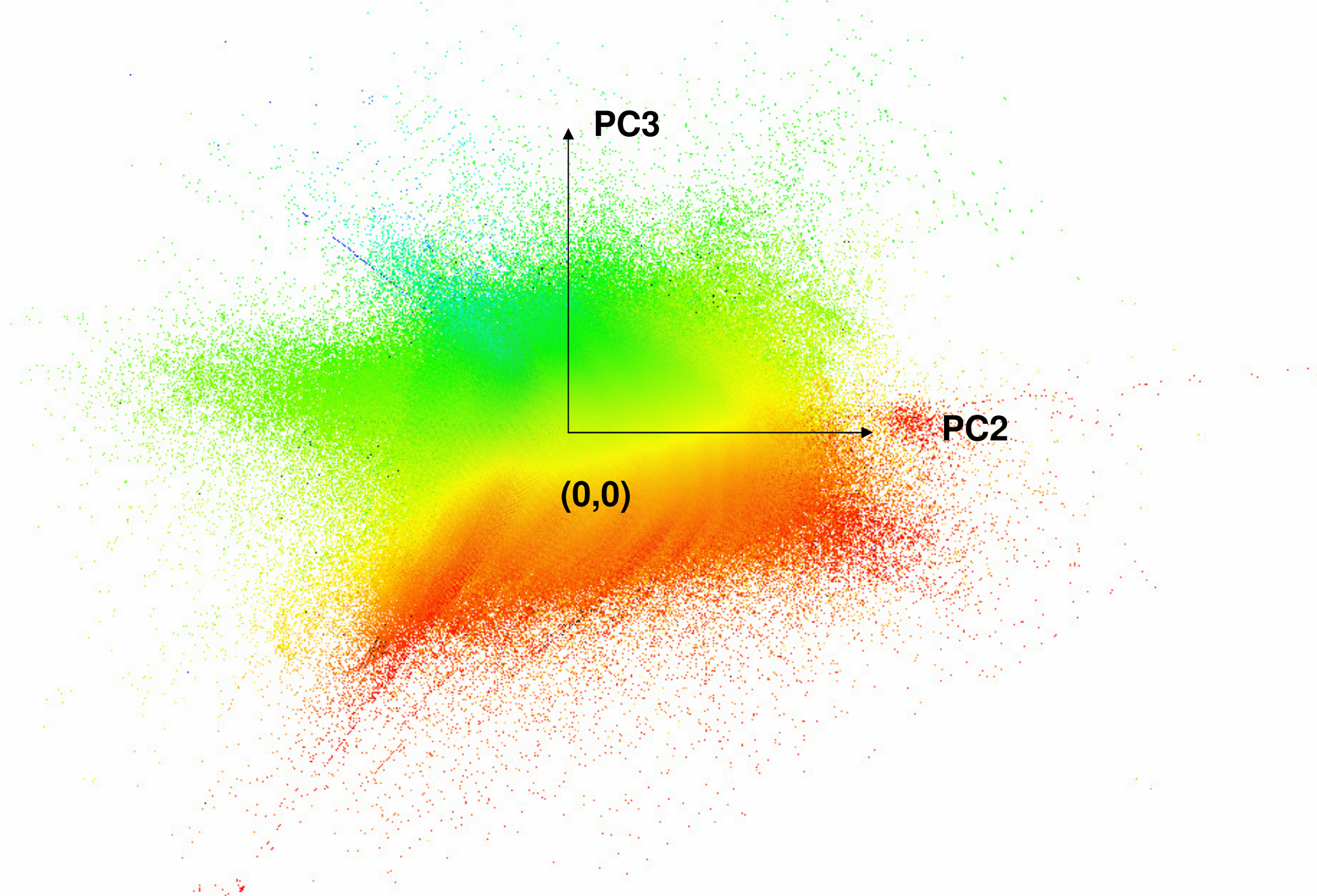


# Cyclic double bond ratio

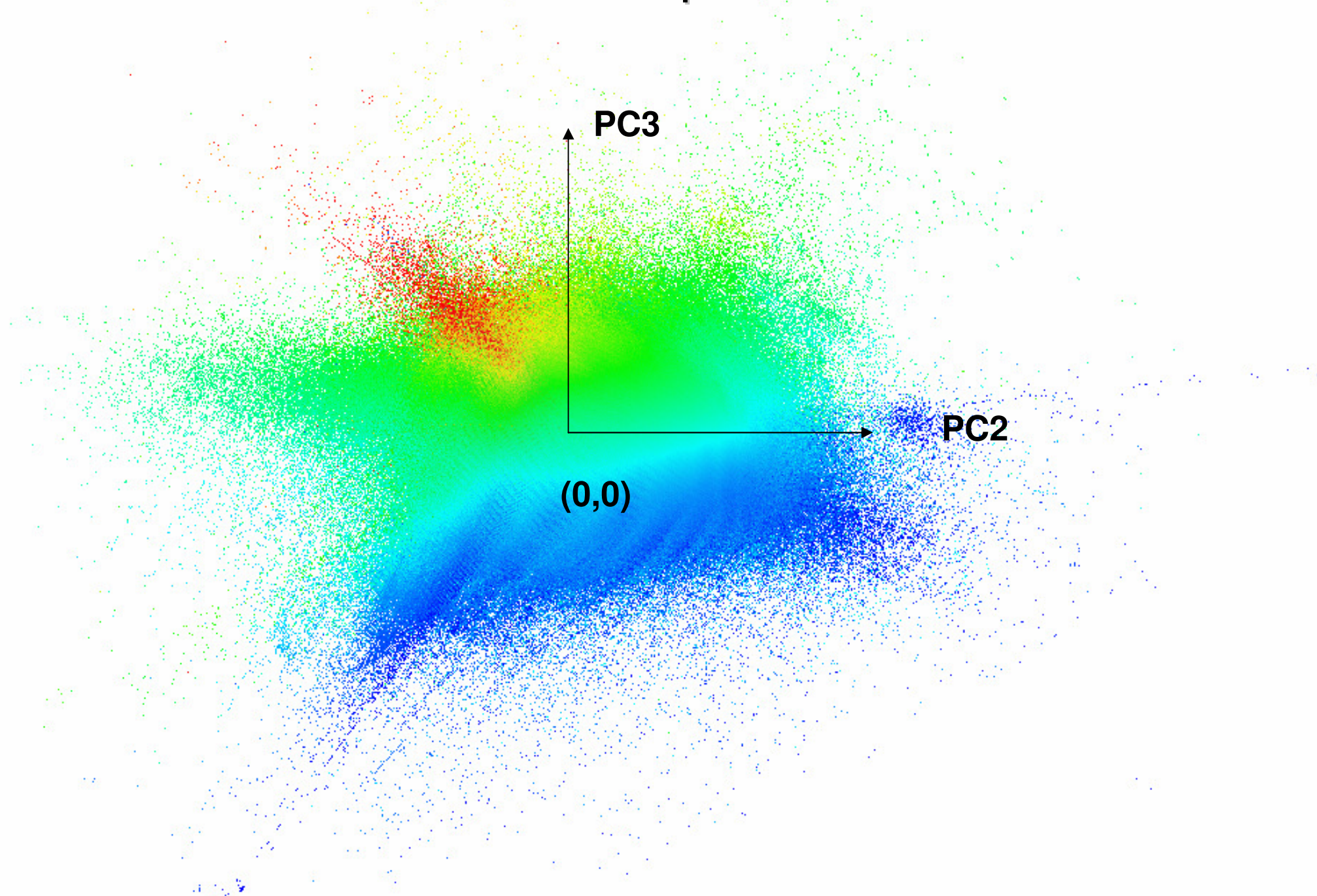




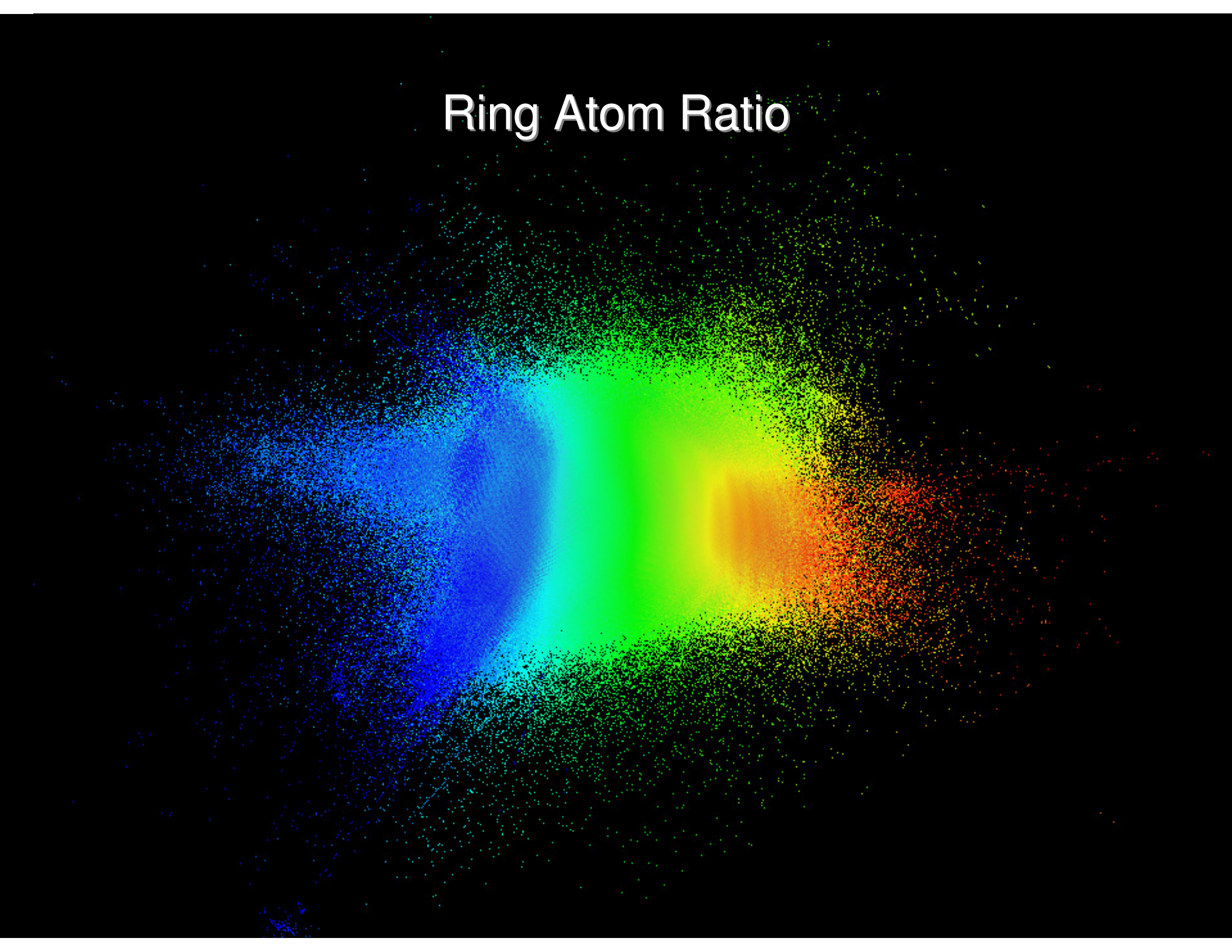
# Carbon ratio

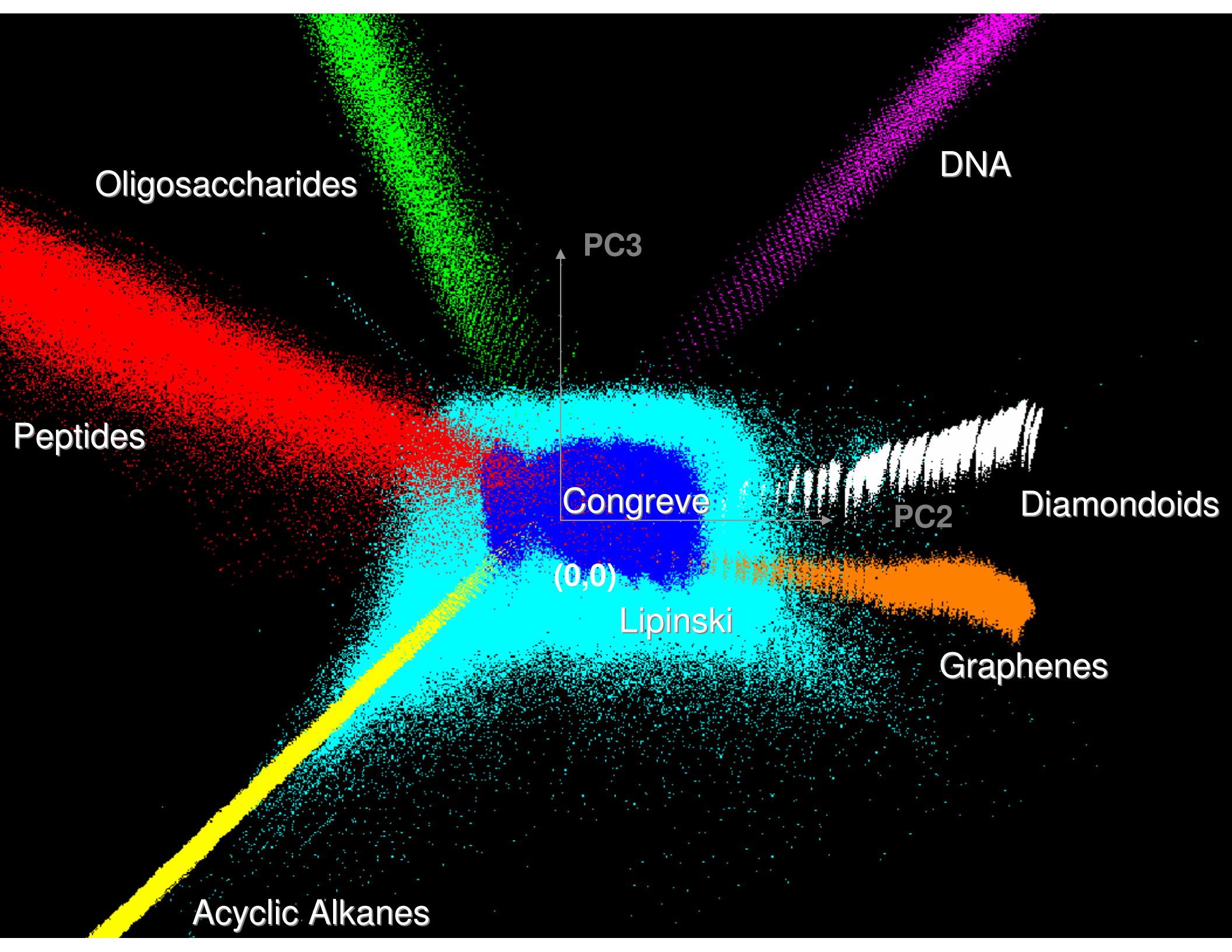


# H-Bond acceptor ratio



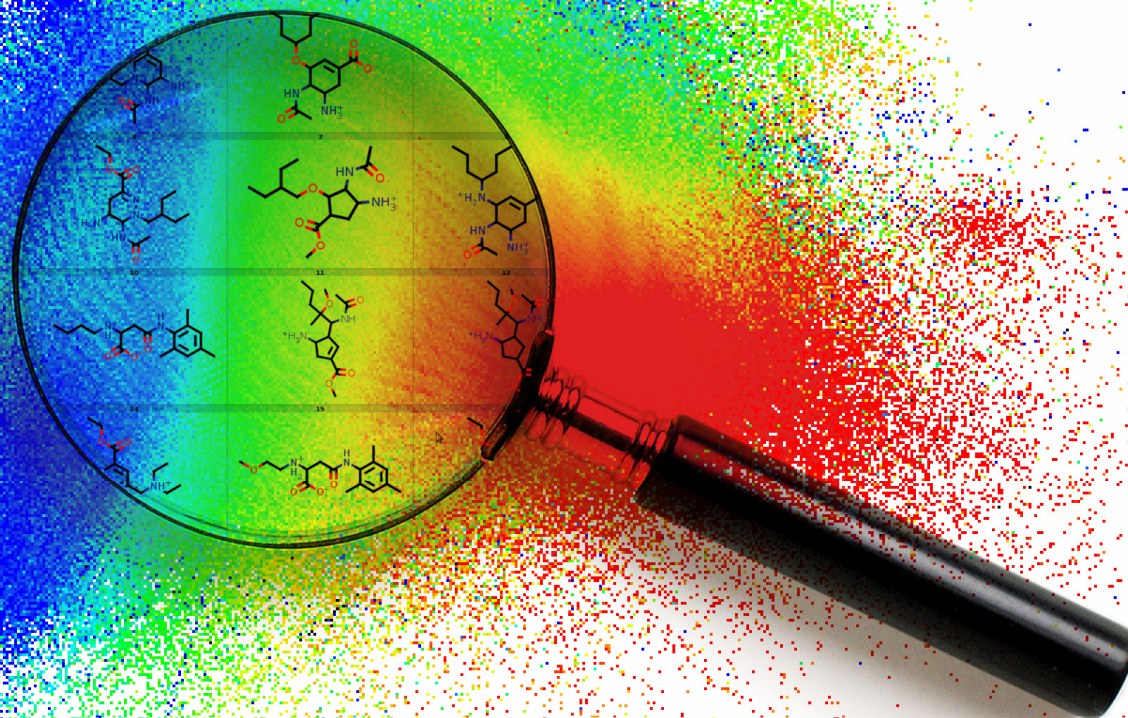
# Ring Atom Ratio





# Scoring Chemical Space

1. GDB, CST
2. MQN
- 3. Scoring**



R. van Deursen, L. C. Blum, J.-L. Reymond, *J. Chem. Inf. Model.* **2010**, *50*, 1924-1934

**Lorenz Blum (GDB subsets)**

E. Luethi et al., *J. Med. Chem.* **2010**, *53*, 7236, and N. Garcia-Delgado et al., *ACS Med. Chem. Lett.* **2010**, *1*, 422-426

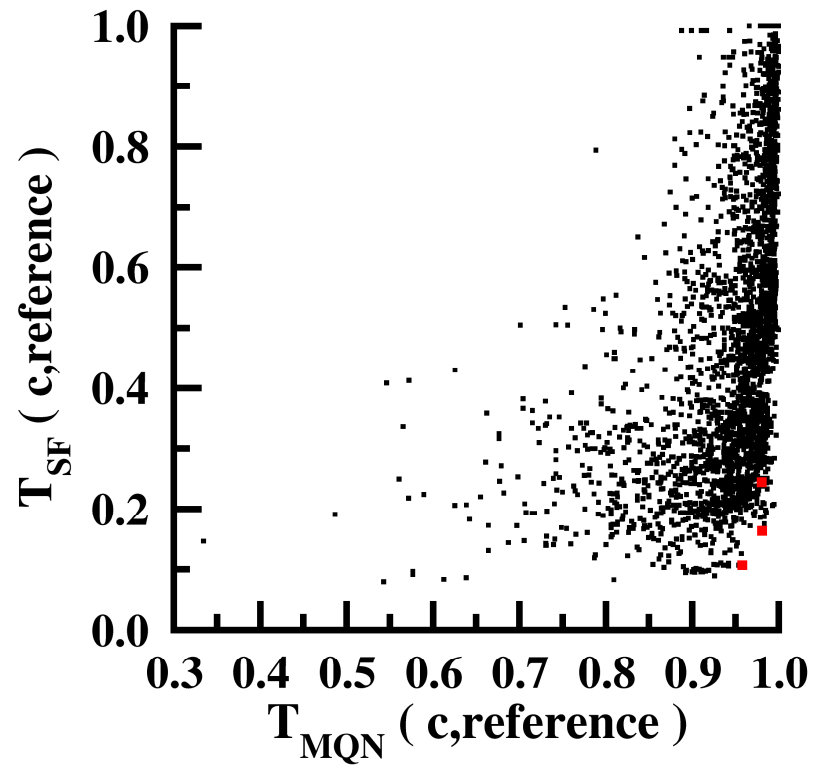
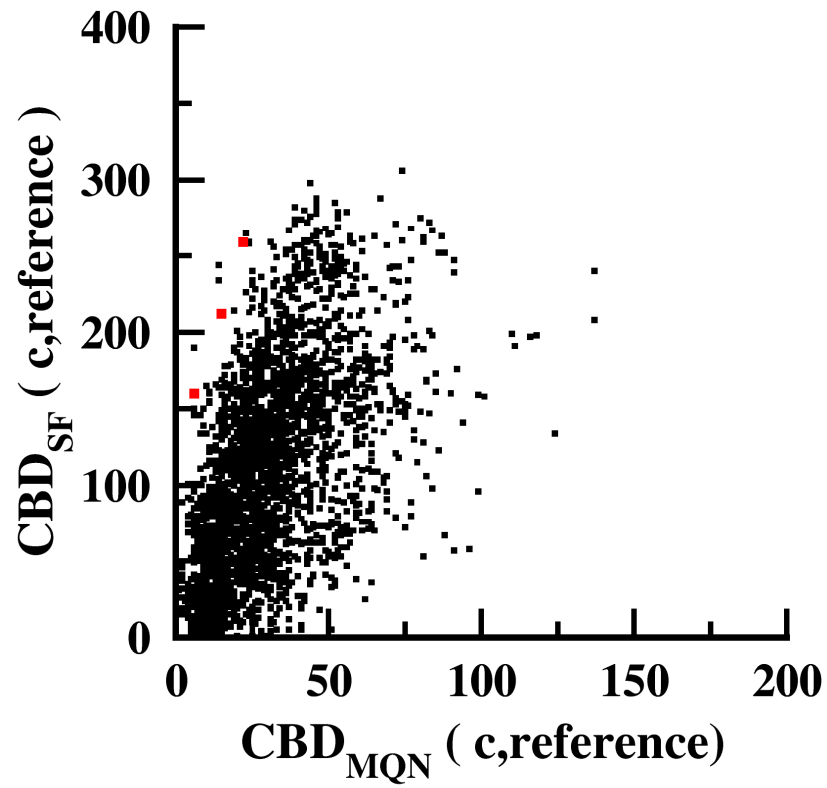
# Enriching the DUD actives from Pubchem

	nr. of actives <sup>b)</sup>	EF <sub>0.1</sub> CBD <sub>MON</sub>	T <sub>MON</sub>	CBD <sub>SF</sub>	T <sub>SF</sub>	EF <sub>1</sub> CBD <sub>MON</sub>	T <sub>MON</sub>	CBD <sub>SF</sub>	T <sub>SF</sub>
<i>Nuclear Hormone Receptors</i>									
AR	79	379.5	379.5	265.6	265.6	46.8	48.1	43.0	41.8
ERagonist	67	507.1	507.1	432.5	387.8	56.7	50.7	58.2	47.8
ERantagonist	39	358.7	358.7	333.1	333.1	51.3	51.3	35.9	41.0
GR	78	166.6	140.9	538.1	563.7	60.2	55.1	53.8	65.4
MR	15	666.2	666.2	466.3	466.3	86.7	86.7	80.0	86.7
PPARg	85	728.9	623.1	870.0	905.3	87.0	84.7	89.4	91.7
PR	27	592.2	555.2	592.2	629.2	59.2	59.2	59.2	70.4
RXRa	20	849.4	599.6	849.4	849.4	95.0	85.0	85.0	100.0
<i>Kinases</i>									
CDK2	72	111.0	83.3	138.8	138.8	20.8	18.1	15.3	16.7
EGFR	475	90.5	67.3	126.2	132.5	25.5	20.8	20.2	27.2
FGFr1	120	191.5	183.2	624.6	724.5	29.2	22.5	74.2	85.0
HSP90	37	378.1	378.1	648.2	702.2	54.0	37.8	70.3	70.3
P38 MAP	454	424.8	380.8	691.2	783.6	59.5	55.9	79.5	89.4
PDGFrb	170	64.7	52.9	82.3	82.3	20.0	18.2	18.2	20.0
SRC	159	188.5	182.3	590.8	659.9	27.7	22.0	71.1	74.2
TK	22	545.1	499.7	726.8	726.8	81.8	81.8	95.4	86.3
VEGFr2	88	102.2	102.2	136.3	193.0	20.5	19.3	20.5	30.7

Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking.  
*J. Med. Chem.* **2006**, *49*, 6789-6801.

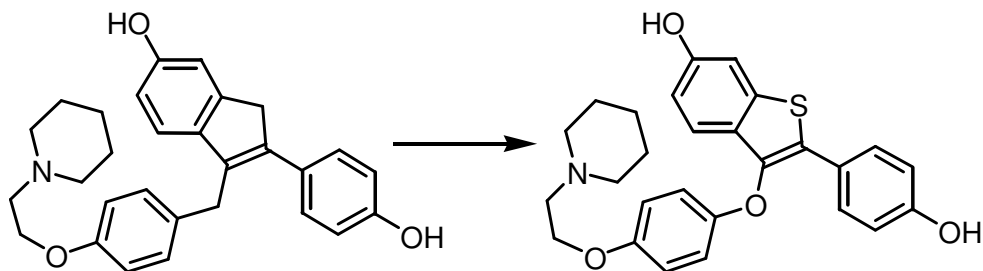


# Lead Hop(p)ing



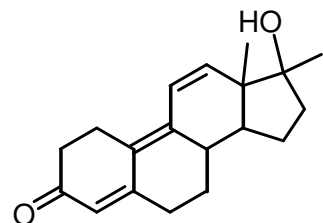


# Lead Hop(p)ing



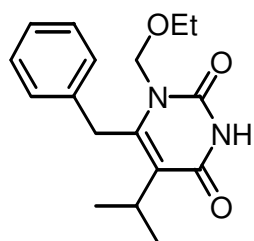
Lead ER antagonist

$CBD_{MQN} = 6$   
( $CBD_{SF} = 160$ )

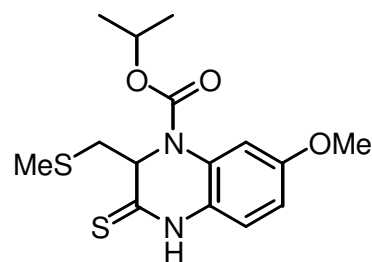


Lead AR inhibitor

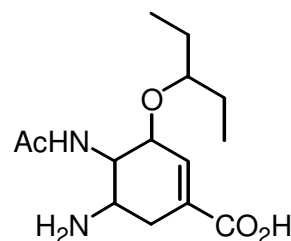
$T_{MQN} = 0.958$   
( $T_{SF} = 0.107$ )



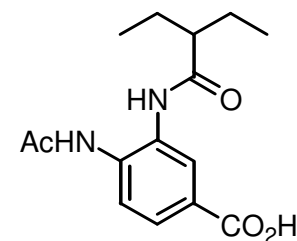
Lead HIV RT inhibitor



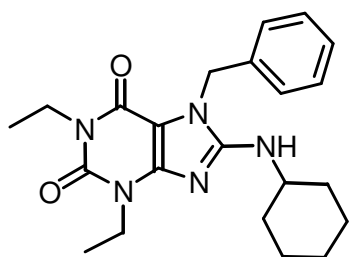
$CBD_{MQN} = 15$   
( $CBD_{SF} = 212$ )



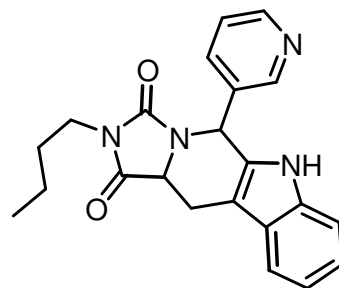
Lead NA inhibitor



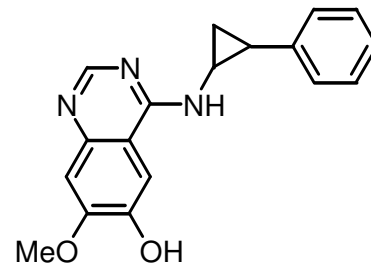
$T_{MQN} = 0.981$   
( $T_{SF} = 0.163$ )



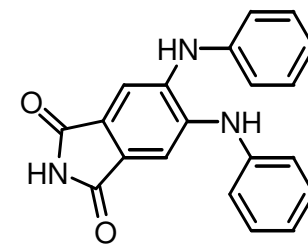
Lead pde5 inhibitor



$CBD_{MQN} = 22$   
( $CBD_{SF} = 259$ )



Lead EGFR inhibitor



$T_{MQN} = 0.980$   
( $T_{SF} = 0.244$ )

