Supporting Information

Activity of Sulfonium Bisphosphonates on Tumor Cell Lines

by

Yonghui Zhang[†], Michael P. Hudock[§], Kilannin Krysiak[†],
Rong Cao[§], Kyle Bergan[†], Fenglin Yin[§], Annette Leon[§] and Eric Oldfield[†,§,*]

[†]Department of Chemistry,

University of Illinois at Urbana-Champaign,

600 South Mathews Avenue, Urbana, Illinois 61801

and

[§]Center for Biophysics and Computational Biology,

University of Illinois at Urbana-Champaign,

607 South Mathews Avenue, Urbana, Illinois 61801

Table of Contents

Supporting Methods:

QSAR Descriptors: In order to provide sufficient granularity, sub-divided van der Waals surface area descriptors[1] (as implemented in MOE) were utilized in the molar refractivity and positive charge terms, as opposed to whole molecule-based descriptors. Subdivided surface area descriptors map specific molecular properties to a fraction of the van der Waals surface of the molecule[2], so for example, a whole molecule formal charge descriptor would report back the net charge of the entire molecule, whereas a sub-divided surface area descriptor for charge would compute total surface area for ranges of charge (i.e. $0.10 < q < 0.15$). Although molecular descriptors, such as those described, do not provide explicit feature locations in three-dimensional space, they are often more than sufficient to provide a "fingerprint" and point of comparison between homologous series of compounds. Specifically, in our case, molar refractivity and partial positive charge descriptors were handled in this manner.

Human FPPS inhibition. Human FPPS was expressed and purified as described previously[3]. Human FPPS assays were carried out using 96 well plates with 200 μL reaction mixture in each well. The condensation of geranyl diphosphate and isopentenyl diphosphate was monitored by a continuous spectrophotometric assay for phosphate releasing enzymes. The reaction buffer contained 50 mM Tris-HCl, 1 mM $MgCl_2$, at pH 7.4. The compounds investigated were pre-incubated with the enzyme for 30 minutes at 20°C. The IC_{50} values were obtained by fitting the inhibition data to a normal dose-response curve using Origin 6.1 (OriginLab Corporation, Northampton, MA, www.OriginLab.com).

Crystallization and x-ray data collection for human FPPS·bisphosphonate complex. Crystals of human FPPS complexed with Mg and **43** were obtained based on the methods described by Kavanagh *et al.*[3], with slight modification. FPPS was incubated with 2.5 mM bisphosphonate and 2.5

mM $MgCl_2$ overnight on ice before setting up the drops. Crystals were grown at room temperature in sitting drops by mixing 2 μL of protein solution with 1 μL of precipitant, which consisted of 40% (v/v) of either polyethylene glycol 2,000 or 4,000 and 0.1 M phosphate/citrate buffer, pH 4.2. Diffraction data were obtained at 100 K using an ADSC Q315 CCD detector at the Brookhaven National Synchrotron Light Source, beamline X29 (λ=1.1 Å).

Crystallization and x-ray data collection of *T.brucei* FPPS·bisphosphonate complex. Crystallization was carried out essentially as reported previously by Mao *et al,*[4] with **43** used as the inhibitor. Diffraction data were obtained at 100 K using an ADSC Q4 CCD detector at the Brookhaven National Synchrotron Light Source beamline X8C (λ=1.1 Å).

Structure determination of human FPPS·bisphosphonate complex. For structure determination, the human FPPS structure (1YV5)[3] minus the risedronate ligand was used as a search model using the molecular replacement method. Rigid body refinement was applied to the model obtained using AMoRe[5]. The crystal structure was then further refined by using Shelxl-97[6], with bisphosphonate densities readily identified. Rebuilding and fitting the ligand in the *2Fo-Fc* electron density map was carried out by using the program *O*[7].

Structure determination of *T.brucei* FPPS-bisphosphonate complex. The crystal structure of the *T.brucei* FPPS bisphoshponate complex was determined by using the molecular replacement method using the program AMoRe[5]**.** The previously solved *T.brucei* FPPS structure (2EWG)[8] minus the minodronate ligand was used as a starting model. The densities of the ligands were readily identified. After iterative rounds of refinement using CNS[9] and rebuilding and ligand fitting using *Coot*[10], the structures had the final refinement statistics shown in Table 4 in the text. As expected the refinement statistics are better than those of the human enzyme, due to the lower resolution of the human FPPS dataset.

Supporting Figures:

Figure S1: Experimental and predicted activity correlation matrix. Correlation between data sets represented as an R value. Experimental tumor cell activity and CoMSIA modeling results, both alignments, all three lines shown.

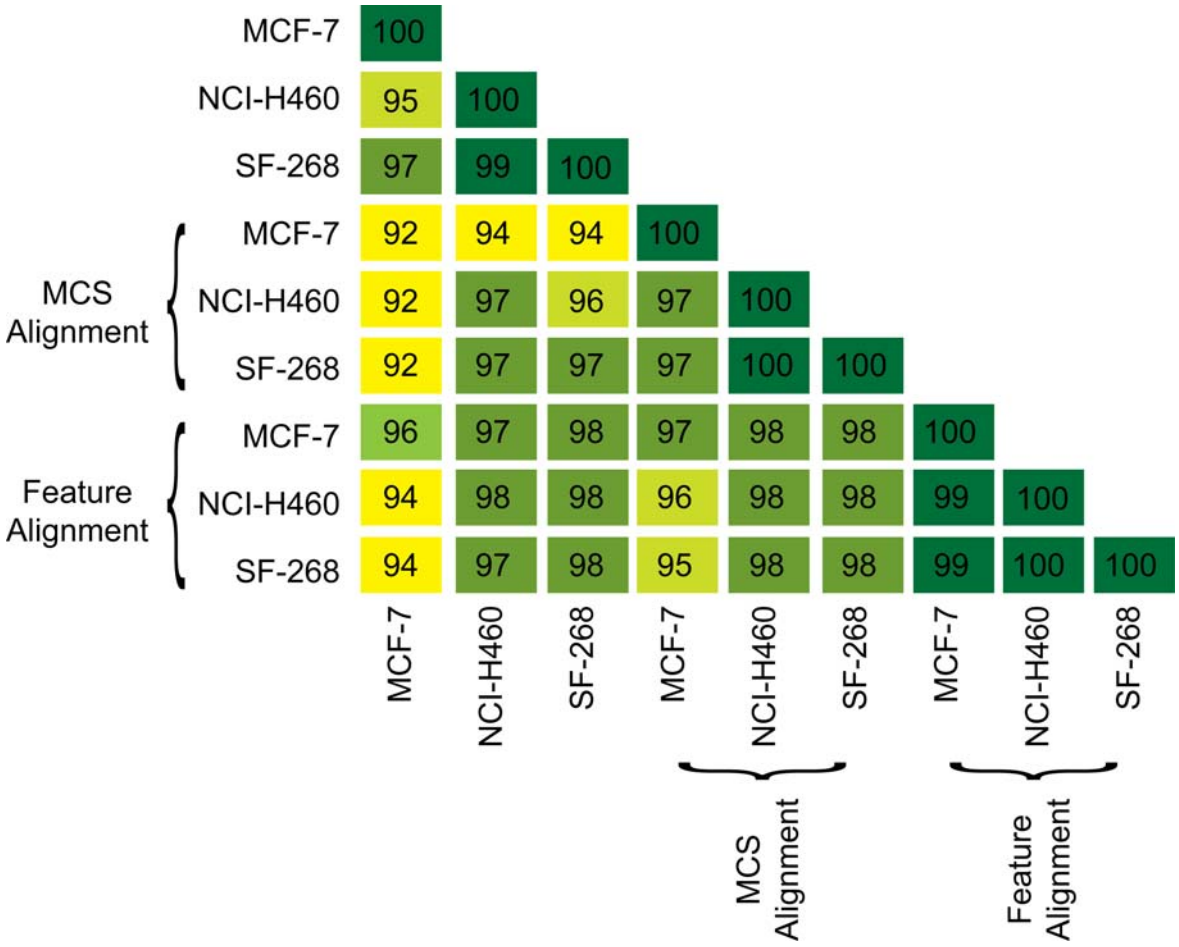


Figure S2: Molecular Descriptor QSAR results for MCF-7 and SF-268 cell lines showing overall good correlation between experimental and predicted pIC_{50} (r^2 = 0.82 and 0.78 respectively)(pIC_{50} = -$\log_{10}$[IC_{50} M])(● = training set, ○ = test set).

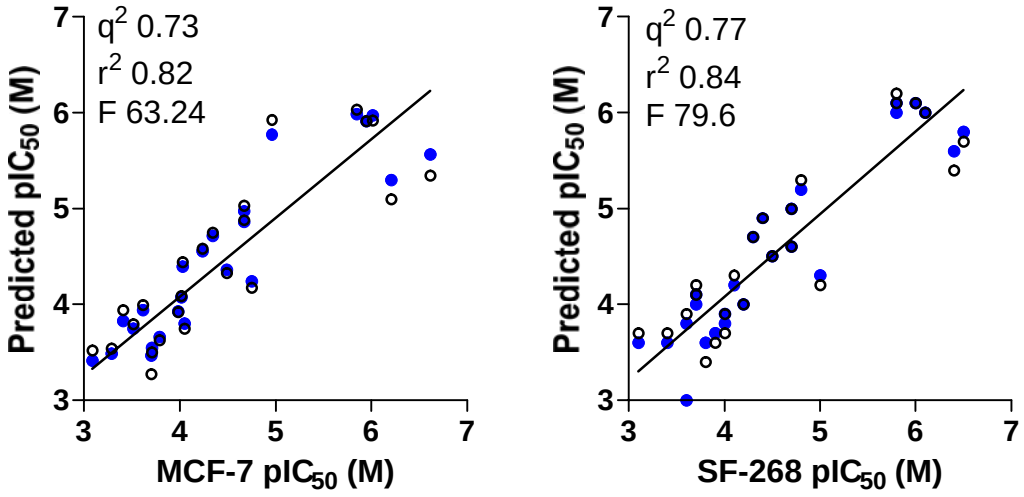


Figure S3: Hologram QSAR (HQSAR) results for MCF-7 and SF-268 cell lines, showing experimental versus predicted r^2 of 0.76 and 0.72 respectively.

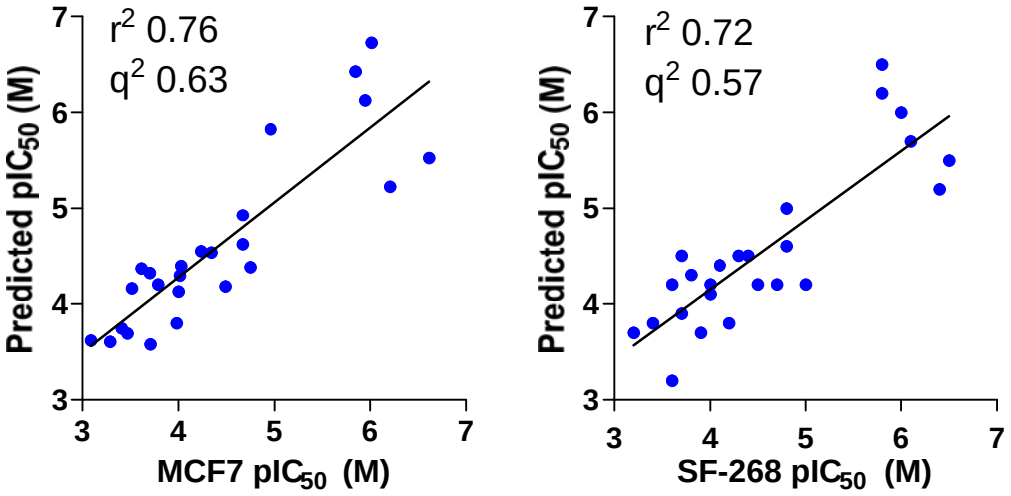
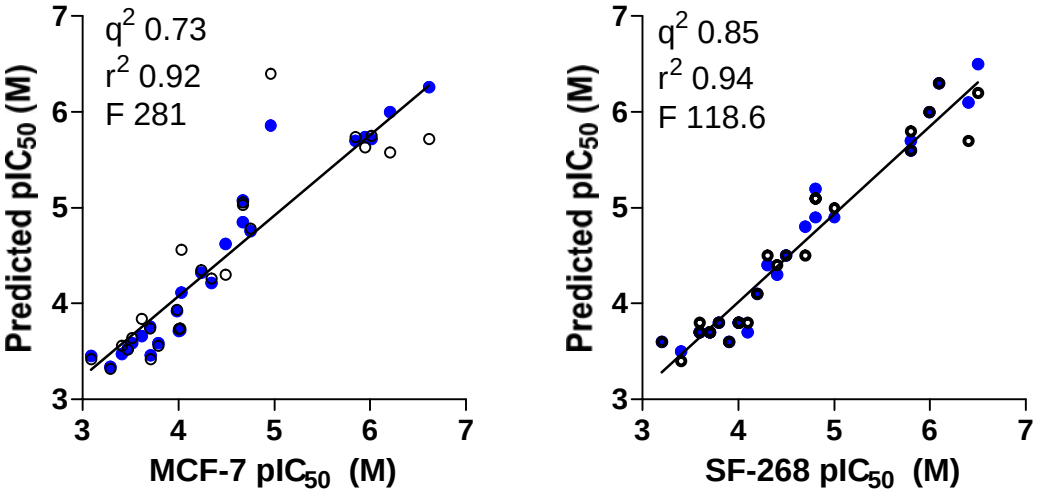


Figure S4: CoMSIA results for MCF-7 and SF-268 cell lines, q^2 = 0.73 and 0.94 respectively. (pIC_{50} = -$\log_{10}$[IC_{50} M])(● = training set, ○ = test set).



Supporting Tables:

Table S1: Microanalysis

Cpd	Compound Name	% Carbon		% Hyrodgen		% Nitrogen	
		Theory	Found	Theory	Found	Theory	Found
14	2-[tetrahydrothiophenium-1-yl]-1,1-bisphosphonic acid	25.67	25.39	5.21	5	0	0
15	1-(2-Hydroxy-2,2-bis-phosphono-ethyl)-tetrahydro-thiophenium	24.66	24.73	4.83	4.81	0	0
16	1-(3-Hydroxy-3,3-bis-phosphono-propyl)-tetrahydro-thiophenium	27.46	26.9	5.27	5.21	0	0
17	1-(3-Hydroxy-3,3-bis-phosphono-propyl)-tetrahydro-thiopyranium	27.46	27.12	5.27	5.25	0	0
18	1-(3-Hydroxy-3,3-bis-phosphono-propyl)-tetrahydro-thiopyranium	29.18	28.66	5.82	5.49	0	0
19	1-(4-Hydroxy-4,4-bis-phosphono-butyl)-tetrahydro-thiophenium	29.34	28.79	5.76	5.21	0	0
20	(2-Hydroxy-2,2-bis-phosphono-ethyl)-methyl-(3-phenyl-propyl)-sulfonium	38.18	37.81	5.55	5.06	0	0
21	(2-Hydroxy-2,2-bis-phosphono-ethyl)-methyl-(3-phenoxy-propyl)-sulfonium	36.46	36.12	5.35	5.21	0	0
22	2-(phenylpentylmethylsulfonium-1-yl)ethylidene-1,1-bisphosphonic acid	42.97	42.9	6.44	6.27	0	0
23	(phenylbutylmethylsulfonium-1-yl)ethylidene-1,1-bisphosphonic acid	41.78	41.79	6.1	5.96	0	0
24	2-(methylphenylpropylsulfonium-1-yl)-ethylidene-1,1-bisphosphonic acid	39.67	39.91	5.83	5.62	0	0
25	(3-Hydroxy-3,3-bis-phosphono-propyl)-methyl-(3-phenyl-propyl)-sulfonium	39.7	39.64	5.89	5.64	0	0
26	(3-Hydroxy-3,3-bis-phosphono-propyl)-methyl-(3-phenoxy-propyl)-sulfonium	35.46	35.71	5.26	5.37	0	0
27	(3-Hydroxy-3,3-bis-phosphono-propyl)-methyl-phenethyl-sulfonium	35.13	34.67	5.16	5.09	0	0
28	1-(3-Hydroxy-3,3-bis-phosphono-propyl)-3-phenyl-tetrahydro-thiophenium	34.75	34.57	5.38	5.35	0	0
29	2-(S-methyl,S-4butyric acidphenyl-sulfonium-1-yl)ethylidene-1,1-bisphosphonic acid	44.06	44.24	6.31	6.2	0	0
30	2(methyldodecylsulfonium-1-yl)ethylidene-1,1-bisphosphonic acid	43.96	43.95	8.51	8.54	0	0
31	2-(methyldecylsulfonium-1-yl)ethylidene-1,1-bisphosphonic acid	40.9	40.58	8.08	7.76	0	0
32	2-(methyltetradecylsulfonium-1-yl)ethylidene-1,1-bisphosphonic acid	46.25	46.23	8.9	8.82	0	0
33	2-(methyltridecylsulfonium-1-yl)-1,1-bisphosphonic acid	43.85	43.6	8.79	8.46	0	0
34	2-(methylundecylsulfonium-1-yl)-1,1-bisphosphonic acid	41.17	41.32	8.39	8.09	0	0

35	2-(methylicosylsulfonium-1-yl)-1,1-bisphosphonic acid	50.11	50.02	9.22	9.44	0	0
36	2-(methyloctadecylsulfonium-1-yl)-1,1-bisphosphonic acid	46.11	45.88	8.4	8.81	0	0
37	2-(methylhexadecylsulfonium-1-yl)ethylidene-1,1-bisphosphonic acid	47.33	47.18	9.28	8.91	0	0
38	2-(methyloctylsufonium-1-yl)ethylidien-1,1-bisphosphonic acid	37.93	37.77	7.52	7.58	0	0
40	(2-Hydroxy-2,2-bis-phosphono-ethyl)-methyl-propyl-sulfonium	18.98	18.7	4.94	4.56	0	0
41	2-(methylpropylsulfonium-1-yl)-ethylidene-1,1-bisphosphonic acid	25.9	25.54	5.8	5.74	0	0
42	(3-Hydroxy-3,3-bis-phosphono-propyl)-dimethyl-sulfonium	22.3	22.64	5.44	5.08	0	0
43	(2-Hydroxy-2,2-bis-phosphono-ethyl)-dimethyl-sulfonium	18.05	18.43	4.54	4.7	0	0
44	2-(methylarmylsulfonium-1-yl)ethylidene 1,1-bisphosphonic acid	30.13	29.81	6.76	6.34	0	0
46	(3-Hydroxy-3,3-bis-phosphono-propyl)-methyl-propyl-sulfonium	25.46	25.67	5.19	5.54	0	0

Table S2: Relationship between human FPPS inhibition and molecular diameter[c].

Cpd[a]	Human FPPS IC_{50} (μM)	Human FPPS pIC_{50} (M)[b]	diameter[c]
43	1.03	6.0	5
42	9.00	5.0	6
41	0.53	6.3	7
44	1.08	6.0	9
38	0.36	6.4	12
31	0.24	6.6	14
30	0.26	6.6	16
32	8.47	5.1	18
36	7.45	5.1	20
37	45.60	4.3	20
36	7.45	5.1	22
35	27.40	4.6	24

[a] Structures shown in Figures 1-3

[b] pIC_{50} = -log10[IC_{50} (M)]

[c] Largest vertex eccentricity, Petitjean, M. J. Chem. Info., 32, 331-337, 1992

Table S3: Molecular Descriptor QSAR output for NCI-H460 model (summary of these results shown in Table 2 in the text).

```
QuaSAR-Model(PLS): /Volumes/hudock/MOE/sulfonium/June2007/qsar_1_062107.mdb
Sat Jul 21 11:51:43 2007

Activity Field               : NCI_pIC50
Weight Field                 :
Condition Limit              : 1e+06
Component Limit              : 0

Observations                 : 25
Descriptors                  : 4
Components Used              : 4
Condition Number             : 597321.55

ROOT MEAN SQUARE ERROR (RMSE): 0.39158
CORRELATION COEFFICIENT (R2) : 0.84897
CROSS-VALIDATED RMSE         : 0.48660
CROSS-VALIDATED R2           : 0.77138

ESTIMATED LINEAR MODEL

NCI_pIC50 =
         4.74682
        +0.32175 * b_Cont_Rot
        -0.01398 * SMR_VSA7
        -0.02456 * PEOE_VSA+2
        -0.00083 * weinerPath

ESTIMATED NORMALIZED LINEAR MODEL (SD = Standard Deviation)

NCI_pIC50 / SD(NCI_pIC50) =
         4.71101
        +1.82212 * b_Cont_Rot / SD(b_Cont_Rot)
        -0.36315 * SMR_VSA7 / SD(SMR_VSA7)
        -0.26633 * PEOE_VSA+2 / SD(PEOE_VSA+2)
        -0.90460 * weinerPath / SD(weinerPath)

RELATIVE IMPORTANCE OF DESCRIPTORS

1.000000  b_Cont_Rot
0.199303  SMR_VSA7
0.146166  PEOE_VSA+2
0.496457  weinerPath
```

b_cont_Rot = Number of contiguous rotatable bonds.
SMR_VSA7 = Molar refractivity using the sub-divided molecular surface area method.
PEOE_VSA+2 = Partial Gasteiger charge using the sub-divided molecular surface area method.
weinerPath = Describes the compactness, branching and substitution of the molecule.

Table S4: Molecular Descriptor QSAR Results for MCF-7 cell line.

Cpd[a]	Contiguous Rotatable Bonds	Weiner Path	Molar Refractivity (SMR VSA7)	Gasteiger Charge (VSA+2)	MCF7 pIC_{50} (M)	Predicted pIC_{50} (M)	CV Predicted[b] pIC_{50} (M)	Residual
30	14	1818	110.91148	12.37335	6.6	5.6	5.3	1.3
31	12	1353	110.91148	12.37335	6.2	5.3	5.1	1.1
35	22	4718	110.91148	12.37335	6.0	6.0	5.9	0.1
37	18	3044	110.91148	12.37335	6.0	5.9	5.9	0.0
36	20	3821	110.91148	12.37335	5.8	6.0	6.0	-0.2
32	16	2379	110.91148	12.37335	5.0	5.8	5.9	-1.0
20	6	1154	77.585464	10.677375	4.8	4.2	4.2	0.6
22	9	1464	77.585464	12.37335	4.7	4.9	4.9	-0.2
38	10	976	110.91148	12.37335	4.7	5.0	5.0	-0.4
39	10	1261	110.91148	30.056063	4.5	4.4	4.3	0.2
23	8	1249	77.585464	12.37335	4.3	4.7	4.7	-0.4
24	7	1056	77.585464	12.37335	4.2	4.6	4.6	-0.3
15	2	397	45.894287	21.35475	4.1	3.8	3.7	0.3
21	7	1360	77.585464	10.677375	4.0	4.4	4.4	-0.4
14	3	351	45.894287	23.050724	4.0	4.1	4.1	-0.1
25	7	1362	77.585464	30.056063	4.0	3.9	3.9	0.1
40	4	413	110.91148	10.677375	3.8	3.7	3.6	0.2
16	3	499	45.894287	40.733437	3.7	3.5	3.5	0.2
43	2	270	109.27663	0	3.7	3.5	3.3	0.4
41	5	366	110.91148	12.37335	3.6	3.9	4.0	-0.4
17	2	478	45.894287	21.35475	3.5	3.7	3.8	-0.3
28	3	1129	58.534134	40.733437	3.5	3.0	2.7	0.8
45	7	756	110.91148	30.056063	3.4	3.8	3.9	-0.5
18	3	595	45.894287	40.733437	3.3	3.5	3.5	-0.3
46	5	516	110.91148	30.056063	3.1	3.4	3.5	-0.4

[a] Structures shown in Figures 1-3

[b] Predicted values obtained from Leave-One-Out Cross Validation

Table S5: Molecular Descriptor QSAR output for MCF-7 model

```
QuaSAR-Model(PLS): /Volumes/hudock/MOE/sulfonium/June2007/qsar_1_062107.mdb
Sat Jul 21 11:51:11 2007

Activity Field               : MCF7_pIC50
Weight Field                 :
Condition Limit              : 1e+06
Component Limit              : 0

Observations                 : 25
Descriptors                  : 4
Components Used              : 4
Condition Number             : 597321.55

ROOT MEAN SQUARE ERROR (RMSE): 0.41459
CORRELATION COEFFICIENT (R2) : 0.82124
CROSS-VALIDATED RMSE         : 0.51079
CROSS-VALIDATED R2           : 0.73327

ESTIMATED LINEAR MODEL

MCF7_pIC50 =
         4.67720
        +0.28611 * b_Cont_Rot
        -0.01473 * SMR_VSA7
        -0.02422 * PEOE_VSA+2
        -0.00065 * weinerPath

ESTIMATED NORMALIZED LINEAR MODEL (SD = Standard Deviation)

MCF7_pIC50 / SD(MCF7_pIC50) =
         4.76983
        +1.66491 * b_Cont_Rot / SD(b_Cont_Rot)
        -0.39320 * SMR_VSA7 / SD(SMR_VSA7)
        -0.26986 * PEOE_VSA+2 / SD(PEOE_VSA+2)
        -0.72485 * weinerPath / SD(weinerPath)

RELATIVE IMPORTANCE OF DESCRIPTORS

1.000000  b_Cont_Rot
0.236166  SMR_VSA7
0.162089  PEOE_VSA+2
0.435367  weinerPath
```

b_cont_Rot = Number of contiguous rotatable bonds.
SMR_VSA7 = Molar refractivity using the sub-divided molecular surface area method.
PEOE_VSA+2 = Partial Gasteiger charge using the sub-divided molecular surface area method.
weinerPath = Describes the compactness, branching and substitution of the molecule.

Table S6: Molecular Descriptor QSAR Results for SF-268 cell line.

Cpd[a]	Contiguous Rotatable Bonds	Weiner Path	Molar Refractivity (SMR VSA7)	Gasteiger Charge (VSA+2)	SF-268 pIC_{50} (M)	Predicted pIC_{50} (M)	CV Predicted[b] pIC_{50} (M)	Residual
30	14	1818	110.91148	12.37335	6.5	5.8	5.7	0.9
31	12	1353	110.91148	12.37335	6.4	5.6	5.4	1.1
32	16	2379	110.91148	12.37335	6.1	6	6	0.1
37	18	3044	110.91148	12.37335	6	6.1	6.1	-0.1
35	22	4718	110.91148	12.37335	5.8	6	6.1	-0.3
36	20	3821	110.91148	12.37335	5.8	6.1	6.2	-0.4
20	6	1154	77.585464	10.677375	5	4.3	4.2	0.7
38	10	976	110.91148	12.37335	4.8	5.2	5.3	-0.5
22	9	1464	77.585464	12.37335	4.7	5	5	-0.3
39	10	1261	110.91148	30.056063	4.7	4.6	4.6	0.1
21	7	1360	77.585464	10.677375	4.5	4.5	4.5	0
23	8	1249	77.585464	12.37335	4.4	4.9	4.9	-0.5
24	7	1056	77.585464	12.37335	4.3	4.7	4.7	-0.4
25	7	1362	77.585464	30.056063	4.2	4	4	0.2
14	3	351	45.894287	23.050724	4.1	4.2	4.3	-0.2
15	2	397	45.894287	21.35475	4	3.9	3.9	0.1
40	4	413	110.91148	10.677375	4	3.8	3.7	0.3
16	3	499	45.894287	40.733437	3.9	3.7	3.6	0.3
43	2	270	109.27663	0	3.8	3.6	3.4	0.4
41	5	366	110.91148	12.37335	3.7	4.1	4.2	-0.6
45	7	756	110.91148	30.056063	3.7	4	4.1	-0.4
17	2	478	45.894287	21.35475	3.6	3.8	3.9	-0.3
28	3	1129	58.534134	40.733437	3.6	3	2.7	0.9
18	3	595	45.894287	40.733437	3.4	3.6	3.7	-0.3
46	5	516	110.91148	30.056063	3.1	3.6	3.7	-0.6

[a] Structures shown in Figures 1-3

[b] Predicted values obtained from Leave-One-Out Cross Validation

Table S7: Molecular Descriptor QSAR output for SF-268 model

```
QuaSAR-Model(PLS): /Volumes/hudock/MOE/sulfonium/June2007/qsar_1_062107.mdb
Sat Jul 21 11:52:20 2007

Activity Field               : SF268_pIC50
Weight Field                 :
Condition Limit              : 1e+06
Component Limit              : 0

Observations                 : 25
Descriptors                  : 4
Components Used              : 4
Condition Number             : 597321.55

ROOT MEAN SQUARE ERROR (RMSE): 0.38182
CORRELATION COEFFICIENT (R2) : 0.84820
CROSS-VALIDATED RMSE         : 0.47782
CROSS-VALIDATED R2           : 0.76793

ESTIMATED LINEAR MODEL

SF268_pIC50 =
         4.76774
        +0.32257 * b_Cont_Rot
        -0.01496 * SMR_VSA7
        -0.02305 * PEOE_VSA+2
        -0.00084 * weinerPath

ESTIMATED NORMALIZED LINEAR MODEL (SD = Standard Deviation)

SF268_pIC50 / SD(SF268_pIC50) =
         4.86504
        +1.87817 * b_Cont_Rot / SD(b_Cont_Rot)
        -0.39958 * SMR_VSA7 / SD(SMR_VSA7)
        -0.25704 * PEOE_VSA+2 / SD(PEOE_VSA+2)
        -0.93930 * weinerPath / SD(weinerPath)

RELATIVE IMPORTANCE OF DESCRIPTORS

1.000000  b_Cont_Rot
0.212750  SMR_VSA7
0.136858  PEOE_VSA+2
0.500113  weinerPath
```

b_cont_Rot = Number of contiguous rotatable bonds.
SMR_VSA7 = Molar refractivity using the sub-divided molecular surface area method.
PEOE_VSA+2 = Partial Gasteiger charge using the sub-divided molecular surface area method.
weinerPath = Describes the compactness, branching and substitution of the molecule.

Table S8: Hologram (HQSAR) results and output for MCF-7, NCI-H460 and SF-268 cell lines. Average experiment vs. predicted $r^2 = 0.76$, $q^2 = 0.6$.

Cpd	NCI-H460 pIC_{50} (M)	Predicted pIC_{50} (M)	Residual	MCF-7 pIC_{50} (M)	Predicted pIC_{50} (M)	Residual	SF-268 pIC_{50} (M)	Predicted pIC_{50} (M)	Residual
30	6.7	5.5	1.2	6.6	5.4	1.2	6.5	5.5	1.1
31	6.5	5.2	1.3	6.2	5.1	1.1	6.4	5.2	1.2
32	6.3	5.7	0.6	5.0	5.7	-0.8	6.1	5.7	0.3
35	5.9	6.6	-0.6	6.0	6.6	-0.6	5.8	6.5	-0.7
36	5.9	6.3	-0.4	5.9	6.3	-0.5	5.8	6.2	-0.5
37	5.9	6.0	-0.2	6.0	6.0	-0.1	6.0	6.0	0.0
38	4.9	4.9	0.0	4.7	4.8	-0.2	4.8	5.0	-0.1
20	4.9	4.3	0.5	4.8	4.4	0.3	5.0	4.2	0.8
21	4.8	4.3	0.5	4.0	4.3	-0.3	4.5	4.2	0.2
22	4.6	4.7	0.0	4.7	4.6	0.1	4.8	4.6	0.2
39	4.6	4.2	0.4	4.5	4.2	0.3	4.7	4.2	0.5
23	4.3	4.6	-0.3	4.3	4.5	-0.2	4.4	4.5	-0.1
24	4.3	4.6	-0.3	4.2	4.5	-0.3	4.3	4.5	-0.2
25	4.2	3.9	0.3	4.0	3.9	0.1	4.2	3.8	0.4
40	4.1	4.1	0.0	3.8	4.2	-0.4	4.0	4.2	-0.2
14	4.1	4.3	-0.2	4.0	4.2	-0.2	4.1	4.4	-0.3
41	4.0	4.4	-0.4	3.6	4.3	-0.7	3.7	4.5	-0.8
15	3.9	4.0	-0.1	4.0	4.1	-0.1	4.0	4.1	-0.1
16	3.8	3.6	0.2	3.7	3.6	0.1	3.9	3.7	0.2
43	3.7	4.2	-0.5	3.7	4.3	-0.6	3.8	4.3	-0.5
45	3.7	3.8	-0.1	3.4	3.8	-0.3	3.7	3.9	-0.2
17	3.7	4.1	-0.4	3.5	4.1	-0.6	3.6	4.2	-0.6
28	3.6	3.3	0.3	3.5	3.6	-0.2	3.6	3.2	0.3
18	3.5	3.7	-0.2	3.3	3.6	-0.3	3.4	3.8	-0.4
46	3.1	3.7	-0.6	3.1	3.6	-0.5	3.2	3.7	-0.6
	q2	0.6		q2	0.6		q2	0.57	
	r2	0.76		r2	0.76		r2	0.75	
	Length	151		Length	353		Length	353	

	N	3	N	3	N	3

Table S9: CoMSIA output for NCI-H460 model (full table shown in the main text).

Maximum common substructure alignment:

```
Regression Equation(s)
Use COMFA FIELD RETRIEVE/LIST/GRAPH or EVA RETRIEVE/LIST/GRAPH
CoMFA/EVA coefficients.
All columns are COMFA.

Relative Contributions
 #                        Norm.Coeff. Fraction
 -                        ----------- ---------
 1 COMSIA_ST (1368 vars)       0.538     0.234
 2 COMSIA_EL (1368 vars)       0.087     0.038
 3 COMSIA_HY (1368 vars)       0.872     0.380
 4 COMSIA_DO (1368 vars)       0.436     0.190
 5 COMSIA_AC (1368 vars)       0.364     0.159

Input Selections: PLS Analysis
  Minimum Sigma to use Column:  0.0000   Missing Values: COLUMN_MEAN_DEFAULT
  Row Weighting: SAME_WEIGHTS_FOR_ALL
  Neither BOOTSTRAPPED nor CROSSVALIDATED
  COMPONENTS: 4
  Scaling: COMFA_STANDARD    Intercept forced through 0.0: NO
  NIPALS: (Max Iter = 100; EPS = 0.000100)
  Target (Y) Variable(s):
   2: NCI_pIC50
  Explanatory (X) Variable(s): (Actual terms: 6840 requested, 6372 used)
   4: COMSIA_ST             5: COMSIA_EL             6: COMSIA_HY
   7: COMSIA_DO             8: COMSIA_AC
  Rows in Analysis: (25 rows)
   1: 546                   2: 569                   3: 572
   4: 575                   5: 576                   6: 584
   7: 585                   8: 589                   9: 594
  10: 682                  11: 683                  12: 685
  13: 686                  14: 687                  15: 688
  16: 689                  17: 690                  18: 694
  19: 695                  20: 696                  21: 697
  22: 527                  23: 580                  24: 547
  25: 574

Summary output
Standard Error of Estimate           0.273
R squared                          0.941
F values     ( n1= 4, n2=20 )     80.178
Prob.of R2=0 ( n1= 4, n2=20 )      0.000
```

Feature-based alignment:

```
Regression Equation(s)
Use COMFA FIELD RETRIEVE/LIST/GRAPH or EVA RETRIEVE/LIST/GRAPH
CoMFA/EVA coefficients.
All columns are COMFA.

Relative Contributions
 #                        Norm.Coeff. Fraction
 -                        ----------- ---------
 1 COMSIA_ST (1872 vars)       0.699     0.330
 2 COMSIA_EL (1872 vars)       0.081     0.038
 3 COMSIA_HY (1872 vars)       1.100     0.519
 4 COMSIA_DO (1872 vars)       0.183     0.086
 5 COMSIA_AC (1872 vars)       0.056     0.026

Input Selections: PLS Analysis
  Minimum Sigma to use Column:  0.0000   Missing Values: COLUMN_MEAN_DEFAULT
```

```
  Row Weighting: SAME_WEIGHTS_FOR_ALL
  Neither BOOTSTRAPPED nor CROSSVALIDATED
  COMPONENTS: 4
  Scaling: COMFA_STANDARD    Intercept forced through 0.0: NO
  NIPALS: (Max Iter = 100; EPS = 0.000100)
  Target (Y) Variable(s):
   2: NCI_pIC50
  Explanatory (X) Variable(s): (Actual terms: 9360 requested, 8830 used)
   4: COMSIA_ST            5: COMSIA_EL            6: COMSIA_HY
   7: COMSIA_DO            8: COMSIA_AC
  Rows in Analysis: (25 rows)
   1: 527                  2: 546                  3: 569
   4: 572                  5: 575                  6: 576
   7: 580                  8: 584                  9: 585
  10: 589                 11: 594                 12: 682
  13: 683                 14: 685                 15: 686
  16: 687                 17: 688                 18: 689
  19: 690                 20: 694                 21: 695
  22: 696                 23: 697                 24: 547
  25: 574

Summary output
Standard Error of Estimate          0.216
R squared                         0.963
F values     ( n1= 4, n2=20 )   131.391
Prob.of R2=0 ( n1= 4, n2=20 )     0.000
```

Table S10: CoMSIA QSAR results for MCF-7 cell line.

	MCF-7 Experimental Activity		CoMSIA pIC_{50} Predictions							
				Test Sets						
Cpd	IC_{50} (μM)	pIC_{50} (M)	Training Set	1	2	3	4	5	Pred	Residual
30	0.24	6.6	6.3	6.5	6.3	6.0	**5.7**	6.3	5.7	-0.54
31	0.62	6.2	6.0	6.1	6.1	**5.6**	5.8	6.0	5.6	-0.42
35	0.95	6.0	5.7	5.8	5.7	5.8	5.9	5.6	5.7	0.03
37	1.12	6.0	5.7	6.1	5.7	5.8	5.7	**5.6**	5.6	-0.11
36	1.41	5.9	5.7	5.9	5.7	5.8	5.8	5.6	5.7	0.04
32	10.96	5.0	5.9	**6.4**	5.9	5.8	5.5	5.8	6.4	0.54
20	17.78	4.8	4.7	4.7	4.7	4.9	4.8	4.8	4.8	0.03
38	21.38	4.7	5.1	5.1	5.1	4.9	5.1	5.1	5.1	-0.02
22	21.38	4.7	4.8	4.8	4.8	4.8	**5.0**	4.8	5.0	0.18
39	32.36	4.5	4.6	4.5	**4.3**	4.6	4.7	4.6	4.3	-0.32
23	45.71	4.3	4.2	4.2	4.2	4.3	4.3	4.2	4.3	0.05
24	57.54	4.2	4.3	4.3	4.3	4.5	4.4	4.3	4.4	0.03
21	93.33	4.0	4.1	4.1	4.1	**4.6**	4.1	4.1	4.6	0.45
14	95.50	4.0	3.7	3.8	3.7	3.7	3.8	3.7	3.7	0.02
15	100.00	4.0	3.7	3.7	3.7	3.8	3.7	3.7	3.7	0.02
25	104.71	4.0	3.9	3.9	4.0	3.9	3.9	3.9	3.9	0.01
40	162.18	3.8	3.6	3.6	**3.6**	3.7	3.6	3.6	3.6	-0.03
16	194.98	3.7	3.5	3.5	3.4	3.4	3.4	3.5	3.4	-0.04
43	199.53	3.7	3.8	3.8	3.7	3.7	3.8	3.7	3.7	-0.02
41	239.88	3.6	3.7	**3.8**	3.7	3.6	3.7	3.7	3.8	0.18
17	302.00	3.5	3.6	3.6	3.6	3.8	3.6	3.6	3.6	0.05
28	338.84	3.5	3.5	3.5	3.6	3.5	3.5	3.5	3.5	-0.01
45	389.05	3.4	3.5	3.5	3.4	3.4	3.4	**3.6**	3.6	0.10
18	512.86	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	-0.02
46	812.83	3.1	3.4	3.5	3.4	3.3	3.4	3.5	3.4	-0.03
		q^2	0.73	0.75	0.73	0.75	0.76	0.7		
		R^2	0.93	0.83	0.9	0.841	0.845	0.81		
		F test	61.789	5	60	53	54.3	41.7		

N	4	4	4	4	4	4
n	25	23	23	23	23	23

Table S11: CoMSIA output for MCF-7 model.

Maximum common substructure alignment:

```
Regression Equation(s)
Use COMFA FIELD RETRIEVE/LIST/GRAPH or EVA RETRIEVE/LIST/GRAPH
CoMFA/EVA coefficients.
All columns are COMFA.

Relative Contributions
 #                       Norm.Coeff. Fraction
 -                       ----------- ---------
 1 COMSIA_ST (1368 vars)       0.513     0.205
 2 COMSIA_EL (1368 vars)       0.099     0.040
 3 COMSIA_HY (1368 vars)       0.855     0.342
 4 COMSIA_DO (1368 vars)       0.647     0.259
 5 COMSIA_AC (1368 vars)       0.384     0.154

Input Selections: PLS Analysis
  Minimum Sigma to use Column:  0.0000   Missing Values: COLUMN_MEAN_DEFAULT
  Row Weighting: SAME_WEIGHTS_FOR_ALL
  Neither BOOTSTRAPPED nor CROSSVALIDATED
  COMPONENTS: 4
  Scaling: COMFA_STANDARD    Intercept forced through 0.0: NO
  NIPALS: (Max Iter = 100; EPS = 0.000100)
  Target (Y) Variable(s):
   1: MCF7_pIC50
  Explanatory (X) Variable(s): (Actual terms: 6840 requested, 6372 used)
   4: COMSIA_ST             5: COMSIA_EL             6: COMSIA_HY
   7: COMSIA_DO             8: COMSIA_AC
  Rows in Analysis: (25 rows)
   1: 546                   2: 569                   3: 572
   4: 575                   5: 576                   6: 584
   7: 585                   8: 589                   9: 594
  10: 682                  11: 683                  12: 685
  13: 686                  14: 687                  15: 688
  16: 689                  17: 690                  18: 694
  19: 695                  20: 696                  21: 697
  22: 527                  23: 580                  24: 547
  25: 574

Summary output
Standard Error of Estimate          0.326
R squared                          0.912
F values     ( n1= 4, n2=20 )     51.682
Prob.of R2=0 ( n1= 4, n2=20 )      0.000
```

Feature-based alignment:

```
Regression Equation(s)
Use COMFA FIELD RETRIEVE/LIST/GRAPH or EVA RETRIEVE/LIST/GRAPH
CoMFA/EVA coefficients.
All columns are COMFA.

Relative Contributions
 #                       Norm.Coeff. Fraction
 -                       ----------- ---------
 1 COMSIA_ST (1872 vars)       0.652     0.306
 2 COMSIA_EL (1872 vars)       0.072     0.034
 3 COMSIA_HY (1872 vars)       1.029     0.484
 4 COMSIA_DO (1872 vars)       0.313     0.147
 5 COMSIA_AC (1872 vars)       0.062     0.029

Input Selections: PLS Analysis
```

```
  Minimum Sigma to use Column:  0.0000   Missing Values: COLUMN_MEAN_DEFAULT
  Row Weighting: SAME_WEIGHTS_FOR_ALL
  Neither BOOTSTRAPPED nor CROSSVALIDATED
  COMPONENTS: 4
  Scaling: COMFA_STANDARD    Intercept forced through 0.0: NO
  NIPALS: (Max Iter = 100; EPS = 0.000100)
  Target (Y) Variable(s):
   1: MCF7_pIC50
  Explanatory (X) Variable(s): (Actual terms: 9360 requested, 8830 used)
   4: COMSIA_ST             5: COMSIA_EL             6: COMSIA_HY
   7: COMSIA_DO             8: COMSIA_AC
  Rows in Analysis: (25 rows)
   1: 527                   2: 546                   3: 569
   4: 572                   5: 575                   6: 576
   7: 580                   8: 584                   9: 585
  10: 589                  11: 594                  12: 682
  13: 683                  14: 685                  15: 686
  16: 687                  17: 688                  18: 689
  19: 690                  20: 694                  21: 695
  22: 696                  23: 697                  24: 547
  25: 574

Summary output
Standard Error of Estimate          0.300
R squared                          0.925
F values     ( n1= 4, n2=20 )     61.789
Prob.of R2=0 ( n1= 4, n2=20 )      0.000
```

Table S12: CoMSIA QSAR results for SF-268 cell line.

	SF-268 Experimental Activity		CoMSIA pIC_{50} Predictions							
				Test Sets						
Cpd	IC_{50} (μM)	pIC_{50} (M)	Training Set	1	2	3	4	5	Pred	Residual
30	0.32	6.5	6.5	6.6	6.5	6.2	**6.2**	6.5	6.2	-0.27
31	0.32	6.4	6.1	6.2	6.1	**5.7**	6.0	6.1	5.7	-0.39
32	0.32	6.1	6.3	6.4	6.4	6.2	6.2	6.4	6.3	-0.02
37	0.32	6.0	6.0	6.1	6.0	6.0	6.0	**6.0**	6.0	-0.02
35	0.32	5.8	5.6	5.6	5.6	5.6	5.6	5.6	5.6	0.02
36	0.32	5.8	5.7	**5.8**	5.7	5.8	5.8	5.7	5.8	0.07
20	0.32	5.0	4.9	5.0	4.9	5.0	5.0	4.9	5.0	0.02
38	0.32	4.8	5.2	5.2	5.1	5.0	5.2	5.2	5.1	-0.01
22	0.32	4.8	4.9	4.9	4.8	4.9	**5.1**	4.9	5.1	0.19
39	0.32	4.7	4.8	4.8	**4.5**	4.7	4.9	4.7	4.5	-0.29
21	0.32	4.5	4.5	4.5	4.4	4.6	4.5	4.5	4.5	0.03
23	0.32	4.4	4.3	4.3	4.4	4.4	4.4	4.3	4.4	0.02
24	0.32	4.3	4.4	4.4	4.4	**4.5**	4.5	4.4	4.5	0.09
25	0.32	4.2	4.1	4.1	4.1	4.1	4.1	4.1	4.1	0.01
14	0.32	4.1	3.7	3.8	3.7	3.7	3.8	3.8	3.8	0.01
40	0.32	4.0	3.8	3.8	3.7	3.9	3.8	3.8	3.8	0.01
15	0.32	4.0	3.8	3.8	**3.8**	3.8	3.7	3.8	3.8	0.05
16	0.32	3.9	3.6	3.6	3.6	3.5	3.6	3.6	3.6	-0.03
43	0.32	3.8	3.8	3.9	3.8	3.8	3.8	3.8	3.8	-0.02
45	0.32	3.7	3.7	**3.7**	3.7	3.7	3.7	3.7	3.7	0.00
41	0.32	3.7	3.7	3.8	3.7	3.7	3.7	3.7	3.7	0.03
17	0.32	3.6	3.7	3.7	3.7	3.9	3.7	3.8	3.8	0.02
28	0.32	3.6	3.7	3.7	3.7	3.7	3.6	**3.7**	3.7	0.02
18	0.32	3.4	3.5	3.4	3.4	3.4	3.4	3.5	3.4	-0.01
46	0.32	3.2	3.6	3.6	3.5	3.5	3.6	3.6	3.6	-0.03
		q^2	0.85	0.78	0.844	0.82	0.82	0.82		
		R^2	0.96	0.95	0.96	0.96	0.96	0.96		

F test	118.6	84.4	119.7	104.7	110.2	97.7
N	4	4	4	4	4	4
n	25	23	23	23	23	23

Table S13: CoMSIA output for SF-268 model.

Maximum common substructure alignment:

```
Regression Equation(s)
Use COMFA FIELD RETRIEVE/LIST/GRAPH or EVA RETRIEVE/LIST/GRAPH
CoMFA/EVA coefficients.
All columns are COMFA.

Relative Contributions
 #                       Norm.Coeff. Fraction
 -                       ----------- ---------
 1 COMSIA_ST (1368 vars)       0.557     0.241
 2 COMSIA_EL (1368 vars)       0.095     0.041
 3 COMSIA_HY (1368 vars)       0.882     0.382
 4 COMSIA_DO (1368 vars)       0.429     0.186
 5 COMSIA_AC (1368 vars)       0.345     0.150

Input Selections: PLS Analysis
  Minimum Sigma to use Column:  0.0000   Missing Values: COLUMN_MEAN_DEFAULT
  Row Weighting: SAME_WEIGHTS_FOR_ALL
  Neither BOOTSTRAPPED nor CROSSVALIDATED
  COMPONENTS: 4
  Scaling: COMFA_STANDARD    Intercept forced through 0.0: NO
  NIPALS: (Max Iter = 100; EPS = 0.000100)
  Target (Y) Variable(s):
   3: SF268_pIC50
  Explanatory (X) Variable(s): (Actual terms: 6840 requested, 6372 used)
   4: COMSIA_ST             5: COMSIA_EL             6: COMSIA_HY
   7: COMSIA_DO             8: COMSIA_AC
  Rows in Analysis: (25 rows)
   1: 546                   2: 569                   3: 572
   4: 575                   5: 576                   6: 584
   7: 585                   8: 589                   9: 594
  10: 682                  11: 683                  12: 685
  13: 686                  14: 687                  15: 688
  16: 689                  17: 690                  18: 694
  19: 695                  20: 696                  21: 697
  22: 527                  23: 580                  24: 547
  25: 574

Summary output
Standard Error of Estimate          0.267
R squared                          0.940
F values     ( n1= 4, n2=20 )     79.023
Prob.of R2=0 ( n1= 4, n2=20 )      0.000
```

Feature-based alignment:

```
Regression Equation(s)
Use COMFA FIELD RETRIEVE/LIST/GRAPH or EVA RETRIEVE/LIST/GRAPH
CoMFA/EVA coefficients.
All columns are COMFA.

Relative Contributions
 #                       Norm.Coeff. Fraction
 -                       ----------- ---------
 1 COMSIA_ST (1872 vars)       0.678     0.320
 2 COMSIA_EL (1872 vars)       0.077     0.037
 3 COMSIA_HY (1872 vars)       1.064     0.502
 4 COMSIA_DO (1872 vars)       0.240     0.113
 5 COMSIA_AC (1872 vars)       0.060     0.028

Input Selections: PLS Analysis
```

```
  Minimum Sigma to use Column:  0.0000   Missing Values: COLUMN_MEAN_DEFAULT
  Row Weighting: SAME_WEIGHTS_FOR_ALL
  Neither BOOTSTRAPPED nor CROSSVALIDATED
  COMPONENTS: 4
  Scaling: COMFA_STANDARD    Intercept forced through 0.0: NO
  NIPALS: (Max Iter = 100; EPS = 0.000100)
  Target (Y) Variable(s):
   3: SF268_pIC50
  Explanatory (X) Variable(s): (Actual terms: 9360 requested, 8830 used)
   4: COMSIA_ST             5: COMSIA_EL             6: COMSIA_HY
   7: COMSIA_DO             8: COMSIA_AC
  Rows in Analysis: (25 rows)
   1: 527                   2: 546                   3: 569
   4: 572                   5: 575                   6: 576
   7: 580                   8: 584                   9: 585
  10: 589                  11: 594                  12: 682
  13: 683                  14: 685                  15: 686
  16: 687                  17: 688                  18: 689
  19: 690                  20: 694                  21: 695
  22: 696                  23: 697                  24: 547
  25: 574

Summary output
Standard Error of Estimate          0.220
R squared                           0.960
F values     ( n1= 4, n2=20 )     118.549
Prob.of R2=0 ( n1= 4, n2=20 )       0.000
```

References:

1. Labute, P., A widely applicable set of descriptors. *J. Mol. Graph. Modelling* **2000,** 18, (4-5), 464-477.
2. Guba, W., Descriptors and Represenation of Chemicals. In *Predictive Toxicology*, C, H., Ed. Taylor & Francis: CRC London, 2005; pp 11-36.
3. Kavanagh, K. L.; Guo, K.; Dunford, J. E.; Wu, X.; Knapp, S.; Ebetino, F. H.; Rogers, M. J.; Russell, R. G.; Oppermann, U., The molecular mechanism of nitrogen-containing bisphosphonates as antiosteoporosis drugs. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** 103, (20), 7829-34.
4. Mao, J.; Gao, Y. G.; Odeh, S.; Robinson, H.; Montalvetti, A.; Docampo, R.; Oldfield, E., Crystallization and Preliminary X-ray Diffraction Study of the Farnesyl Diphosphate Synthase from *Trypanosoma brucei*. *Acta. Crystallogr. D Biol. Crystallogr.* **2004,** 60, (Pt 10), 1863-6.
5. Navaza, J., AMoRe: an automated package for molecular replacement. *Acta Crystallog. Sect. A* **1994,** 50, 157-163.
6. Sheldrick, G.; Schneider, T., SHELXL: High Resolution Refinement. *Methods in Enzymology* **1997,** 277, 319-343.
7. Jones, T. A.; Zou, J. Y.; Cowan, S. W.; Kjeldgaard, M., Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. Sect. A* **1991,** 47, 110-119.
8. Mao, J.; Mukherjee, S.; Zhang, Y.; Cao, R.; Sanders, J. M.; Song, Y.; Zhang, Y.; Meints, G. A.; Gao, Y. G.; Mukkamala, D.; Hudock, M. P.; Oldfield, E., Solid-state NMR, crystallographic, and computational investigation of bisphosphonates and farnesyl diphosphate synthase-bisphosphonate complexes. *J. Am. Chem. Soc.* **2006,** 128, (45), 14485-97.
9. Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **1998,** 54, (Pt 5), 905-21.
10. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004,** 60, (Pt 12 Pt 1), 2126-32.