Bisphosphonate Inhibition of a *Plasmodium* Farnesyl Diphosphate Synthase and a General Method for Predicting Cell-Based Activity from Enzyme Data

Dushyant Mukkamala[#,1], Joo Hwan No[#,1], Lauren M. Cass[2], Ting-Kai Chang[3], and Eric Oldfield[1,3,†]

Supporting Information

[1]Center for Biophysics & Computational Biology, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801, USA; [2]School of Molecular and Cellular Biology, University of Illinois at Urbana-Champaign, 505 South Goodwin Avenue, Urbana IL 61801; [3]Department of Chemistry, University of Illinois at Urbana-Champaign, 600 S. Mathews Avenue, Urbana, IL 61801, USA

Table of Contents

Table S1. List of descriptors used in the Combinatorial Descriptor Search[a]

Code	Descripton	Number of Occurences[b]
log(PAMPA)	Log of parallel artificial membrane permeability	10
log(Caco-2)	Log of permeability coefficient for transport into Caco-2 cells	1
log P-neutral	Log of the concentrations of un-ionized compound between water and octanol	
log D-pH 7.4	Log of the ratio of the sum of concentrations of solute's various forms between water and octanol	
clogP	Log P calculated by Sybyl 7.3[1]	
mlogP	Measured log P calculated by Sybyl 7.3[1]	
diameter	Largest value in the distance matrix	
petitjean	Value of (diameter - radius) / diameter.	
petitjeanSC	Petitjean graph Shape Coefficient[2]: (diameter - radius) / radius.	
VDistEq	If m is the sum of the distance matrix entries then VdistEq is defined to be the sum of $\log_2 m - p_i \log_2 p_i / m$ where p_i is the number of distance matrix entries equal to i.	
VDistMa	If m is the sum of the distance matrix entries then VDistMa is defined to be the sum of $\log_2 m - D_{ij} \log_2 D_{ij} / m$ over all i and j.	
weinerPath	Wiener path number: half the sum of all the distance matrix entries[3, 4]	
weinerPol	Wiener polarity number: half the sum of all the distance matrix entries with a value of 3[3]	
BCUT_PEOE_0 BCUT_PEOE_1 BCUT_PEOE_2 BCUT_PEOE_3	The BCUT descriptors[5] are calculated from the eigenvalues of a modified adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqrt(b_{ij}) where b_{ij} is the formal bond order between bonded atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported.	1
BCUT_SLOGP_0 BCUT_SLOGP_1 BCUT_SLOGP_2 BCUT_SLOGP_3	The BCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge[6].	1
BCUT_SMR_0 BCUT_SMR_1 BCUT_SMR_2 BCUT_SMR_3	The BCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge[6].	
GCUT_PEOE_0 GCUT_PEOE_1 GCUT_PEOE_2 GCUT_PEOE_3	The GCUT descriptors are calculated from the eigenvalues of a modified graph distance adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqr(d_{ij}) where d_{ij} is the (modified) graph distance between atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported.	

Code	Descripton	Number of Occurences[b]
GCUT_SLOGP_0 GCUT_SLOGP_1 GCUT_SLOGP_2 GCUT_SLOGP_3	The GCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge[6].	3
GCUT_SMR_0 GCUT_SMR_1 GCUT_SMR_2 GCUT_SMR_3	The GCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge[6].	1
a_count	Number of atoms (including implicit hydrogens). This is calculated as the sum of $(1 + h_i)$ over all non-trivial atoms i.	
a_IC	Atom information content (total). This is calculated to be a_ICM times n.	
a_ICM	Atom information content (mean). This is the entropy of the element distribution in the molecule (including implicit hydrogens but not lone pair pseudo-atoms). Let n_i be the number of occurrences of atomic number i in the molecule. Let $p_i = n_i / n$ where n is the sum of the n_i. The value of a_ICM is the negative of the sum over all i of $p_i \log p_i$.	
b_1rotN	Number of rotatable single bonds. Conjugated single bonds are not included (e.g., ester and peptide bonds).	
b_1rotR	Fraction of rotatable single bonds: b_1rotN divided by b_heavy.	1
b_count	Number of bonds (including implicit hydrogens). This is calculated as the sum of $(d_i/2 + h_i)$ over all non-trivial atoms i.	
b_rotN	Number of rotatable bonds. A bond is rotatable if it has order 1, is not in a ring, and has at least two heavy neighbors.	
b_rotR	Fraction of rotatable bonds: b_rotN divided by b_heavy.	1
b_single	Number of single bonds (including implicit hydrogens). Aromatic bonds are not considered to be single bonds.	
chi0v	Atomic valence connectivity index (order 0)[7, 8]. This is calculated as the sum of $1/\mathrm{sqrt}(v_i)$ over all heavy atoms i with $v_i > 0$.	
chi0v_C	Carbon valence connectivity index (order 0). This is calculated as the sum of $1/\mathrm{sqrt}(v_i)$ over all carbon atoms i with $v_i > 0$.	
chi1v	Atomic valence connectivity index (order 1)[7, 8]. This is calculated as the sum of $1/\mathrm{sqrt}(v_i v_j)$ over all bonds between heavy atoms i and j where $i < j$.	
chi1v_C	Carbon valence connectivity index (order 1). This is calculated as the sum of $1/\mathrm{sqrt}(v_i v_j)$ over all bonds between carbon atoms i and j where $i < j$.	

Code	Descripton	Number of Occurences[b]
Weight	Molecular weight (including implicit hydrogens) with atomic weights taken from *CRC Handbook of Chemistry and Physics*, CRC Press (1994).	
b_heavy	Number of bonds between heavy atoms.	
a_nC	Number of carbon atoms: $\#\{Z_i \mid Z_i = 6\}$.	
chi0	Atomic connectivity index (order 0)[7, 8]. This is calculated as the sum of 1/sqrt(d_i) over all heavy atoms i with $d_i > 0$.	
chi0_C	Carbon connectivity index (order 0). This is calculated as the sum of 1/sqrt(d_i) over all carbon atoms i with $d_i > 0$.	
chi1	Atomic connectivity index (order 1)[7, 8]. This is calculated as the sum of 1/sqrt($d_i d_j$) over all bonds between heavy atoms i and j where $i < j$.	
chi1_C	Carbon connectivity index (order 1). This is calculated as the sum of 1/sqrt($d_i d_j$) over all bonds between carbon atoms i and j where $i < j$.	1
VAdjEq	Vertex adjacency information (equality): $-(1-f)\log_2(1-f) - f \log_2 f$ where $f = (n^2 - m) / n^2$, n is the number of heavy atoms and m is the number of heavy-heavy bonds. If f is not in the open interval (0,1), then 0 is returned.	
VAdjMa	Vertex adjacency information (magnitude): $1 + \log_2 m$ where m is the number of heavy-heavy bonds. If m is zero, then zero is returned.	
zagreb	Zagreb index: the sum of d_i^2 over all heavy atoms i.	
balabanJ	Balaban's connectivity topological index[9].	
PC+ Q_PC+ PEOE_PC+	Total positive partial charge: the sum of the positive q_i. Q_PC+ is identical to PC+ which has been retained for compatibility.	2
PC- Q_PC- PEOE_PC-	Total negative partial charge: the sum of the negative q_i. Q_PC- is identical to PC- which has been retained for compatibility.	
RPC+ Q_RPC+ PEOE_RPC+	Relative positive partial charge: the largest positive q_i divided by the sum of the positive q_i. Q_RPC+ is identical to RPC+ which has been retained for compatibility.	1
RPC- Q_RPC- PEOE_RPC-	Relative negative partial charge: the smallest negative q_i divided by the sum of the negative q_i. Q_RPC- is identical to RPC- which has been retained for compatibility.	5
PEOE_VSA+0	Sum of v_i where q_i is in the range [0.00,0.05).	2
PEOE_VSA+1	Sum of v_i where q_i is in the range [0.05,0.10).	
PEOE_VSA-0	Sum of v_i where q_i is in the range [-0.05,0.00).	
PEOE_VSA-1	Sum of v_i where q_i is in the range [-0.10,-0.05).	1

Code	Descripton	Number of Occurences[b]
Q_VSA_FHYD PEOE_VSA_FHYD	Fractional hydrophobic van der Waals surface area. This is the sum of the v_i such that $\|q_i\|$ is less than or equal to 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.	3
Q_VSA_FNEG PEOE_VSA_FNEG	Fractional negative van der Waals surface area. This is the sum of the v_i such that q_i is negative divided by the total surface area. The v_i are calculated using a connection table approximation.	3
Q_VSA_FPNEG PEOE_VSA_FPNEG	Fractional negative polar van der Waals surface area. This is the sum of the v_i such that q_i is less than -0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.	1
Q_VSA_FPOL PEOE_VSA_FPOL	Fractional polar van der Waals surface area. This is the sum of the v_i such that $\|q_i\|$ is greater than 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.	4
Q_VSA_FPOS PEOE_VSA_FPOS	Fractional positive van der Waals surface area. This is the sum of the v_i such that q_i is non-negative divided by the total surface area. The v_i are calculated using a connection table approximation.	3
Q_VSA_FPPOS PEOE_VSA_FPPOS	Fractional positive polar van der Waals surface area. This is the sum of the v_i such that q_i is greater than 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.	
Q_VSA_HYD PEOE_VSA_HYD	Total hydrophobic van der Waals surface area. This is the sum of the v_i such that $\|q_i\|$ is less than or equal to 0.2. The v_i are calculated using a connection table approximation.	2
Q_VSA_NEG PEOE_VSA_NEG	Total negative van der Waals surface area. This is the sum of the v_i such that q_i is negative. The v_i are calculated using a connection table approximation.	1
Q_VSA_PNEG PEOE_VSA_PNEG	Total negative polar van der Waals surface area. This is the sum of the v_i such that q_i is less than -0.2. The v_i are calculated using a connection table approximation.	
Q_VSA_POL PEOE_VSA_POL	Total polar van der Waals surface area. This is the sum of the v_i such that $\|q_i\|$ is greater than 0.2. The v_i are calculated using a connection table approximation.	1
Q_VSA_POS PEOE_VSA_POS	Total positive van der Waals surface area. This is the sum of the v_i such that q_i is non-negative. The v_i are calculated using a connection table approximation.	
Q_VSA_PPOS	Total positive polar van der Waals surface area. This is the sum of the v_i such that q_i is greater than 0.2. The v_i are calculated using a connection table approximation.	1

Code	Descripton	Number of Occurences[b]
lip_acc	The number of O and N atoms.	
opr_nrot	The number of rotatable bonds[10].	
E	Value of the potential energy. The state of all term enable flags will be honored (in addition to the term weights). This means that the current potential setup accurately reflects what will be calculated.	1
E_ang	Angle bend potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	
E_ele	Electrostatic component of the potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	2
E_nb	Value of the potential energy with all bonded terms disabled. The state of the non-bonded term enable flags will be honored (in addition to the term weights).	3
E_oop	Out-of-plane potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	5
E_sol	Solvation energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	1
E_stb	Bond stretch-bend cross-term potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	
E_str	Bond stretch potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	1
E_strain	Local strain energy: the current energy minus the value of the energy at a near local minimum. The current energy is calculated as for the E descriptor. The local minimum energy is the value of the E descriptor after first performing an energy minimization. Current chirality is preserved and charges are left undisturbed during minimization. The structure in the database is not modified (results of the minimization are discarded).	1
E_tor	Torsion (proper and improper) potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	3
E_vdw	van der Waals component of the potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.	
Kier1	First kappa shape index: $(n\text{-}1)^2 / m^2$ [8].	
Kier2	Second kappa shape index: $(n\text{-}1)^2 / m^2$ [8].	1
Kier3	Third kappa shape index: $(n\text{-}1)\ (n\text{-}3)^2 / p_3^2$ for odd n, and $(n\text{-}3)\ (n\text{-}2)^2 / p_3^2$ for even n [8].	1

Code	Descripton	Number of Occurences[b]
KierA1	First alpha modified shape index: $s\,(s\text{-}1)^2 / m^2$ where $s = n + a$ [8].	2
KierA2	Second alpha modified shape index: $s\,(s\text{-}1)^2 / m^2$ where $s = n + a$ [8].	1
KierA3	Third alpha modified shape index: $(n\text{-}1)\,(n\text{-}3)^2 / p_3^2$ for odd n, and $(n\text{-}3)\,(n\text{-}2)^2 / p_3^2$ for even n where $s = n + a$ [8].	1
KierFlex	Kier molecular flexibility index: (KierA1) (KierA2) / n [8].	
logs	Log of the aqueous solubility This property is calculated from an atom contribution linear atom type model[11] with $r^2 = 0.90$, ~1,200 molecules.	7
apol	Sum of the atomic polarizabilities (including implicit hydrogens) with polarizabilities taken from *CRC Handbook of Chemistry and Physics*, CRC Press (1994).	1
bpol	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens) with polarizabilities taken from *CRC Handbook of Chemistry and Physics*, CRC Press (1994).	
mr	Molecular refractivity (including implicit hydrogens). This property is calculated from an 11 descriptor linear model with $r^2 = 0.997$, RMSE = 0.168 on 1,947 small molecules.	
dipole	Dipole moment calculated from the partial charges of the molecule.	
dipoleX	The x component of the dipole moment (external coordinates).	1
dipoleY	The y component of the dipole moment (external coordinates).	2
dipoleZ	The z component of the dipole moment (external coordinates).	
pmi	Principal moment of inertia.	5
pmiX	x component of the principal moment of inertia (external coordinates).	5
pmiY	y component of the principal moment of inertia (external coordinates).	3
pmiZ	z component of the principal moment of inertia (external coordinates).	3
rgyr	Radius of gyration.	
vsa_acc	Approximation to the sum of VDW surface areas of pure hydrogen bond acceptors (not counting acidic atoms and atoms that are both hydrogen bond donors and acceptors such as -OH).	
vsa_hyd	Approximation to the sum of VDW surface areas of hydrophobic atoms.	1

Code	Descripton	Number of Occurences[b]
vsa_pol	Approximation to the sum of VDW surface areas of polar atoms (atoms that are both hydrogen bond donors and acceptors), such as -OH.	
SlogP	Log of the octanol/water partition coefficient (including implicit hydrogens). This property is an atomic contribution model[6] that calculates logP from the given structure; i.e., the correct protonation state (washed structures). Results may vary from the logP(o/w) descriptor. The training set for SlogP was ~7000 structures.	11
SlogP_VSA0	Sum of v_i such that L_i <= -0.4.	
SlogP_VSA2	Sum of v_i such that L_i is in (-0.2,0].	
SlogP_VSA7	Sum of v_i such that L_i is in (0.25,0.30].	
SlogP_VSA8	Sum of v_i such that L_i is in (0.30,0.40].	
SlogP_VSA9	Sum of v_i such that L_i > 0.40.	
SMR	Molecular refractivity (including implicit hydrogens). This property is an atomic contribution model[6] that assumes the correct protonation state (washed structures). The model was trained on ~7000 structures and results may vary from the mr descriptor.	
SMR_VSA2	Sum of v_i such that R_i is in (0.26,0.35].	
SMR_VSA3	Sum of v_i such that R_i is in (0.35,0.39].	
SMR_VSA5	Sum of v_i such that R_i is in (0.44,0.485].	
SMR_VSA7	Sum of v_i such that R_i > 0.56.	
ASA	Water accessible surface area calculated using a radius of 1.4 A for the water molecule. A polyhedral representation is used for each atom in calculating the surface area.	1
ASA+	Water accessible surface area of all atoms with positive partial charge (strictly greater than 0).	
ASA-	Water accessible surface area of all atoms with negative partial charge (strictly less than 0).	
ASA_H	Water accessible surface area of all hydrophobic ($\|q_i\|$<0.2) atoms.	4
ASA_P	Water accessible surface area of all polar ($\|q_i\|$>=0.2) atoms.	1
CASA+	Positive charge weighted surface area, ASA+ times max { q_i > 0 }[12].	
CASA-	Negative charge weighted surface area, ASA- times max { q_i < 0 }[12].	
DASA	Absolute value of the difference between ASA+ and ASA-.	1
DCASA	Absolute value of the difference between CASA+ and CASA-[12].	2

Code	Descripton	Number of Occurences[b]
FASA+	Fractional ASA+ calculated as ASA+ / ASA.	
FASA-	Fractional ASA- calculated as ASA- / ASA.	
FASA_H	Fractional ASA_H calculated as ASA_H / ASA.	4
FASA_P	Fractional ASA_P calculated as ASA_P / ASA.	4
FCASA+	Fractional CASA+ calculated as CASA+ / ASA.	
FCASA-	Fractional CASA- calculated as CASA- / ASA.	2
VSA	van der Waals surface area. A polyhedral representation is used for each atom in calculating the surface area.	2
TPSA	Polar surface area calculated using group contributions to approximate the polar surface area from connection table information only. The parameterization is that of Ertl *et al.* [13].	
density	Molecular mass density: Weight divided by vdw_vol.	
vdw_area	Area of van der Waals surface calculated using a connection table approximation.	
vdw_vol	van der Waals volume calculated using a connection table approximation.	1
dens	Mass density: molecular weight divided by van der Waals volume as calculated in the vol descriptor.	
glob	Globularity, or inverse condition number (smallest eigenvalue divided by the largest eigenvalue) of the covariance matrix of atomic coordinates. A value of 1 indicates a perfect sphere while a value of 0 indicates a two- or one-dimensional object.	
std_dim1	Standard dimension 1: the square root of the largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.	
std_dim2	Standard dimension 2: the square root of the second largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.	1
std_dim3	Standard dimension 3: the square root of the third largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.	
vol	van der Waals volume calculated using a grid approximation (spacing 0.75 A).	
logP(o/w)	Log of the octanol/water partition coefficient (including implicit hydrogens). This property is calculated from a linear atom type model [LOGP 1998] with $r^2 = 0.931$, RMSE=0.393 on 1,827 molecules.	4

[a] Descriptors searched and their descriptions are generally taken from MOE[14]
[b] Times that the descriptor occurred in the top ranked descriptor combinations

Table S2. Top ten "enzyme (FPPS) plus 2-descriptor" combinations with their coefficients and relative contributions for the *D. discoideum* cell pIC_{50} predictions.

Dictyostelium discoideum (FPPS) pIC (cell) =	R^2	Relative Importance of pIC_{50}(enzyme)	Relative Importance of Descriptor B	Relative Importance of Descriptor C
-20.76307 +0.48341 * pIC_{50}Enzyme -14.18055 * GCUT_PEOE_0 +3.20333 * BCUT_PEOE_3	0.69692	1.000000	0.484251	0.729151
-12.32173 +0.57298 * pIC_{50}Enzyme +4.56066 * BCUT_SMR_3 +0.01669 * E_vdw	0.68450	1.000000	0.416094	0.388190
-6.15020 +0.50365 * pIC_{50}Enzyme +3.02716 * BCUT_PEOE_3 +0.000724775 * CASA+	0.67415	1.000000	0.661355	0.393880
-6.94591 +0.49467 * pIC_{50}Enzyme +3.32192 * BCUT_PEOE_3 +0.00390 * ASA+	0.67085	1.000000	0.738924	0.417312
-3.89017 +0.52832 * pIC_{50}Enzyme +2.05150 * BCUT_PEOE_3 +0.01401 * E_vdw	0.66951	1.000000	0.427270	0.353525
2.94011 +0.46041 * pIC_{50}Enzyme -0.00635 * ASA_P +0.63821 * PEOE_PC+	0.66825	1.000000	0.392190	0.638878
-8.79712 +0.50884 * pIC_{50}Enzyme +2.97502 * BCUT_PEOE_3 -0.62648 * PEOE_PC-	0.66574	1.000000	0.643330	0.362862
-17.05598 +0.56036 * pIC_{50}Enzyme +6.27128 * BCUT_SMR_3 +0.000757926 * CASA+	0.66511	1.000000	0.585055	0.370214
-8.75301 +0.48545 * pIC_{50}Enzyme +3.48930 * BCUT_SMR_3 +0.45159 * PEOE_PC+	0.66286	1.000000	0.375754	0.428755
-6.09568 +0.49120 * pIC_{50}Enzyme +3.03006 * BCUT_PEOE_3 +0.35326 * FCASA+	0.66042	1.000000	0.678777	0.361338

Table S3. Top ten "enzyme (FPPS) plus 2-descriptor" combinations with their coefficients and relative contributions for the *L. donovani* cell pIC_{50} predictions.

Leishmania donovani (FPPS) pIC (cell) =	R^2 [a]	Relative Importance of pIC_{50}(enzyme) [a]	Relative Importance of Descriptor B [a]	Relative Importance of Descriptor C [a]
12.82134 +0.74773 * pIC_{50}Enzyme -1.23653 * FCASA- -14.34579 * Q_VSA_FPOS	0.86952	0.269777	0.915872	1.000000
-1.52445 +0.74773 * pIC_{50}Enzyme -1.23653 * FCASA- +14.34579 * Q_VSA_FNEG	0.86952	0.269777	0.915872	1.000000
2.63995 +0.56184 * pIC_{50}Enzyme +0.10038 * SlogP -2.72841 * b_rotR	0.80829	1.000000	0.615015	0.987653
2.63995 +0.10038 * SlogP -2.72841 * b_1rotR +0.56184 * pIC_{50}Enzyme	0.80829	0.615015	0.987653	1.000000
2.43168 +0.67597 * pIC_{50}Enzyme +0.01062 * PEOE_VSA+0 -4.40941 * Q_VSA_FPOS	0.80745	1.000000	0.806196	0.793464
-1.97773 +0.67597 * pIC_{50}Enzyme +0.01062 * PEOE_VSA+0 +4.40940 * Q_VSA_FNEG	0.80745	0.793464	0.806196	1.000000
2.78569 +0.67570 * pIC_{50}Enzyme -3.01289 * Q_VSA_FPOS +0.14382 * SlogP	0.80578	1.000000	0.861479	0.732693
-0.22720 +0.67570 * pIC_{50}Enzyme +0.14382 * SlogP +3.01289 * Q_VSA_FNEG	0.80578	1.000000	0.732693	0.861479
5.23713 +0.35909 * pIC_{50}Enzyme +0.00796 * Q_VSA_HYD -0.19768 * KierA1	0.80375	0.319957	0.726844	1.000000
4.83355 +0.30356 * pIC_{50}Enzyme -0.22903 * KierA1 +0.00805 * VSA	0.80300	0.233458	1.000000	0.738323

[a] Identical values are obtained when descriptors are linearly dependent

Table S4. Top ten "enzyme (UPPS) plus 2-descriptor" combinations with their coefficients and relative contributions for the *S. pneumoniae* cell pIC_{50} predictions.

Streptococcus pneumoniae (UPPS) pIC (cell) =	R^2	Relative Importance of pIC_{50}Enzyme	Relative Importance of Descriptor B	Relative Importance of Descriptor C
2.36039 +0.09521 * pIC_{50}Enzyme +5.57167 * E_oop +0.30641 * logS	0.70859	0.206740	0.770366	1.000000
2.73916 -0.02058 * pIC_{50}Enzyme -0.25615 * dipoleY -0.50252 * SlogP	0.65294	0.036773	0.664638	1.000000
4.02259 -0.10469 * pIC_{50}Enzyme +0.44931 * logS -0.00111 * pmiZ	0.63628	0.155024	1.000000	0.380149
1.93265 +0.02581 * pIC_{50}Enzyme +5.82244 * E_oop -0.38762 * SlogP	0.63029	0.059775	0.858685	1.000000
2.91917 +0.05651 * pIC_{50}Enzyme -0.16171 * dipoleY +0.33943 * logS	0.61395	0.110774	0.460388	1.000000
3.29294 +0.11608 * pIC_{50}Enzyme +8.55584 * E_oop -0.01199 * vsa_hyd	0.60705	0.213066	1.000000	0.682444
3.00970 -0.02670 * pIC_{50}Enzyme +0.30685 * logS -0.000145748 * pmiY	0.59586	0.057891	1.000000	0.472137
2.72405 +0.01829 * pIC_{50}Enzyme +0.01022 * E_nb +0.37898 * logS	0.59540	0.032109	0.334330	1.000000
0.97384 +0.21900 * pIC_{50}Enzyme -0.000223701 * pmiY +6.34230 * E_oop	0.59461	0.542273	0.827566	1.000000
2.46641 +0.03924 * pIC_{50}Enzyme +0.12363 * E_str +0.40087 * logS	0.59423	0.065125	0.332503	1.000000

Table S5. Top ten "enzyme (MurI) plus 2-descriptor" combinations with their coefficients and relative contributions for the *S. pneumoniae* cell pIC_{50} predictions.

Streptococcus pneumoniae (MurI) pIC (cell) =	R^2 [a]	Relative Importance of pIC_{50}(enzyme) [a]	Relative Importance of Descriptor B [a]	Relative Importance of Descriptor C [a]
10.52552 +0.67868 * pIC_{50}Enzyme -8.00526 * FASA_H -9.58941 * Q_VSA_FPOL	0.72033	0.760086	0.930312	1.000000
2.52026 +0.67868 * pIC_{50}Enzyme -9.58941 * Q_VSA_FPOL +8.00526 * FASA_P	0.72033	0.760086	1.000000	0.930312
0.93611 +0.67868 * pIC_{50}Enzyme -8.00526 * FASA_H +9.58941 * Q_VSA_FHYD	0.72033	0.760086	0.930312	1.000000
-7.06915 +0.67868 * pIC_{50}Enzyme +9.58941 * Q_VSA_FHYD +8.00526 * FASA_P	0.72033	0.760086	1.000000	0.930312
1.11064 +0.73485 * pIC_{50}Enzyme +0.00116 * pmiZ -0.00189 * DCASA	0.70598	1.000000	0.352718	0.412965
0.96166 +0.74218 * pIC_{50}Enzyme -0.01333 * ASA_H +0.02650 * Q_VSA_HYD	0.70104	0.683156	1.000000	0.978195
1.07105 +0.73955 * pIC_{50}Enzyme +0.00121 * pmiZ -0.00932 * DASA	0.69820	1.000000	0.367214	0.401603
5.68232 +0.68410 * pIC_{50}Enzyme -5.36646 * Q_VSA_FPOL -0.00565709 * ASA_H	0.69080	1.000000	0.730430	0.673985
0.31586 +0.68410 * pIC_{50}Enzyme -0.00565708 * ASA_H +5.36646 * Q_VSA_FHYD	0.69080	1.000000	0.673985	0.730430
-0.26123 +0.77505 * pIC_{50}Enzyme -0.00154943 * DCASA +10.54584 * PEOE_RPC+	0.68937	1.000000	0.261402	0.321590

[a] Identical values are obtained when descriptors are linearly dependent

Table S6. Top ten "enzyme (NS3 protease) plus 2-descriptor" combinations with their coefficients and relative contributions for the HCV replicon cell pIC_{50} predictions.

Hepatitis C Virus (NS3 protease) pIC (cell) =	R^2	Relative Importance of pIC_{50}(enzyme)	Relative Importance of Descriptor B	Relative Importance of Descriptor C
-1.29239 +0.91301 * pIC_Enzyme +0.58847 * log(PAMPA) -0.00017357 * pmiY	0.77744	0.730339	1.000000	0.305259
-2.44335 +0.92235 * pIC_{50}Enzyme +0.55294 * log(PAMPA) +0.25772 * log(Caco-2)	0.74101	0.785205	1.000000	0.261151
0.11561 +0.84966 * pIC_{50}Enzyme +0.55981 * log(PAMPA) -0.56438 * std_dim2	0.73967	0.714448	1.000000	0.231869
-2.11082 +0.91600 * pIC_{50}Enzyme +0.58198 * log(PAMPA) -0.16375 * dipoleX	0.73819	0.740897	1.000000	0.222535
-1.07166 +0.92801 * pIC_{50}Enzyme +0.62367 * log(PAMPA) -0.24639 * PEOE_PC+	0.73569	0.700433	1.000000	0.198231
-1.43622 +0.92293 * pIC_{50}Enzyme +0.65322 * log(PAMPA) -0.06097 * KierA2	0.73564	0.665088	1.000000	0.197280
-1.52542 +0.92611 * pIC_{50}Enzyme +0.64820 * log(PAMPA) -0.04420 * Kier2	0.73518	0.672545	1.000000	0.195686
-1.58484 +0.91365 * pIC_{50}Enzyme +0.65058 * log(PAMPA) -0.05945 * Kier3	0.73454	0.661071	1.000000	0.193573
-1.85996 +0.96397 * pIC_{50}Enzyme +0.61988 * log(PAMPA) -0.00374 * Q_VSA_PPOS	0.73415	0.732031	1.000000	0.203505
-1.61354 +0.95796 * pIC_{50}Enzyme +0.61521 * log(PAMPA) -0.00310 * Q_VSA_POL	0.73394	0.732983	1.000000	0.203058

Table S7. Top ten "enzyme (HIV-1 integrase) plus 2-descriptor" combinations with their coefficients and relative contributions for the HIV-1 integrase replicon cell pIC_{50} predictions.

MT4 cells (HIV-1 integrase) pIC (cell) =	R^2	Relative Importance of pIC_{50}(enzyme)	Relative Importance of Descriptor B	Relative Importance of Descriptor C
-4.83173 +0.98896 * pIC50_enzyme +1.36334 * pCC50 -0.00697 * Weight	0.68808	0.892564	1.000000	0.526629
-3.90054 +0.79998 * pIC50_enzyme +1.24225 * pCC50 -0.00990 * vsa_hyd	0.68552	0.792381	1.000000	0.520038
-5.92926 +0.89119 * pIC50_enzyme +1.37520 * pCC50 -0.000147 * pmi	0.68063	0.797385	1.000000	0.489095
-5.67093 +0.84883 * pIC50_enzyme +1.35713 * pCC50 -0.000166 * pmiX	0.67775	0.769595	1.000000	0.482305
-2.59105 +0.71970 * pIC50_enzyme +1.22060 * pCC50 -0.30594 * chi1v	0.67651	0.725509	1.000000	0.509833
-5.53671 +0.85017 * pIC50_enzyme +1.34149 * pCC50 -0.19817 * dipole	0.67241	0.779798	1.000000	0.477547
-5.45844 +0.84242 * pIC50_enzyme +1.33341 * pCC50 -0.19829 * dipoleX	0.67078	0.777374	1.000000	0.475734
-2.07656 +0.67785 * pIC50_enzyme +1.16006 * pCC50 -0.00792 * vol	0.66766	0.718981	1.000000	0.519278
-2.88642 +0.78860 * pIC50_enzyme +1.21043 * pCC50 -0.18150 * chi0v	0.66619	0.801639	1.000000	0.497274
-6.15351 +0.91486 * pIC50_enzyme +1.35583 * pCC50 +0.20861 * logS	0.66278	0.830251	1.000000	0.471726

Table S8. Top ten "enzyme (hlGP) plus 2-descriptor" combinations with their coefficients and relative contributions for the rat hepatocytes pIC_{50} predictions.

Rat Hepatocytes (hlGP) pIC (cell) =	R^2	Relative Importance of pIC_{50}(enzyme)	Relative Importance of Descriptor B	Relative Importance of Descriptor C
-2.48325 +0.74485 * pIC_{50}Enzyme +0.02112 * PEOE_VSA_PNEG +0.00797 * Q_VSA_HYD	0.71312	1.000000	0.671022	0.456010
-2.61987 +0.81722 * pIC_{50}Enzyme +0.01567 * PEOE_VSA_PNEG +0.38040 * E_str	0.70675	1.000000	0.453849	0.360041
-2.47800 +0.77133 * pIC_{50}Enzyme +0.02157 * PEOE_VSA_PNEG +0.00356 * ASA_H	0.70330	1.000000	0.661853	0.414954
-3.30393 +0.77214 * pIC_{50}Enzyme +0.02442 * PEOE_VSA_PNEG +3.06864 * FASA_H	0.70198	1.000000	0.748269	0.455019
-0.23529 +0.77214 * pIC_{50}Enzyme +0.02442 * PEOE_VSA_PNEG -3.06864 * FASA_P	0.70198	1.000000	0.748269	0.455019
-7.57118 +0.89039 * pIC_{50}Enzyme +0.01755 * PEOE_VSA_PNEG -2.55393 * BCUT_PEOE_0	0.70019	1.000000	0.466411	0.329457
-8.83556 +0.83245 * pIC_{50}Enzyme +0.01546 * PEOE_VSA_PNEG +2.90552 * GCUT_SMR_3	0.69768	1.000000	0.439538	0.327327
-7.31859 +2.61295 * GCUT_SLOGP_3 +0.79188 * pIC_{50}Enzyme +0.01647 * PEOE_VSA_PNEG	0.69753	1.000000	0.345396	0.492133
-7.31859 +0.79188 * pIC_{50}Enzyme +0.01647 * PEOE_VSA_PNEG +2.61295 * GCUT_SLOGP_3	0.69753	1.000000	0.492133	0.345396
-2.00842 +0.73115 * pIC_{50}Enzyme +0.01897 * PEOE_VSA_PNEG +0.00634 * vsa_hyd	0.69664	1.000000	0.614097	0.397957

Table S9. Top ten "enzyme (KDR kinase) plus 2-descriptor" combinations with their coefficients and relative contributions for the NIH3T3 cell pIC_{50} predictions.

NIH 3T3 (KDR, type III receptor tyrosine kinase) pIC (cell) =	R^2	Relative Importance of pIC_{50}(enzyme)	Relative Importance of Descriptor B	Relative Importance of Descriptor C
0.82478 +0.85331 * pIC_{50}Enzyme +6.56543 * PEOE_VSA_FPOL +0.05504 * E_tor	0.60006	1.000000	0.494191	0.718790
0.66511 +0.86124 * pIC_{50}Enzyme +0.05066 * E_tor +12.60955 * PEOE_VSA_FPNEG	0.59907	1.000000	0.655445	0.485462
9.32007 +0.79676 * pIC_{50}Enzyme -18.85532 * PEOE_RPC- -0.06827 * apol	0.59487	0.766096	0.760710	1.000000
10.76167 +0.82694 * pIC_{50}Enzyme -19.03382 * PEOE_RPC- -0.00815127 * ASA	0.59352	0.795265	0.768062	1.000000
10.23883 +0.80199 * pIC_{50}Enzyme -20.80145 * PEOE_RPC- -0.00895460 * vdw_vol	0.59328	0.730389	0.794900	1.000000
10.31493 +0.79311 * pIC_{50}Enzyme -20.41614 * PEOE_RPC- -0.01264080 * vol	0.59276	0.729847	0.788320	1.000000
10.31493 +0.79311 * pIC_{50}Enzyme -20.41614 * PEOE_RPC- -0.01264080 * vol	0.58789	0.708545	0.798883	1.000000
-0.21236 +0.91159 * pIC_{50}Enzyme +0.03452 * E -0.05235 * E_nb	0.58380	0.500955	0.743845	1.000000
1.32645 +0.89890 * pIC_{50}Enzyme -0.01124 * E_nb +0.04199 * E_tor	0.58330	1.000000	0.434552	0.520573
2.55668 +0.72281 * pIC_{50}Enzyme -0.75267 * KierA3 +1.19503 * PEOE_PC+	0.58114	0.783185	1.000000	0.775189

Table S10. Top ten "enzyme (Akt kinase) plus 2-descriptor" combinations with their coefficients and relative contributions for the MiaPaCa-2 cell pIC_{50} predictions.

MiaPaca-2 human pancreatic cancer cells (Akt Kinase) pIC (cell) =	R^2	Relative Importance of pIC_{50}Enzyme	Relative Importance of Descriptor B	Relative Importance of Descriptor C
4.33712 +0.45265 * pIC_{50}Enzyme -0.0000929650 * pmiX -4.23429 * GCUT_SLOGP_2	0.68326	1.000000	0.556839	0.424186
2.55169 +0.44882 * pIC_Enzyme -0.000099391 * pmi +0.00253 * ASA_H	0.68020	1.000000	0.711354	0.499224
4.44515 +0.39949 * pIC_{50}Enzyme -0.0000873893 * pmiX -0.02504 * E_ele	0.67971	1.000000	0.593093	0.458662
4.23090 +0.42619 * pIC_{50}Enzyme -0.0000879041 * pmiX -1.87949 * FASA_P	0.67557	1.000000	0.559220	0.423219
2.35141 +0.42619 * pIC_{50}Enzyme -0.0000879041 * pmiX +1.87949 * FASA_H	0.67557	1.000000	0.559220	0.423219
4.65687 +0.41086 * pIC_{50}Enzyme -0.0000778479 * pmi -4.34328 * GCUT_SLOGP_2	0.67513	1.000000	0.608637	0.479355
4.54500 +0.38640 * pIC_{50}Enzyme -1.96877 * FASA_P -0.0000739953 * pmi	0.67115	1.000000	0.488967	0.615143
2.57623 +0.38640 * pIC_{50}Enzyme -0.0000739953 * pmi +1.96877 * FASA_H	0.67115	1.000000	0.615143	0.488967
4.34247 +0.40352 * pIC_Enzyme -0.0025494 * ASA_P -0.0000831233 * pmiX	0.67007	1.000000	0.435308	0.558505
4.71925 +0.35970 * pIC_{50}Enzyme -0.0000719880 * pmi -0.02523 * E_ele	0.66794	1.000000	0.642875	0.513363

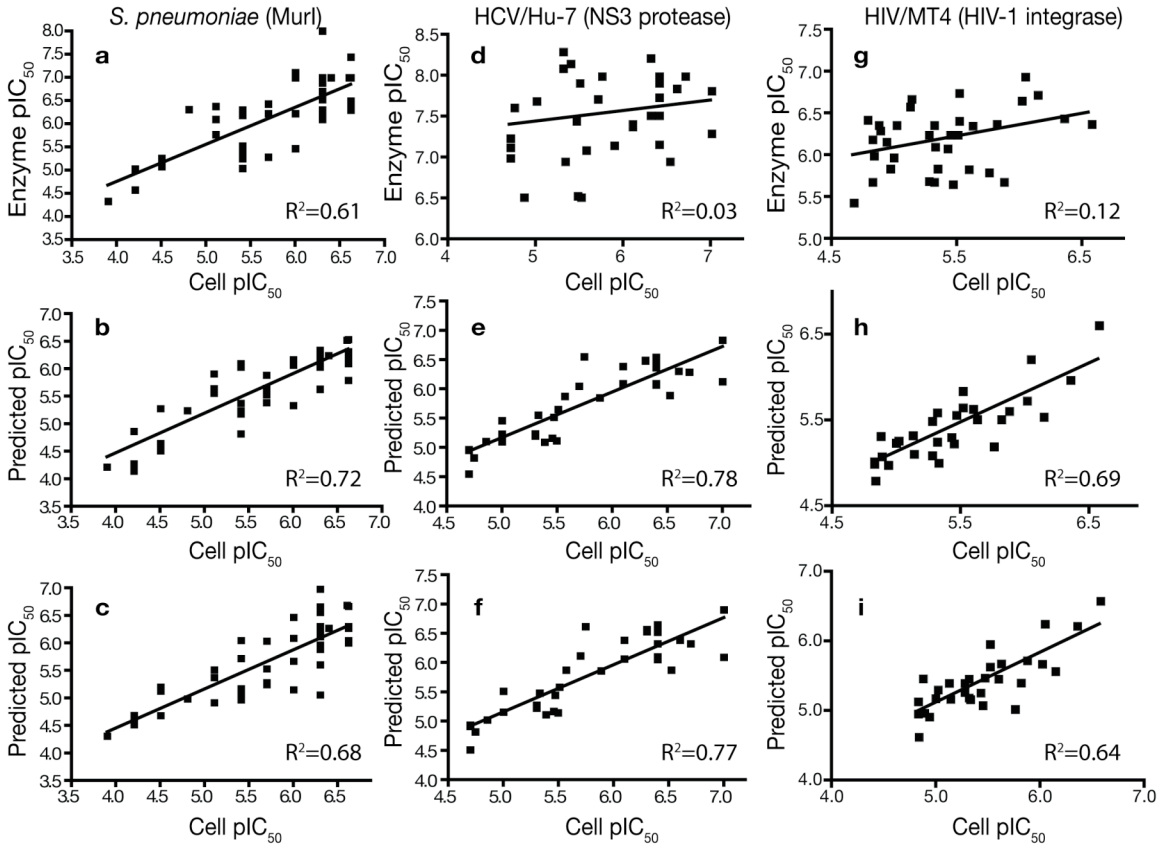

Figure S1. Correlation plots for the cell and enzyme assays, and predicted cell activities from the training and test set data in *Streptococcus pneumoniae* (MurI) (a-c), Hu-7 (HCV NS3 protease replication) (d-f) and MT4 (HIV-1 integrase) (g-i). (a) Plot showing the correlation between cell (*S. pneumoniae*) pIC_{50} and enzyme (MurI) pIC_{50} values. (b) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (c) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis. (d) Plot showing the correlation between cell (HCV/Hu-7) pIC_{50} and enzyme (HCV NS3 protease) pIC_{50} values. (e) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (f) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis. (g) Plot showing the correlation between cell (HIV/MT4) pIC_{50} and enzyme (HIV-1 integrase) pIC_{50} values. (h) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (i) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis.

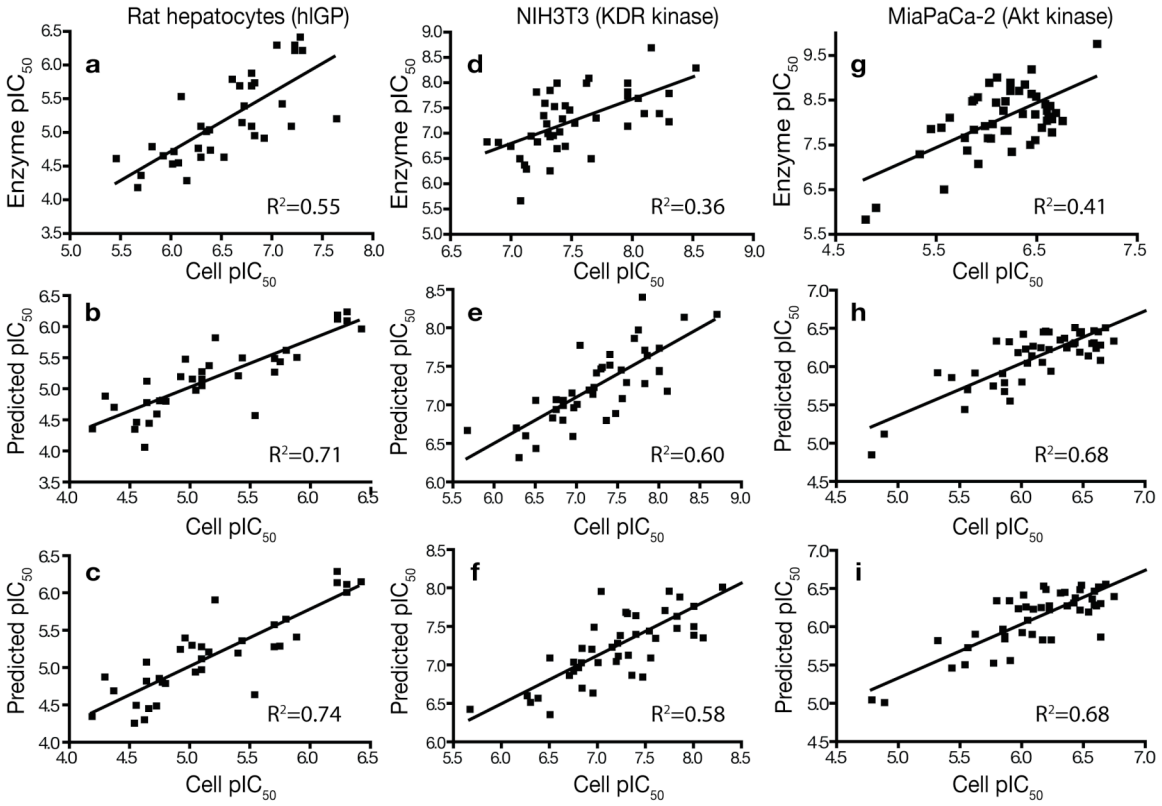

Figure S2. Correlation plots for the cell and enzyme assays, and predicted cell activities from the training and test set data in rat hepatocytes (hlGP) (a-c), NIH3T3 (KDR kinase) (d-f) and MiaPaCa-2 (Akt kinase) (g-i). (a) Plot showing the correlation between cell (Rat hepatocytes) pIC_{50} and enzyme (hlGP) pIC_{50} values. (b) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (c) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis. (d) Plot showing the correlation between cell (NIH 3T3) pIC_{50} and enzyme (KDR kinase) pIC_{50} values. (e) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (f) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis. (g) Plot showing the correlation between cell (MiaPaCa-2) pIC_{50} and enzyme (Akt kinase) pIC_{50} values. (h) Correlation between predicted cell pIC_{50} values (from the best combination of enzyme plus two molecular descriptors) for the training set, with the experimental pIC_{50}. (i) Correlation between test set pIC_{50} predictions obtained from a leave-two-out analysis.

References

1. *Sybyl 7.3*, Tripos, Inc.: St. Louis, MO.

2. Petitjean, M., Applications of the Radius Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical-Compounds. *J. Chem. Inf. Comput. Sci.* **1992,** 32, (4), 331-337.

3. Balaban, A. T., Five New Topological Indices for the Branching of Tree-Like Graphs. *Theoretica Chimica Acta.* **1979,** 53, 355-375.

4. Weiner, H., Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947,** 69, 17-20.

5. Pearlman, R. S.; Smith, K. M., Novel Software Tools for Chemical Diversity. *Persp. Drug. Disc. Desc. 9/10/11* **1998**, 339-353.

6. Wildman, S. A.; Crippen, G. M., Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999,** 39, (5), 868-873.

7. Hall, L. H.; Kier, L. B., The Nature of Structure-Activity Relationships and Their Relation to Molecular Connectivity. *Eur. J. Med. Chem.* **1977,** 12, 307.

8. Hall, L. H.; Kier, L. B., The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Modeling. *Rev. Comp. Chem.* **1991,** 2.

9. Balaban, A. T., Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982,** 89, (5), 399-404.

10. Oprea, T. I., Property Distribution of Drug-Related Chemical Databases. *J. Comp. Aid Mol. Des.* **2000,** 14, 251-264.

11. Hou, T. J.; Xia, K.; Zhang, W.; Xu, X. J., ADME Evaluation in Drug Discovery. 4. Prediction of Aqueous Solubility Based on Atom Contribution Approach. *J. Chem. Inf. Comput. Sci.* **2004,** 44, 266-275.

12. Stanton, D.; Jurs, P., Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Relationship Studies. *Anal. Chem.* **1990,** 62, 2323-2329.

13. Ertl, P.; Rohde, B.; Selzer, P., Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000,** 43, 3714-3717.

14. *Molecular Operating Environment (MOE)*, Chemical Computing Group, Inc.: Montreal, Quebec, 2006.