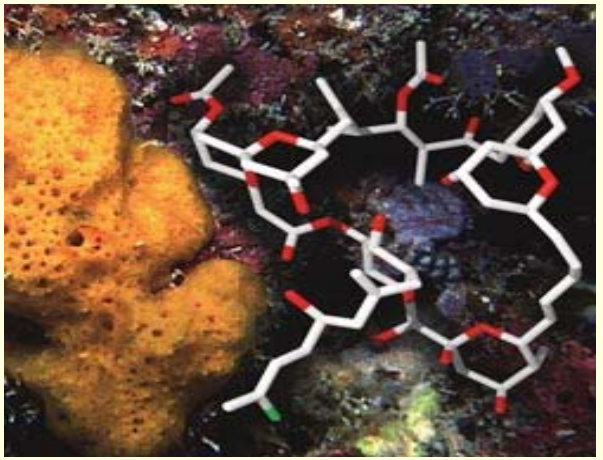
QSAR of Microtubule Stabilizing Dictyostatins

Kia Montgomery
BBSI 2007- University of Pittsburgh
Department of Chemistry, Grambling State University

Billy Day, Ph.D.
Department of Pharmaceutical Sciences, School of Pharmacy
Department of Chemistry
University of Pittsburgh

What is Dictyostatin?

- It is a potent anticancer agent that was discovered from a marine sponge of the genus *Spongia* over a decade ago.
- It is one of the most potent microtubule stabilizing agents discovered to date.



Microtubules

- Polymers made up of α- and β-tubulin heterodimers.
- Tubulin polymerizes at each end with the α-subunit of one tubulin dimer connecting to the β-subunit of the next.
- One end of the microtubule will have the α-subunit (minus or – end) exposed. This is where microtubule shrinking can occur. The other end will have the β-subunit (plus, +, or growing end) exposed.

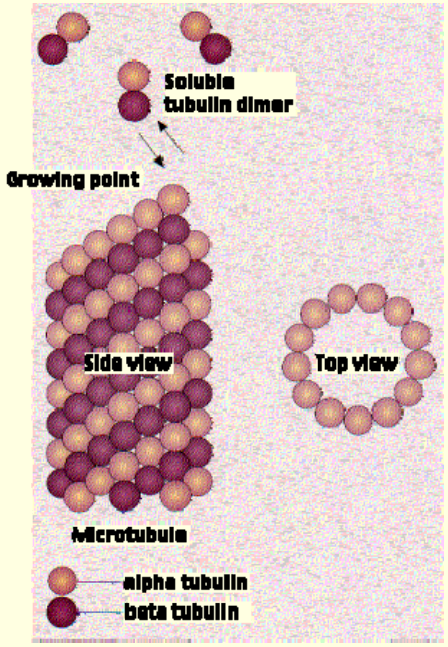


Image from http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/C/Cytoskeleton.html

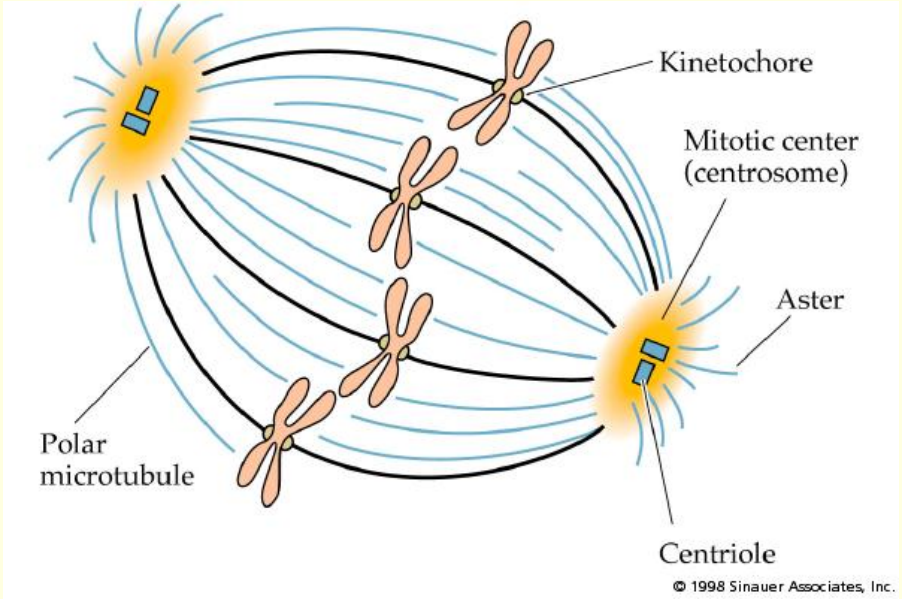


Image from http://www.mie.utoronto.ca/labs/lcdlab/biopic/fig/9.9.jpg

Microtubule Stabilizing Agents

- Some of the most useful chemotherapeutic agents are natural products or natural product analogs. For example, paclitaxel is a natural product that is currently being used to treat patients with breast, lung and ovarian cancers.
- It belongs to a group of chemicals known as taxanes, which function through binding to the β-tubulin subunits of microtubules.
- A number of analogs of paclitaxel, including docetaxel, are also clinically useful anticancer agents.

Discodermolide

- Like paclitaxel, discodermolide, a polyketide natural product, was discovered to be a very potent inhibitor of cancer cell growth. It was isolated in 1990 from the marine sponge *Discodermia dissoluta*. Discodermolide inhibits the growth of human cells by blocking them at G2/ M phase of the cell cycle.

Discodermolide (cont.)

- It was a clinical candidate for cancer chemotherapy due to its high potency in microtubule stabilization and its strong activity against multiple drug resistant cancers.
- Unfortunately, discodermolide only made it to Phase II clinical trials when tested in humans, where it failed due to unexpected toxicity.

Why Is Dictyostatin So Important?

- Because discodermolide showed promising effects, it was important in the fields of chemotherapy and drug discovery to uncover an agent quite similar in structure and activity, but without the undesired toxicity.
- Another marine sponge-derived natural product discovered in 1994, dictyostatin, shares much structural similarity to discodermolide, including identical configurations at all common stereocenters.

Why Dictyostatin?

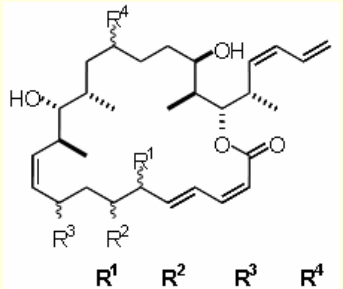
- With the recent withdrawal of discodermolide from clinical development, the importance of uncovering a dictyostatin with the potential for clinical development has increased.
- Several analogs of dictyostatin have been synthesized and some of their biological activities have been measured.

Purpose of Research

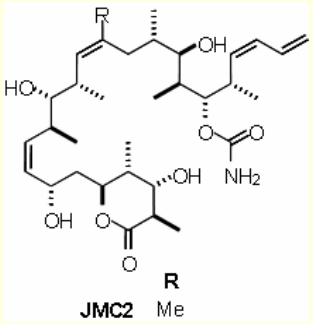
- Using the structures of dictyostatin analogs and their biological activities, along with those of discodermolide and a potent, structurally-related analog, the purpose of this work was to develop a quantitative structure-activity relationship (QSAR) useful in further analog design.

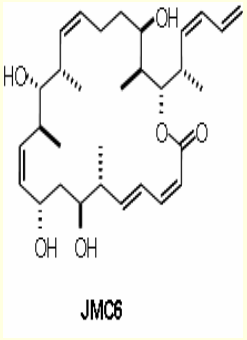
Methodology

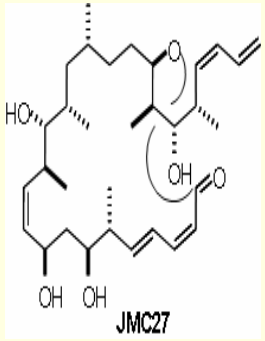
- Molecular models of dictyostatin and its analogs were created (previously) and their global minimum energy conformations were calculated.

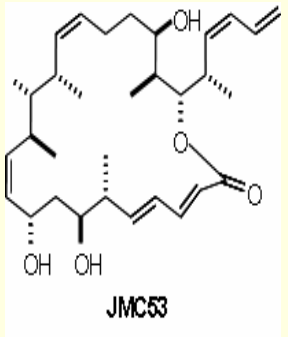


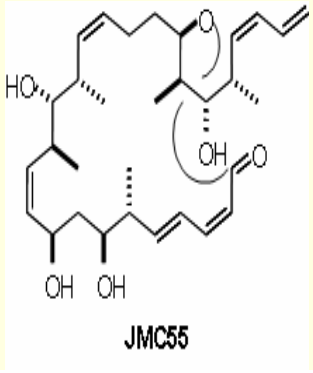
	R^1	R^2	R^3	R^4
JMC1	*R*-Me	*S*-OH	*S*-OH	*S*-Me
JMC4	*R*-Me	*S*-OH	*S*-OH	H
JMC5	*R*-Me	*S*-OH	*S*-OH	*R*-Me
JMC54	*R*-Me	*S*-OH	*R*-OH	H
OL22	*S*-Me	*R*-OH	*S*-OH	*S*-Me
OL23	*S*-Me	*S*-OH	*S*-OH	*S*-Me

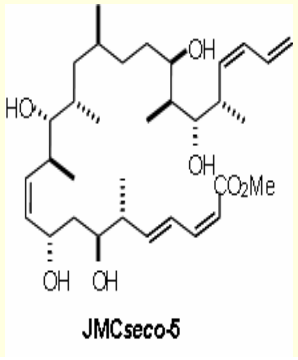






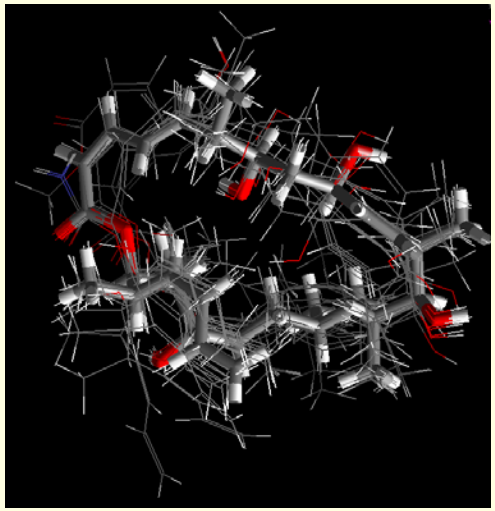




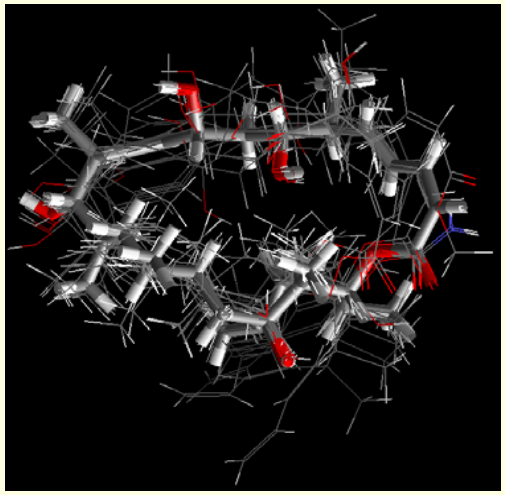


Methodology

- These models were then superimposed to provide maximum structural overlap.



Front View



Back View

- The structural and physiochemical differences for each compound was determined. This was done by calculating a number of **descriptors** for the structures, such as thermodynamic properties, electronic properties, steric (size) properties, linear free energy terms, and many others, using the software suite Cerius^2 (Accelrys, Inc.)

Methodology (descriptors)

- A *descriptor* is any of a number of built-in molecular properties that can be calculated and used to determine QSAR relationships.
- The descriptors for each structure was calculated and then tabulated (in a “study table”)

	Structure	Activity	GFA Predi	GFA Resic	Apol	Dipole-ma	RadOfGyra	Area	MW	Vm	Density	PMI-mag	Rotlbonds	Hbond acc
JMC_1	<56>	9.16	8.70338	0.45662	20175.66	4.414225	4.665468	722.4043	532.7592	555.0742	0.959798	2118.375	7	6
JMC_2	<57>	8.77	8.303794	0.466206	21606.5	1.226052	4.59355	769.202	593.7994	602.036	0.986319	2220.595	21	8
JMC_27	<84>	5.12	5.467358	-0.34736	20175.66	1.813861	4.710114	728.0687	532.7592	555.3413	0.959336	2059.829	9	6
JMC_3	<85>	8.43	8.579341	-0.14934	21093.24	1.025252	4.598209	747.6457	579.7726	585.0822	0.990925	2177.146	21	8
JMC_4	<99>	9.39	9.661477	-0.27148	19662.4	4.458371	4.695098	709.0508	518.7324	539.0338	0.962337	2114.768	7	6
JMC_5	<102>	7.21	7.84152	-0.63152	20175.66	4.151061	4.680878	713.2414	532.7592	555.9184	0.958341	2131.422	7	6
JMC_53	<105>	6.68	7.065049	-0.38505	19977.04	5.869598	4.943129	690.9968	516.7166	532.0684	0.971147	2271.869	7	6
JMC_54	<106>	5.37	5.123087	0.246913	19977.04	4.429769	4.574372	693.1972	516.7166	532.4562	0.97044	1955.15	7	6
JMC_55	<108>	4.3	4.170059	0.129941	20321.84	5.60758	4.722282	696.8469	532.716	541.1891	0.984343	2118.633	9	7
JMC_6	<110>	8.08	7.253235	0.826765	19977.04	4.557536	4.704742	698.6309	516.7166	532.3966	0.970548	2112.345	7	6
JMC_secc	<181>	5.04	5.251131	-0.21113	21174.48	4.204174	4.779849	769.9534	564.8012	593.1836	0.952152	2452.113	26	7
OL_22	<182>	7.77	8.071724	-0.30172	20175.66	4.01984	4.584457	714.7639	532.7592	555.0412	0.959855	2062.113	7	6
OL_23	<186>	8.37	8.14874	0.22126	20175.66	4.520724	4.665046	714.5298	532.7592	556.1132	0.958005	2120.18	7	6
OL_24	<1>	8.52		-0.05235	20175.66	4.022747	4.627701	712.9166	532.7592	555.5934	0.958901	2103.617	7	6

Genetic Function Approximation

- A multiple regression analysis was used to find the most statistically significant descriptors that explain the differences in activity, the fifty percent growth inhibitory concentration (GI_{50}) against 1A9 human ovarian carcinoma cells.
- Because the study table had a far higher number of descriptors than the number of compounds, it was essential to use a non-traditional regression method. Here, the Genetic Function Approximation (GFA) was used to assist in the regression analyses.

Genetic Function Approximation

- The GFA is a method that produces a population of statistically compelling structure-activity models (equations), rather than single models.
- Cerius[2] uses as one available method the GFA to calculate QSARs.[1]
- GFA begins with a population of randomly-constructed QSAR models, where the models are rated using an error measure that estimates each model's relative predictiveness.[1]

GFA: Quantitative Structure-Activity Relationship

- QSAR is a multivariate, mathematical relationship between a set of 2D and 3D physiocochemical properties (descriptors) and biological activity.[1] It allows one to choose the best candidate compounds, based on the biological activity, as well as gain insight into a variety of fundamental biological processes.[1]

Advantages of GFA

- Some advantages of using the GFA approach are:
  - That it is better at discovering combinations of features that take advantage of correlations between multiple features.
  - Use of Freidman's LOF error measure, which estimates the most appropriate number of features, resists over-fitting and allows control over the smoothness of fit.
  - It also provides additional information not available from standard regression analysis, such as the preferred model length and useful partitions of the dataset.[2]

QSAR Equations

The 5 best-scoring equations shared similar information in terms of statistical parameters and descriptors.

Index	Equation (Activity)
1 (100)	= -0.017328 - 0.41287 * "Rotlbonds" + 0.091723* "HF_MOPAC" – 0.223325 * "HF" + 0.0611096 * "Dipole_Mag"
2 (99)	= -26.5019 +0.03692 * "Area" – 0.328785 * "Rotlbonds" – 0.04374 * " HF" + 0.924229 * "Dipole_Mopac"
3 (98)	= -26.3842 + 0.1088 * "Area" + 1.31599 * "Dipole_Mopac" - 0.353032 * "MR" + 0.654779 * LogP
4 (97)	= 1.694349 – 0.305704 * "Rotlbonds" – 0.135425 * "HF" + 0.605182 * "Dipole_Mopac = 0.053736 * "HF_Mopac"
5 (96)	= -48.6935 – 0.200571 * "MR" – 0.270041 * "Rotlbonds" + 1.27617 * "Dipole_Mopac" + 0.117247 * "Area"

- **Rotatable bonds (Rotlbonds) are *structural* descriptors. The number of rotatable bond descriptor counts the number of bonds in the current molecule having rotations that are considered to be significant in molecular mechanics.**
- **The molecular surface area (Area) descriptor is a 3D *spatial* descriptor that describes the van der Waals area of a molecule.**
- **The heat of formation (Hf) is a *thermodynamic* descriptor. It is the enthalpy for forming a molecule from its constituent atoms.**
- **HF_Mopac and Dipole_Mopac (heat of formation and dipole moment calculated by the suite of programs called Mopac) are *quantum mechanical* descriptors.**

Statistical parameters

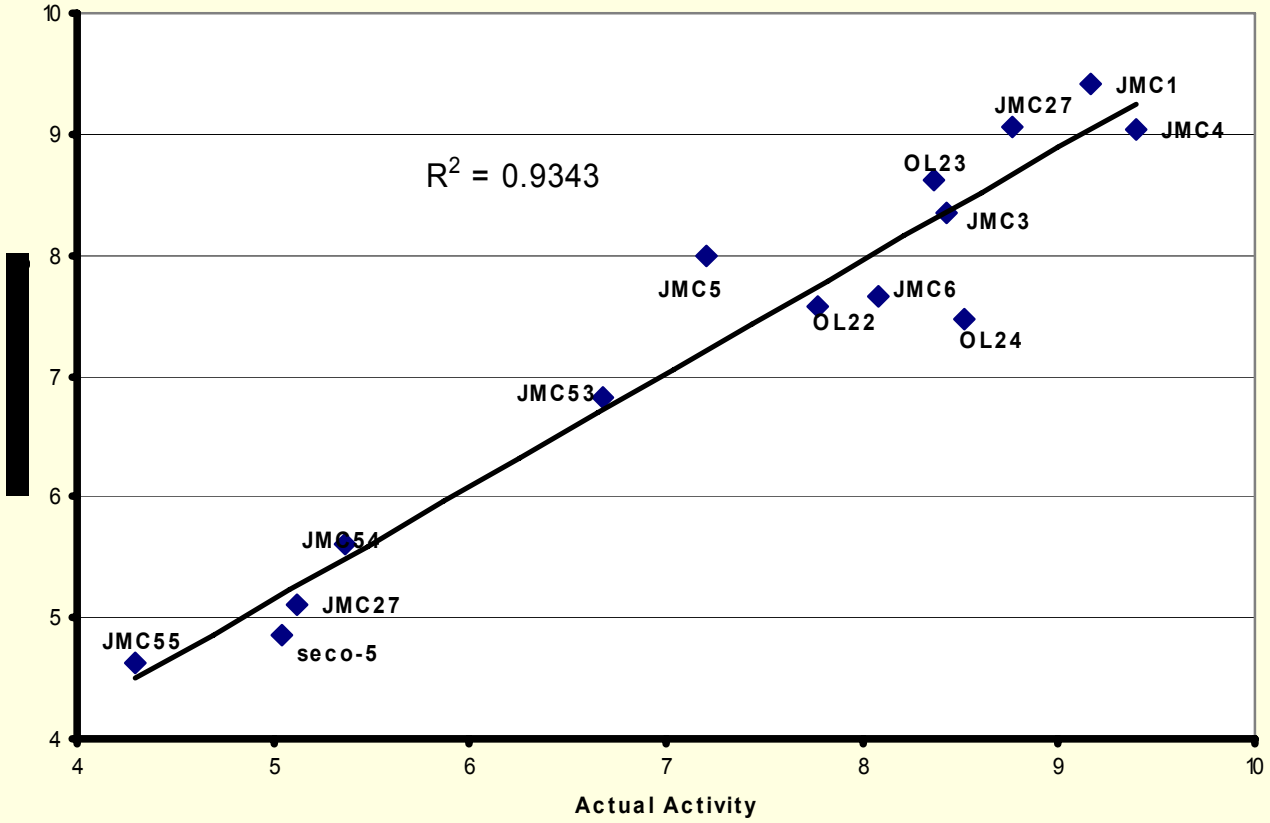
	Lof	r^2	r^2-adj	F-test	N Obs	N Vars	LSE	r	C(P)
1 100)	0.832	0.943	0.918	37.425	14	5	0.153	0.971	- 3.940
2 (99)	0.889	0.939	0.912	34.867	14	5	0.163	0.969	- 3.935
3 (98)	0.930	0.937	0.908	32.210	14	5	0.171	0.968	- 3.932
4 (97)	0.932	0.936	0.908	33.161	14	5	0.171	0.968	- 3.932
5 (96)	0.964	0.934	0.905	31.971	14	5	0.177	0.967	-3. 930

Coefficients of determination

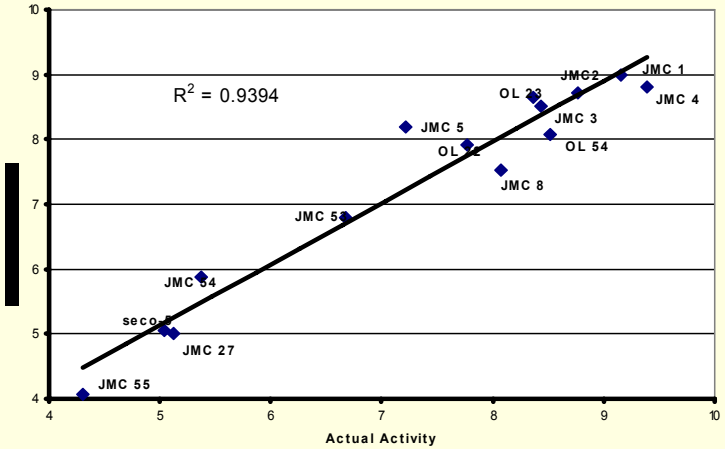
The four-term descriptor equations contained common descriptors such as "Area" and "HF".

Results

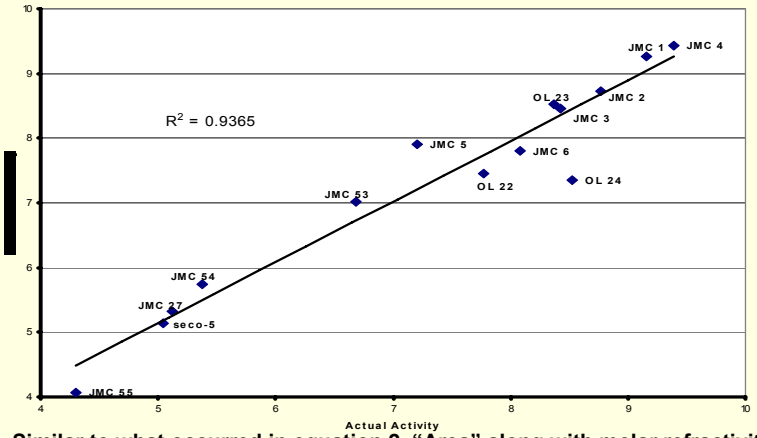
Actual vs. Predicted Activity (antiproliferative potencies against 1A9 human ovarian carcinoma cells) of learning set agents



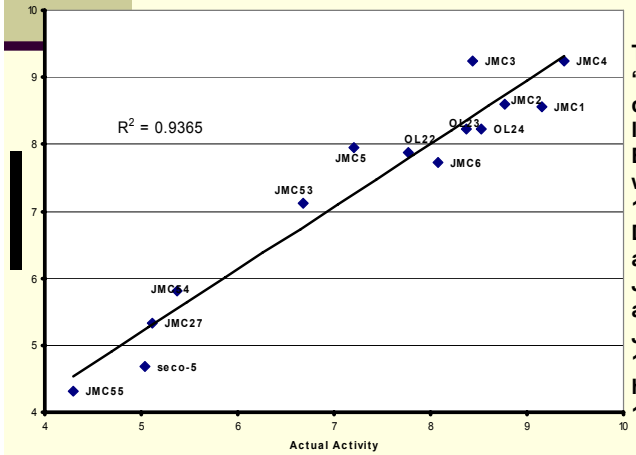
Actual vs. Predicted Activity (antiproliferative potencies against 1A9 human ovarian carcinoma cells) of learning set agents



The descriptor "Area" played a major role in equation 2. Predictions for outliers such as JMC 5 were most wrong (e.g., a D of 25) without taking into account the contribution from the "Area" descriptor. The "Area" descriptor was the most important for correct best predictions of activity for all structures under study in equation 2, such as JMC 54 (D 25), JMC 27 (D 27), and seco-JMC5 (D 28).

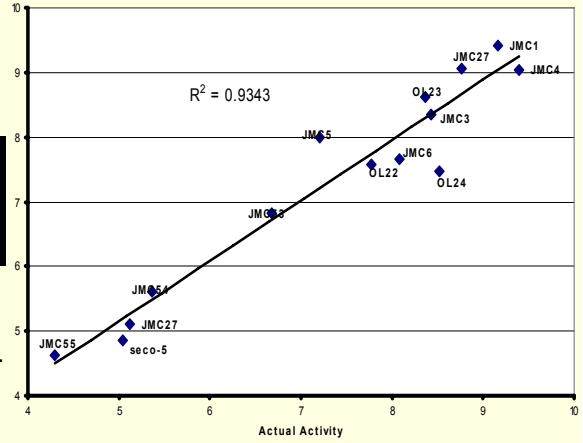


Similar to what occurred in equation 2, "Area" along with molar refractivity "MR" were the most significant contributors to activity predictions in equation 3. For example, JMC 5 had a D of 77 when "Area" was left out, and a D of 55 when "MR" was excluded. JMC 54 had a D of 75 without "Area" and a D of 54 without "MR"; seco-5 had a D of 84 without "Area" and D of 60 without "MR"; and OL 24 had a D of 78 and a D of 55 without "MR".



The heat of formation "HF" and "HF_Mopac" proved to be important descriptors in equation 4, causing the largest chage in values when left out. Examples are JMC 3 with a D of 28 without the presence of "HF" and D of 19 without "HF_MOPAC", JMC 5 had a D of 18 without the presence of "HF" and a D of 14 without "HF_MOPAC", JMC 27 had a D of 19 without the "HF" and a D of 14 without "HF_MOPAC", JMC 54 had a D of 16 "HF" and a D of 13 without "HF_MOPAC", and seco-5 had a D of 26 without the "HF" and D of 17 without "HF_MOPAC".

Like equation 3, the descriptors that were noteworthy in Equation 5 were "Area" and "MR". For example, JMC 5 had a D of 82 without "Area" and D of 31 without "MR"; JMC 27 had a D of 85 without "Area" and D of 31 without "MR"; JMC 54 had a D of 81 without "Area" and D of 30 without "MR"; and seco-5 had a D of 90 without "Area" and a D of 34 without "MR"



Statistics

	Mean	Std. Dev.	Variance	Median	Minimum	Maximum	Range	Sum	Sum Squared	Kurtosis	Skewness	Count
Apol	20345.97	546.597	298768.3	20175.66	19662.4	21606.5	1944.1	284843.5	5.8E+09	1.076432	1.367935	14
Dipole-mag	3.880056	1.47844	2.185786	4.021293	1.025252	5.869598	4.844345	54.32079	239.183	0.245025	-1.01657	14
Area	720.1035	25.68529	659.7343	713.8856	690.9968	769.9534	78.95654	10081.45	7268262	0.189512	1.030932	14
MW	538.3233	24.01645	576.79	532.7592	516.7166	593.7994	77.0828	7536.527	4064586	1.206923	1.435107	14
Vm	556.4663	22.42669	502.9563	555.0577	532.0684	602.036	69.96766	7790.528	4341704	0.014529	0.925979	14
PMI-mag	2144.154	115.4312	13324.37	2118.504	1955.15	2452.113	496.9628	30018.16	64536763	3.516268	1.382571	14
Rotlbonds	10.64286	6.651861	44.24725	7	7	26	19	149	2161	1.174333	1.646504	14
Hbond acceptor	6.428571	0.755929	0.571429	6	6	8	2	90	586	0.936189	1.526395	14
Hbond donor	4.357143	0.744946	0.554945	4	4	6	2	61	273	2.087442	1.874014	14
AlogP98	5.042814	0.405284	0.164255	4.649949	4.185099	5.469	1.283901	70.5994	358.1549	-0.54648	-0.78197	14
AlogP	5.891973	0.456304	0.208214	5.364251	4.896801	6.269001	1.372201	82.48762	488.7215	0.705473	-1.21668	14
Fh2o	-40.5335	4.842275	23.44763	-38.0214	-50.8899	-37.4534	13.43651	-567.469	23306.36	1.068116	-1.55826	14
Foct	-50.0407	2.781917	7.739062	-48.54	-56.29	-48.34	7.950012	-700.57	35157.64	1.752615	-1.74816	14
Hf	-148.786	35.06364	1229.459	-142.804	-215.791	-98.078	117.7129	-2083	325903.3	0.288749	-0.92245	14
LogP	6.971785	1.54549	2.388541	8.084999	3.629999	8.129999	4.5	97.60499	711.532	0.321585	-1.29525	14
MR	156.4131	5.194243	26.98016	155.8457	151.2712	171.2568	19.98552	2189.783	342861.4	4.883425	2.054391	14
MolRef	159.289	3.904368	15.24409	159.1203	154.5717	167.3816	12.80994	2230.046	355419.9	0.666701	1.094516	14
LUMO	2.244553	0.500339	0.250339	2.065011	1.852551	3.39906	1.546509	31.42374	73.78663	3.185064	2.077097	14
Sr	0.9491	0.672305	0.451994	1.483075	0.223771	1.855074	1.631303	13.2874	18.487	-1.99713	0.090325	14
LUMO_MOPAC	-0.30777	0.328563	0.107954	-0.47576	-0.60278	0.44349	1.04627	-4.30879	2.729519	2.86462	1.951972	14
DIPOLE_MOPAC	4.552857	1.021356	1.043169	4.7235	1.669	5.452	3.783	63.74	303.7603	4.319657	-1.87709	14
HF_MOPAC	-260.418	45.76215	2094.175	-254.727	-348.324	-189.994	158.3303	-3645.85	976669.7	0.198485	-0.78139	14
RadOfGyration	4.681778	0.096204	0.009255	4.61992	4.574372	4.943129	0.368757	65.5449	306.987	3.448878	1.542108	14
HOMO	-11.8573	0.125233	0.015683	-11.8844	-12.0127	-11.5527	0.459974	-166.003	1968.554	1.574631	1.226832	14
HOMO_MOPAC	-9.10054	0.089137	0.007945	-9.03203	-9.21342	-8.97368	0.23974	-127.408	1159.581	-1.83546	-0.03774	14
Activity	7.300714	1.703063	2.900423	7.49	4.3	9.39	5.09	102.21	783.9115	-1.14847	-0.59245	14

Conclusion

- The understanding of descriptors used in QSAR equations can provide excellent opportunity for identifying their features and becoming aware of how they affect the activity of each compound.
- Furthermore, the simpler the equation, the easier it is to use that equation to make chemical modifications; and, in general, the more likely it will be useful in drug design.

References

- 1. *C2·GA*. Accelrys Software Inc. 19 Jun. 2007.
- <http://www.accelrys.com/products/datasheets/c2_ga_data.pdf>.
- 2. Cerius2 Modeling Environment, April 1999. San Diego: Molecular Simulations, Inc., 2007.
- 3.(a) Jung, W.-H., Harrison, C., Shin, Y., Fournier, J.-H., Balachandran, R., Raccor, B.S., Sikorski, R.P., Vogt, A., Curran, D.P., Day, B.W. Total synthesis and biological evaluation of C16 analogs of (–)-dictyostatin. *J. Med. Chem.* **2007**, *50*, 2951-2966;
- (b) Fukui, Y., Brückner, A.M., Shin, Y., Balachandran, R., Day, B.W., Curran, D.P. Fluorous mixture synthesis of (–)-dictyostatin and three stereoisomers. *Org. Lett.* **2006**, *8*, 301-304.

Acknowledgments

- Judy Wieber and BBSI faculty and staff
- Billy Day
- University of Pittsburgh
- Grambling State University
- NIH-NIBIB and NSF-EEC, and the National Science Foundation

Thank You

BBSI

It was GREAT!!!

http://www.give-me-money.com/dancing.html