Inhibition of Geranylgeranyl Diphosphate Synthase by Bisphosphonates: A Crystallographic and Computational Investigation

Cammy K.-M. Chen[°†‡], Michael P. Hudock[°||], Yonghui Zhang[°¶],

Rey-Ting Guo[‡§], Rong Cao[||], Joo Hwan No[||], Po-Huang Liang[†‡],

Tzu-Ping Ko[‡], Tao-Hsin Chang[‡§], Shiou-chi Chang[¶], Yongcheng Song[¶],

Jordan Axelson[¶], Anup Kumar[¶], Andrew H.-J. Wang[†‡§*] and Eric Oldfield[¶||*]

Supporting Information

Table of Contents

Supporting Methods S2-S5
Supporting Figures S6-S11
Binding modes of GGPPS inhibitors S6
11 (BPH-23) (PDB File: 2z52) bound to GGPPS superimposed on FPP, FsPP + IPP and GGPP complex structures. S7
Ligand interaction diagrams S8-S9
Binding motif of bisphosphonate-GGPP complex structures S10
Electron densities S11
Supporting Tables S12-S23
Data collection and refinement statistics S12
Comparison of Liaison results for protein target receptors investigated S13
X-ray/Glide docking RMSD calculations S14
Liaison scoring function parameterization output S15-S16
CoMSIA results S17-S22
Microanalysis data S23

Supporting Methods:

2-[3-(3, 7-Dimethyl-octa-2, 6-dienyl)-pyridinium-1-yl]ethylidene-1,1-bisphosphonic acid (44).

Anal ($C_{17}H_{27}NO_6P_2$•$0.5H_2O$) C, H, N. ^{1}H NMR (D_2O, 400 MHz): δ 8.58 (1H, s), 8.45 (1H, d, *J* = 6 Hz), 8.01 (1H, d, *J* = 8Hz), 7.62 (1H, dd, *J* = 6Hz, 12.4 Hz), 5.2 (1H, t, *J* = 7.2Hz), 4.9 (1H, t, *J* = 7.5Hz), 4.6-4.8 (2H, m), 3.42 (2H, d, *J* = 7.2 Hz), 2.10 (1H, tt, *J* = 6.5Hz, 21Hz), 1.89-1.96 (4H, m), 1.52 (3H, s), 1.46 (3H, s), 1.40 (3H, S). ^{31}P NMR (D_2O, 162M Hz): δ 15.3.

2-(3-Decylpyridinium-1-yl)ethylidene-1,1-bisphosphonic acid (17)

Anal ($C_{17}H_{31}NO_6P_2$•$0.5H_2O$) C, H, N. ^{1}H NMR (D_2O, 400 MHz): δ 8.60 (1H,s), 8.55 (1H, d, J = 6.0 Hz), 8.07 (1H, d, J = 7.5Hz), 7.64 (1H, t, J = 6.5Hz), 4.6-4.8 (2H, m), 2.62 (2H, t, J = 7.5Hz), 2.09 (1H, tt, J = 21Hz, 7.0Hz), 1.48-1.52 (2H, m), 1.00-1.19 (10H, m), 0.69 (3H, t, *J* = 6.5Hz). ^{31}P NMR (D_2O, 162M Hz): δ 15.7.

2-(3-Octylsulfonamidopyridinium-1-yl)ethylidene-1,1-bisphosphonic acid (34).

Anal ($C_{15}H_{28}N_2O_8P_2S$) C, H, N. ^{1}H NMR (D_2O, 400 MHz): δ 8.01-8.10 (2H, m), 7.74 (1H, d, J = 8.8Hz), 7.47 (1H, dd, J = 12Hz, 6Hz), 4.53-4.67 (2H, m), 3.0 (2H, t, J = 7.2Hz), 2.15 (1H, tt, J = 21Hz, 7.2Hz), 1.42-1.62 (2H, m), 0.98-1.30 (10H, m), 0.62 (3H, t, *J* = 6.8Hz) . ^{31}P NMR (D_2O, 162M Hz): δ 14.9.

2-(3-Bromo-5-decyloxylpyridinium-1-yl)ethylidiene-1,1-bisphosphonic acid. (18).

Anal ($C_{17}H_{29}BrNNaO_7P_2$) C, H, N. ^{1}H NMR (D_2O, 500 MHz): δ 8.62 (1H, s), 8.54 (1H, s), 8.06 (1H, s), 4.59-4.74 (2H, m), 4.09 (2H, t, *J* = 6.5Hz), 2.25 (1H, tt, *J* = 22.5Hz, 7.5Hz), 1.60-1.68 (2H, m), 1.00-1.29 (14H, m), 0.68 (3H, t, *J* = 7.0Hz). ^{31}P NMR (D_2O, 202M Hz): δ 14.4.

2-[Decyloxy-5-(3, 5-difluoro-phenyl)-pyridinum-1-yl]ethylidene-1,1-bisphosphonic acid. (13).

Anal ($C_{23}H_{33}F_2NO_7P_2$•$0.5H_2O$) C, H, N. ^{1}H NMR (D_2O, 500 MHz): δ 8.82 (1H, s), 8.66 (1H, s), 7.55 (1H, s), 7.05 (1H, d, *J* = 6.0Hz), 6.75 (1H, t, *J* = 9Hz), 4.75-4.83 (2H, m), (1H, s), 4.09 (2H, t, J = 6.5Hz), 2.25

(1H, tt, *J* = 22.5Hz, 7.5Hz), 1.60-1.68 (2H, m), 1.00-1.29 (14H, m), 0.53 (3H, t, J = 7.5Hz). ^{31}P NMR (D_2O, 202M Hz): δ 14.4. ^{19}F NMR (D_2O, 470M Hz): δ -108.6.

2-[3-(3-Butoxyphenyl)pyridinium-1-yl]ethylidene-1,1-bisphosphonic acid (35).

Anal ($C_{17}H_{23}NO_7P_2$) C, H, N. ^{1}H NMR (D_2O, 500 MHz): δ 9.07 (1H, s), 8.73 (1H, d, *J* = 6Hz), 8.43 (1H, d, *J* = 8Hz), 7.81 (1H, t, *J* = 8.0Hz), 7.35 (1H, t, *J*= 8.0Hz), 7.19 (1H, d, J = 8.0Hz), 7.17 (1H, s), 6.97 (1H, d, *J* =8.0Hz), 4.68-4.85 (2H, m), 3.89 (2H, t, *J* = 6.5Hz), 2.28 (1H, tt, *J*= 21Hz, 7.0Hz), 1.55-1.63 (2H, m), 1.25-1.29 (2H, m), 1.11-1.21 (2H, m), 0.69 (3H, t, *J* = 6.0Hz). ^{31}P NMR (D_2O, 202M Hz): δ 15.4.

2-(3-Dodecyloxypyridinium-1-yl)ethylidene-1,1-bisphosphonic acid (36).

Anal ($C_{19}H_{35}NO_7P_2$•$0.5H_2O$) C, H, N. ^{1}H NMR (D_2O, 400 MHz): δ 8.40 (1H, t, *J* = 7.2Hz), 8.25 (1H, d, *J* = 4.0Hz) 7.52-7.61 (2H, m), 4.50-4.63 (2H, m), 3.90 (2H, t, *J* = 6.0Hz), 2.25 (1H, tt, *J* = 22Hz, 7.0Hz), 1.52-1.64 (2H, m), 1.00-1.29 (18H, m), 0.69 (3H, t, *J* = 6.5Hz). ^{31}P NMR (D_2O, 202M Hz): δ 14.3.

2-[3-(3,7-Dimethyl-octyloxy)-pyridinium-1-yl]ethylidene-1,1-bisphosphonic acid (24)

Anal ($C_{17}H_{31}NO_7P_2$•$0.75H_2O$) C, H. ^{1}H NMR (D_2O, 500 MHz): δ 8.89 (1H, s), 8.43 (1H, d, *J* = 6Hz), 7.86 (1H, d, *J* = 8Hz), 7.71 (1H, t, *J* = 8.0Hz), 4.68-4.80 (2H, m), 4.25 (2H, t, *J* = 6.5Hz), 2.23 (1H, tt, *J*= 21Hz, 7.0Hz), 1.78-1.87 (1H, m), 1.56-1.69 (2H, m), 1.40-1.50 (1H, m), 1.00-1.3.2 (6H, m), 0.82 (3H, d, *J* = 6.5Hz), 0.74 (3H, d, *J* = 6.5Hz). ^{31}P NMR (D_2O, 202M Hz): δ 15.4.

2-[3-(2-Ethoxy-ethoxy)propylpyridinium-1-yl]ethylidene-1,1-bisphosphonic acid (54).

Anal. ($C_{14}H_{24}NNaO_8P_2$) C, H. ^{1}H NMR (D_2O, 500 MHz): δ 8.65 (1H, s), 8.58 (1H, d, *J* = 6Hz), 8.23 (1H, d, *J* = 8Hz), 7.78 (1H, t, *J* = 8.0Hz), 4.68-4.80 (2H, m), 3.39 (8H, m), 2.65-2.80 (2H, t, *J* = 6.5Hz, 1H, overlapped, m), 1.82 (2H, m), 1.00 (3H, t, *J* = 7 Hz). ^{31}P NMR (D_2O, 202M Hz): δ 15.4.

2-(3-Decyloxypyridinium-1-yl)ethylidene-1,1-bisphosphonic acid (15).

Anal ($C_{17}H_{31}NO_7P_2$•0.5H_2O) C, H, N. ^{1}H NMR (D_2O, 400 MHz): δ 8.50 (1H, t, *J* = 7.2Hz), 8.25 (1H, d, J = 6.0Hz) 7.71 (2H, d, *J*= 8Hz), 7.60 (1H, t, *J* = 8Hz), 4.50-4.63 (2H, m), 4.00 (2H, t, *J* = 6.5Hz), 2.11 (1H, tt, *J*= 22Hz, 7.0Hz), 1.52-1.64 (2H, m), 1.00-1.29 (14H, m), 0.64 (3H, t, *J* = 6.5Hz). ^{31}P NMR (D_2O, 202M Hz): δ 15.69.

2-(4-Octylpyridinium-1-yl)ethylidene-1,1-bisphosphonic acid (20)

Anal ($C_{15}H_{27}NO_6P_2$•0.75H_2O)C, H, N. ^{1}H NMR (D_2O, 500 MHz): δ 8.52 (2H, d, *J* = 6.5Hz), 7.55 (2H, t, *J* = 6.5Hz), 4.68-4.80 (2H, m), 2.68 (2H, t, *J* = 7.5Hz), 2.15 (1H, t, *J* = 22Hz, 7.0Hz), 1.49-1.53 (2H, m), 1.00-1.21 (10H, m), 0.63 (3H, t, *J* = 6.5Hz). ^{31}P NMR (D_2O, 202M Hz): δ 15.2.

2-(3-Heptyloxypyridinium-1-yl)ethylidene-1,1-bisphosphonic acid (30).

Anal ($C_{14}H_{25}NO_7P_2$•0.5H_2O) C, H, N. ^{1}H NMR (D_2O, 400 MHz): δ 8.42 (1H, t, *J* = 7.2Hz), 8.25 (1H, d, *J* = 6.0Hz) 7.71 (2H, d, *J* = 8.8Hz), 7.60 (1H, t, *J* = 8.8Hz), 4.50-4.63 (2H, m), 4.00 (2H, t, *J* = 6.4Hz), 2.11 (1H, tt, *J* = 21Hz, 6.8Hz), 1.52-1.64 (2H, m), 1.00-1.29 (6H, m), 0.64 (3H, t, *J* = 6.8Hz). ^{31}P NMR (D_2O, 162M Hz): δ 15.64.

2-(3-Decylaminopyridinium-1-yl)ethylidene-1,1-bisphosphonic acid (26).

Anal ($C_{17}H_{32}N_2O_6P_2$•0.75H_2O)C, H, N. ^{1}H NMR (D_2O, 400 MHz): δ 7.92 (1H, s), 7.83 (1H, d, *J* = 5.2Hz), 7.20-7.31 (2H, m), 2.1 (2H, t, *J* = 6.8Hz), 2.20-2.31 (1H, m), 1.52-1.64 (2H, m), 1.00-1.29 (14H, m), 0.68 (3H, t, *J* = 6.8Hz). ^{31}P NMR (D_2O, 162M Hz): δ 14.75.

2-(3-Hexylpyridinium-1-yl)ethylidene-1,1-bisphosphonic acid (42).

Anal. ($C_{13}H_{22}NNAO_6P_2$•0.5H_2O) C, H. N. 1H NMR (500 MHz, D2O): δ 8.53-8.60 (m, 2H), 8.18 (d, *J* = 8 Hz, 1H), 7.72 (t, *J* = 6 Hz, 1H), 4.65-4.80 (m, 2H), 2.55-2.70 (m, 3H), 1.45-1.50 (m, 2H), 1.05-1.20 (m, 6H), 0.66 (t, *J* = 6.5 Hz, 3H). 31P NMR (202 MHz, D2O): δ 14.6.

2-[3-(Octyn-1-yl)pyridinium-1-yl]ethylidene-1,1-bisphosphonic acid (29).

Anal. ($C_{15}H_{23}NO_6P_2$•$0.5H_2O$). 1H NMR (500 MHz, D2O): δ 8.78 (s, 1H), 8.65 (d, *J* = 8 Hz, 1H), 8.27 (d, *J* = 8 Hz, 1H), 7.79 (t, *J* = 6 Hz, 1H), 4.65-4.80 (m, 2H), 2.60-2.70 (m, 1H), 2.33 (t, *J* = 6 Hz, 1H), 1.15-1.50 (m, 8H); 0.69 (t, *J* = 6.5 Hz, 3H); ^{31}P NMR (202 MHz, D2O): δ 14.7.

1-Hydroxy-2-(3-octylpyridinium-1-yl)ethylidene-1,1-bisphosphonic acid (27).

Anal. ($C_{15}H_{25}NO_7P_2Na_2$ •H_2O) C, H. N. 1H NMR (500 MHz, D2O): δ 8.53-8.60 (m, 2H); 8.15 (d, *J* = 8 Hz, 1H), 7.68 (t, *J* = 6 Hz, 1H), 4.65-4.80 (m, 2H), 2.67 (t, *J* = 6.5 Hz, 2H), 1.45-1.50 (m, 2H), 1.05-1.20 (m, 10H), 0.68 (t, *J* = 6.5 Hz, 3H, CH3); ^{31}P NMR (202 MHz, D2O): δ 14.5.

Figure S1: Binding modes observed in *S. cerevisae* GGPPS-bisphosphonate complexes.

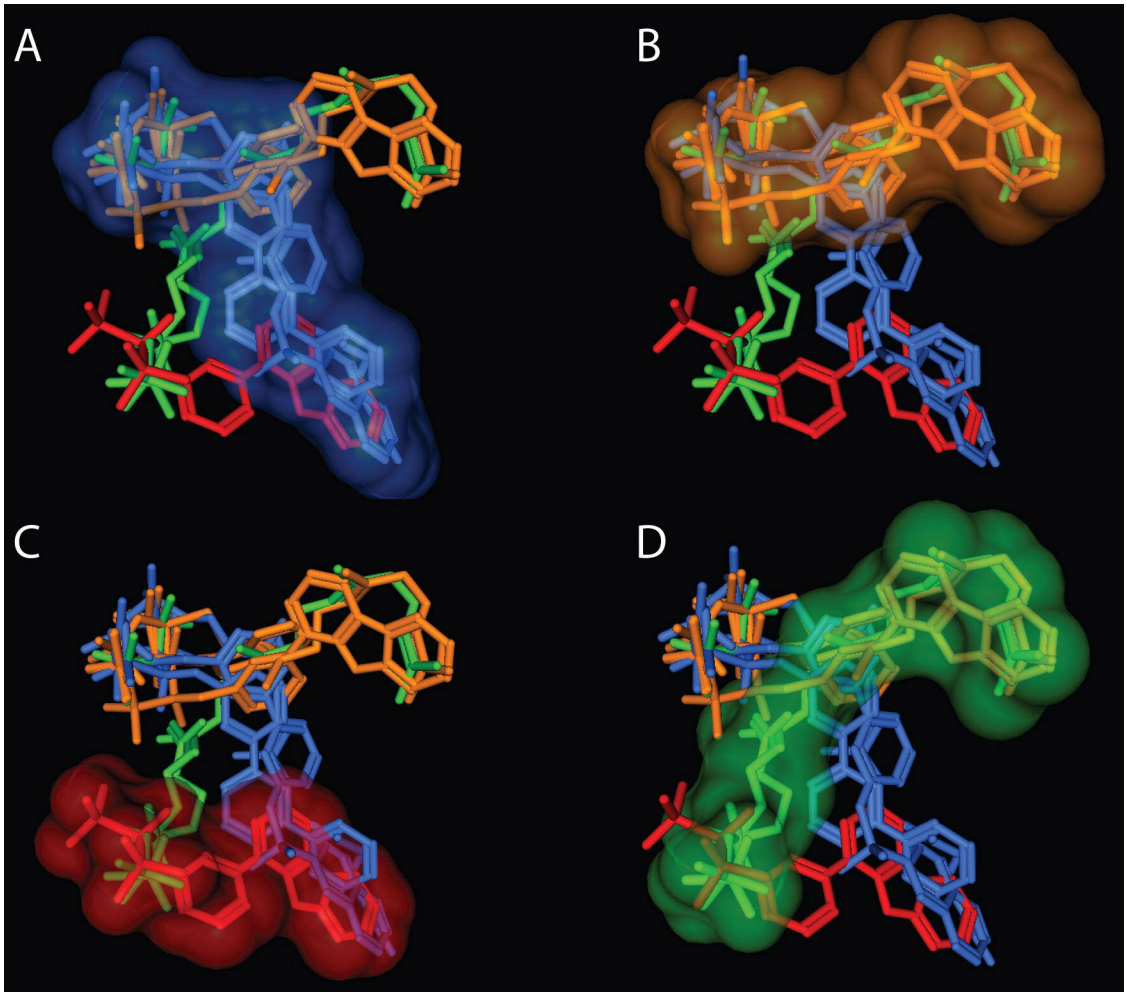

Figure adapted from Guo et al., *PNAS USA*, 104, 10022-10027 (2007).

Figure S2: 11 (BPH-23) (PDB File: 2z52) bound to GGPPS superimposed on FPP, FsPP + IPP and GGPP complex structures (PDB Files: 2e90, 2e8t, 2z4v). The "BPH" designations shown below are those used in the respective PDB File entries.

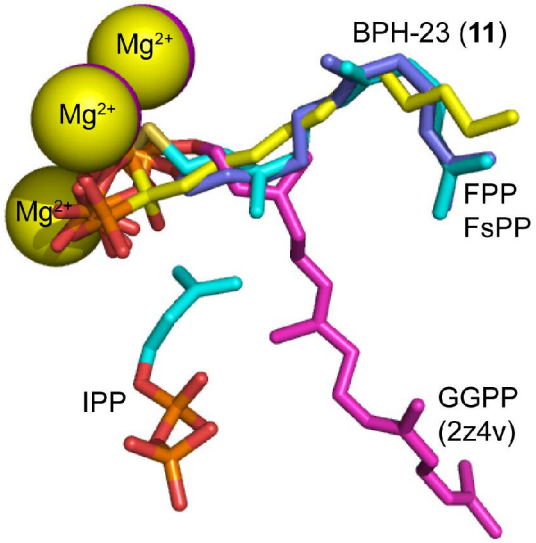

Figure S3: Ligand interactions diagrams for GGPPS-bisphosphonate complexes BPH-252, BPH-252-P21, BPH-23, BPH-28 (PDB Files: 2z4y, 2z4x, 2z52, 2z60). The "BPH" designations shown below are those used in the respective PDB File entries.

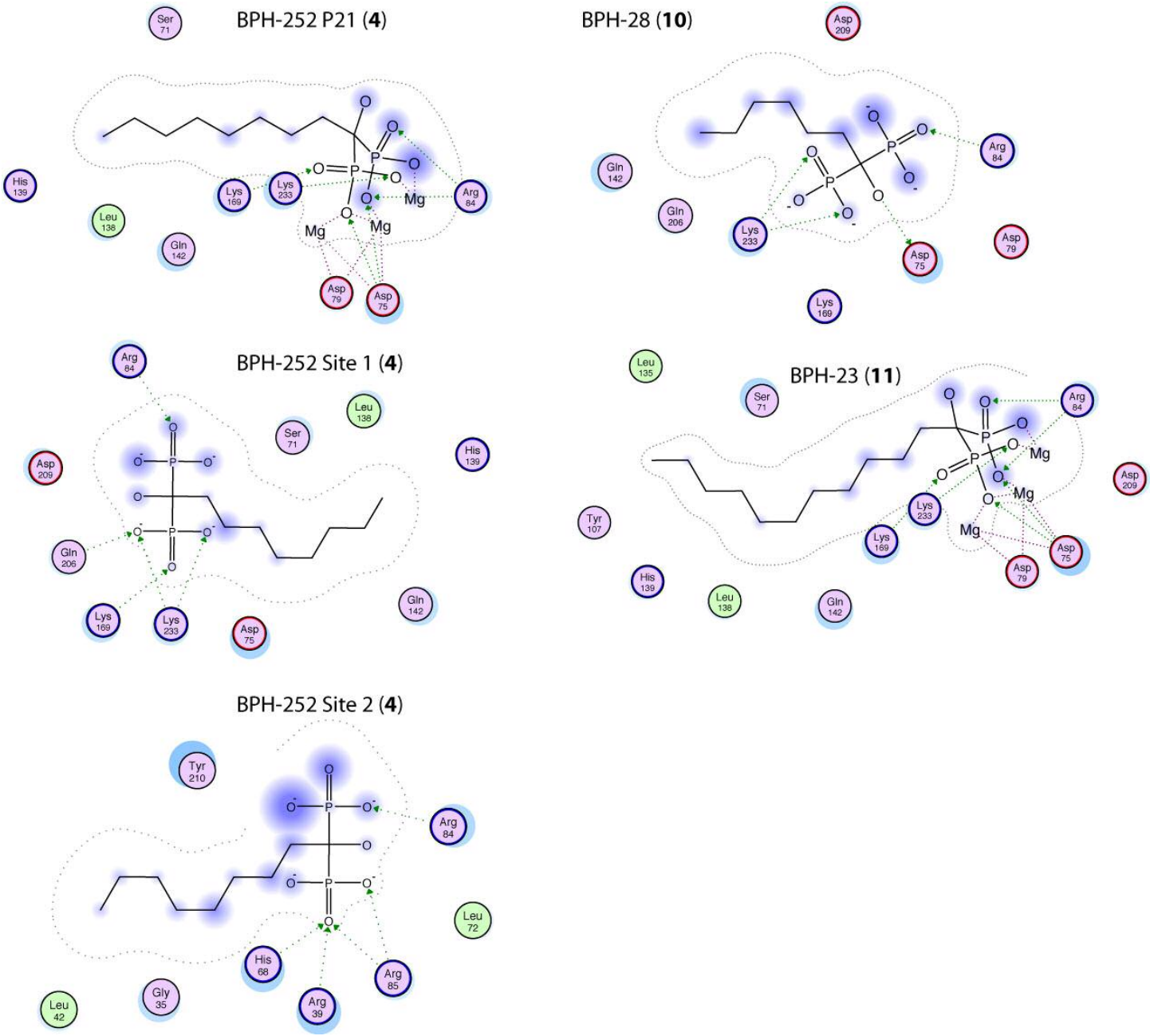

Figure S4: Ligand Interaction diagrams for GGPPS-bisphosphonate complexes BPH-749, BPH-798 and BPH-806 (PDB Files: 2z4w, 2z4z, 2z78). The "BPH" designations shown below are those used in the respective PDB File entries.

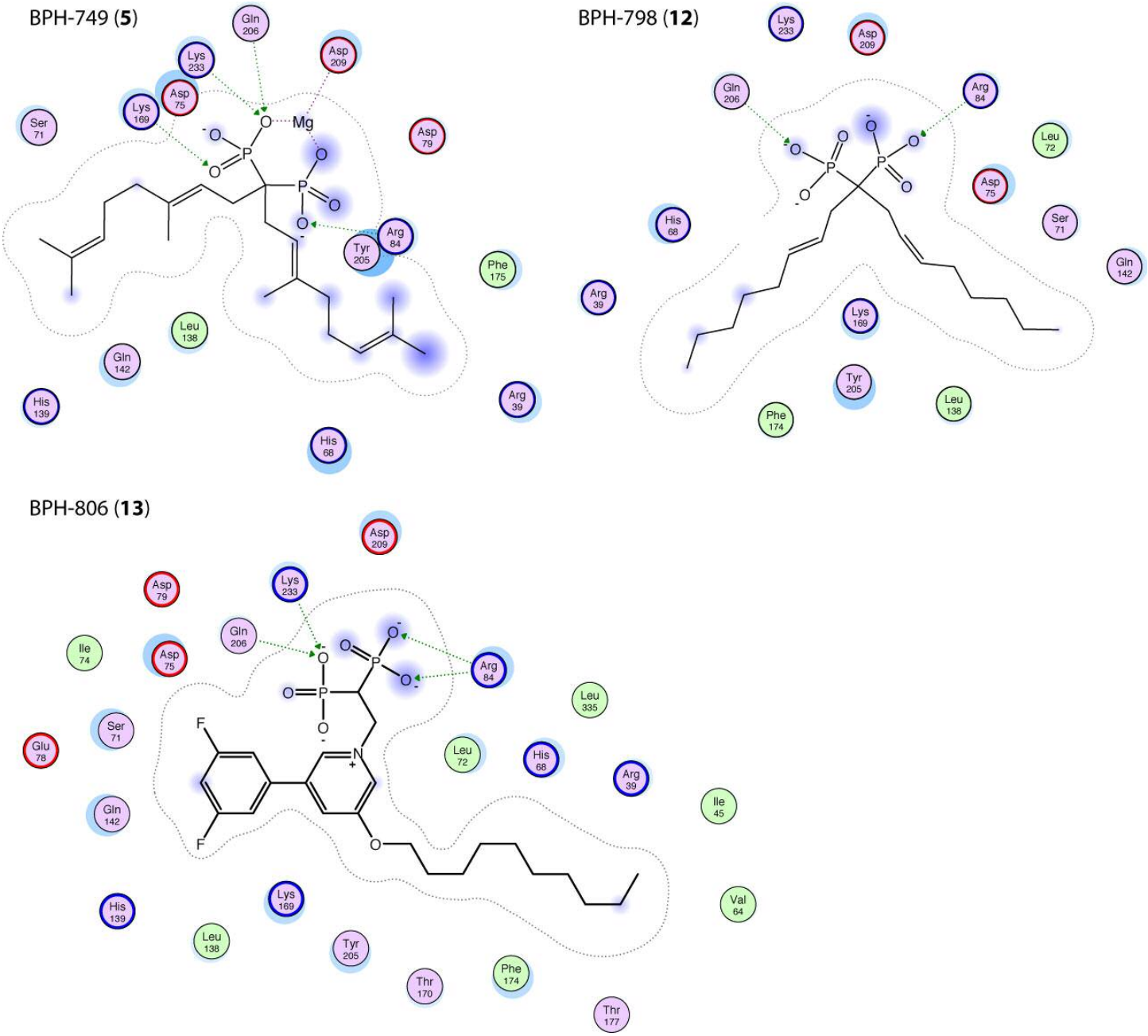

Figure S5: Binding motif of bisphosphonate-GGPP complex structures. A) **10** (BPH-28, PDB File: 2z50) and B) **9** (BPH-742, PDB File: 2z7I) bound to GGPPS superimposed on FsPP + IPP (PDB File: 2e8t) and GGPP (PDB File: 2z4v) complex structures

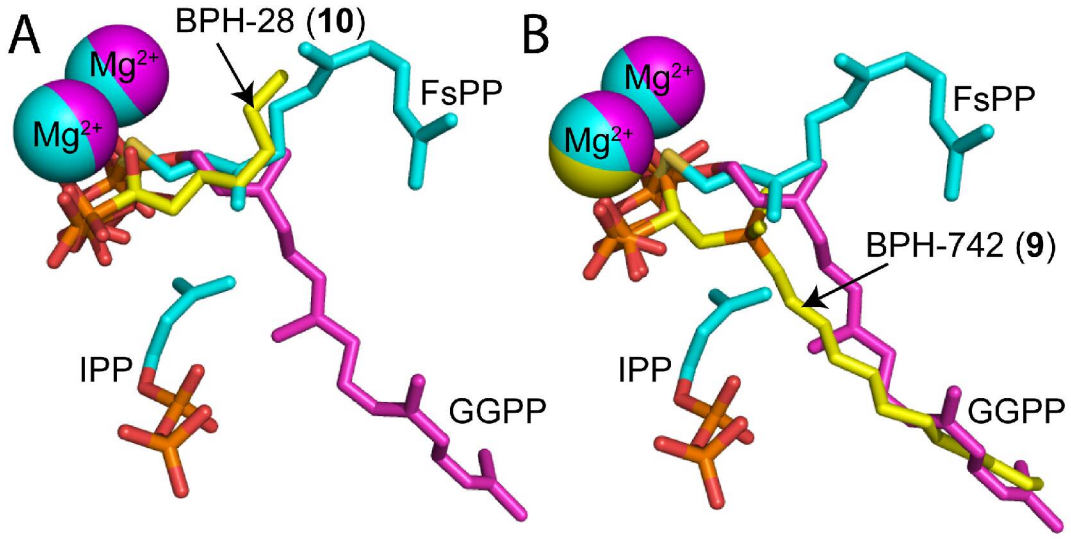

Figure S6: Electron densities (green contoured at 1σ, red at 3σ) for GGPP, **4, 5, 9, 10, 11, 12, 13** bound to GGPPS. The "BPH" designations shown below are those used in the respective PDB File entries.

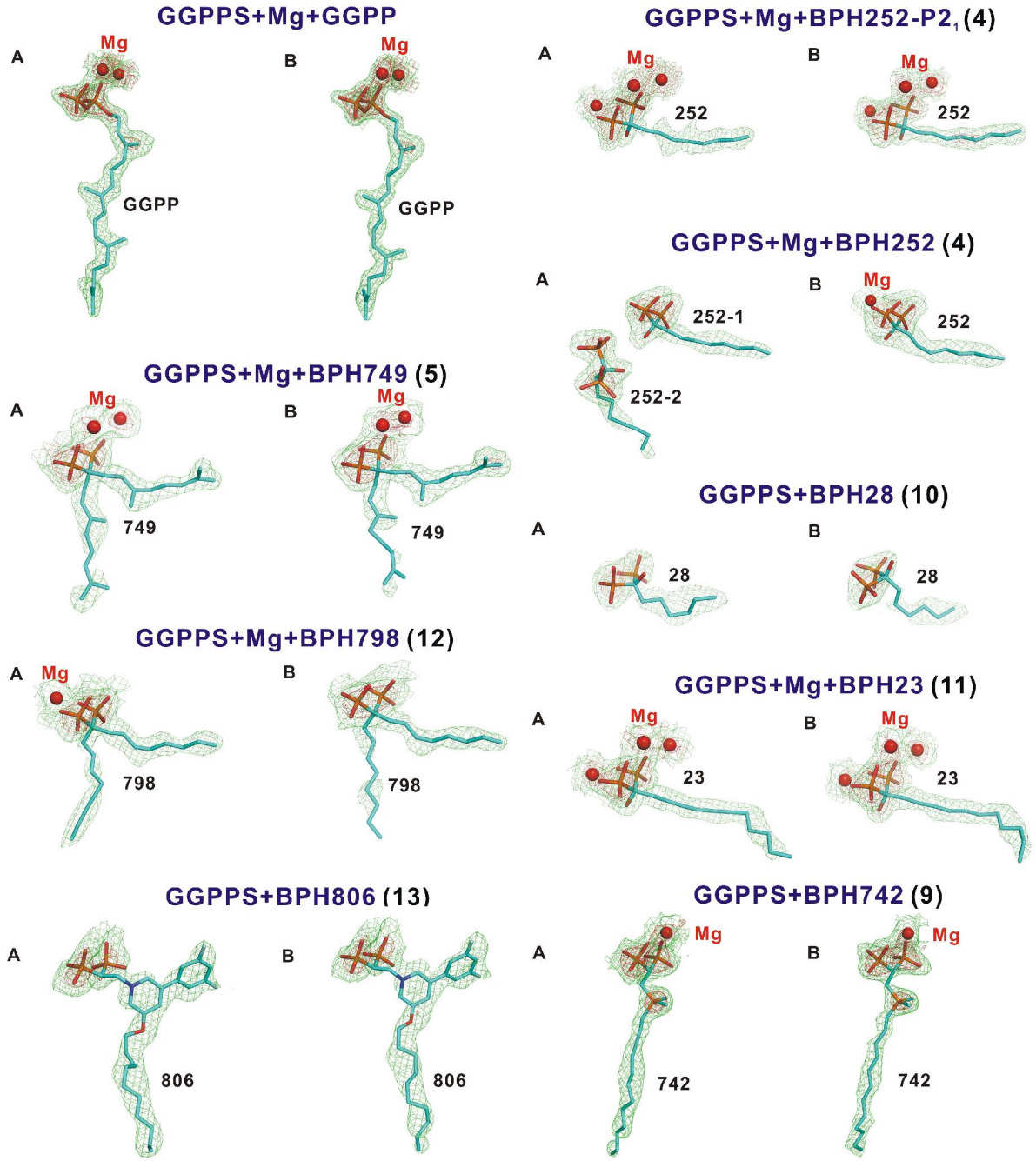

Table S1. Data Collection and Refinement Statistics for GGPP-Bisphosphonate Crystals.

Names	GGPPS-**10**	GGPPS-Mg-**9**
PDB number	2z50	2z7I
Data Collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Resolution (Å)[a]	30-2.01	30-2.10
	(2.08-2.01)	(2.18-2.10)
Unit Cell Dimensions		
a (Å)	47.13	48.11
b (Å)	119.2	116.34
c (Å)	129.71	129.43
No. of reflections		
Observed	204046 (16667)	202498 (35537)
Unique	49435 (4762)	81045 (7897)
Completeness (%)	99.2 (97.0)	99.5 (98.4)
R_{merge} (%)	3.2 (34.6)	5.1 (23.9)
I/s(I)	39.1 (4.4)	29.9(6.5)
Refinement		
No. of reflections	47496 (4286)	41743 (3708)
R_{work} (%)	19.0 (24.1)	18.4 (25.0)
R_{free} (%)	23.9 (28.2)	24.4 (31.3)
Geometry deviations		
Bond lengths (Å)	0.023	0.017
Bond angles (°)	1.7	1.7
No. of all protein atoms	4593	5159
Mean B-values (Å^2)	37.9	36.2
No. of all cofactor atoms	32	52
Mean B-values (Å^2)	60	38.8
No. of water molecules	637	471
Mean B-values (Å^2)	58.8	47.5
Ramachandran plot (%)		
Most favored	96.3	96
Additionally allowed	3.6	4
Generously allowed	0.2	0

[a] Values in the parenthesis are the highest resolution shells.

Table S2: Comparison of Liaison Results for Protein Target Receptors Investigated

	Target Receptor				
	BPH-23 (**12**)	BPH-629 (**8**)	BPH-749 (**5**)	BPH-742 (**9**)	BPH-28 (**10**)
PDB File	2z4z	2e93	2z4w	2z7i	2z5o [f]
Number of Mg^{2+} Ions	3	2	2	1	0
Accuracy[a]	60%	100%	70%	70%	0
r^2	0.7	0.7	0.7	0.7	n/a
q^2 [b]	0.6	0.6	0.7	0.6	n/a
Outliers [c]	1	0	2	1	n/a
Randomization r^2 [d]	0.05	0.05	0.05	0.05	n/a
Regression Coefficients [e]					
van der Waals	-2.10E-02	-2.82E-02	-3.40E-02	-2.12E-02	n/a
Cavity	-1.43E-01	-1.89E-01	-2.32E-01	-1.43E-01	n/a
SlogP	3.07E-01	1.70E-01	2.10E-01	3.07E-01	n/a

[a] Percentage of docked poses residing in the crystallographically observed binding pocket. Computed based on a total of 12 GGPPS-bisphosphonate structures (Table S3).

[b] Cross-validated r^2, representing a metric of the predictivity of the model.

[c] Number of statistical outliers identified by MLR regression module in Strike (Schrödinger, LLC)

[d] Experiment vs. predicted activity correlation upon 100 iterations of data randomization; low values here indicate a good correlation is unlikely to be obtained by chance.

[e] Regression coefficients fit to Liaison (Schrödinger, LLC) parameters. These values serve as the basis for the reparameterized scoring function.

[f] Sufficient structure convergence not obtained to proceed with Liaison calculations.

Table S3: X-ray/Glide docking RMSD calculations using BPH-629 (**8**) grid (all values shown in Å).

pose #	BPH-252 (**4**)	BPH-749 (**5**)	BPH-675 (**6**)	BPH-364 (**7**)	BPH-629 (**8**)	BPH-742 (**9**)	BPH-28 (**10**)	BPH-798 (**12**)
1	1.36	1.06	2.49	1.41	0.65	1.32	1.12	2.45
2	1.01	0.88	2.54	1.64	0.60	1.62	1.12	0.98
3	1.16	0.90	2.39	1.62	0.64	1.38	1.19	1.26
4	1.31	0.99			1.26	1.20	1.26	0.98
5	1.51	1.42			0.98	0.96	1.11	1.13
avg	1.27	1.05	2.47	1.56	0.83	1.29	1.16	1.36

pose #	BPH-806 (**13**)	BPH-715 (**15**)	BPH-91 (**55**)	BPH-261 (**57**)
1	1.54	0.85	1.07	1.15
2	1.79	0.85	1.01	0.79
3	1.41	0.99		0.82
4	1.24	1.11		1.09
5		1.29		1.35
avg	1.49	1.02	1.04	1.04

Table S4: Liaison Scoring Function Parameterization Output

```
##############################################################################
###                                                                        ###
###                          Strike v1.6.308                               ###
###                     Schrodinger, LLC - 2006                            ###
###                                                                        ###
##############################################################################

Job started: Wed Dec  5 09:54:34 2007

Job ID: vengeance-0-4756c9b8

SCHRODINGER: /usr/local/schrodinger
MMSHARE_EXEC: /usr/local/schrodinger/mmshare-v16207/bin/IRIX-mips4
MAESTRO_EXEC: /usr/local/schrodinger/maestro-v80308/bin/IRIX-mips4

Commands input for task number 1
====================================
runMode=train
dataFile=strike_buildqsar_training.csv
csvFile=strike_buildqsar_training-out.csv
normalize=range
delim=,
activityLabel=GGPPS pIC50
model=MLRO
numOptDescript=3
numOptModel=10
columnHeader=1
autoScale=1

Activity column(5) identified is labeled:: GGPPS pIC50
Descriptors employed (alias <== input descriptor title):
=========================================================
Liai.vdw>   <==   Liaison <Uvdw>
Liai.cav>   <==   Liaison <Ucav>
Liai.ele>   <==   Liaison <Uele>
SlogP       <==   SlogP
GGPP.IC50   <==   GGPPS pIC50

Molecules employed:
====================
1:61

Correlation Matrix for input variables:
=======================================
             Liai.vdw>  Liai.cav>  Liai.ele>      SlogP
------------------------------------------------------------
Liai.vdw>       1.0000
Liai.cav>      -0.0517     1.0000
Liai.ele>       0.2990     0.2246     1.0000
SlogP          -0.7375    -0.2389    -0.4016     1.0000
------------------------------------------------------------

For MLR model created using # rows: 61 # columns 3



MLR Regression Statistics:
-------------------------------------------------
     S.D.    R-Squared  F(3.0,57.0)     P
-------------------------------------------------
    0.5683    0.6499      35.3      5.057e-13
```

```
-------------------------------------------------

MLR Regression Coefficients and T-Values:
-------------------------------------------------
      Variable    Coefficient     Std. Err.       T
-------------------------------------------------
     Intercept     6.7722e+00    2.3680e-01    28.5992
Liaison <Uvdw>    -2.8258e-02    7.8063e-03     3.6198
Liaison <Ucav>    -1.8500e-01    6.8980e-02     2.6819
         SlogP     1.4626e-01    5.1754e-02     2.8260
-------------------------------------------------

Cross Validation leave-3-out Results Over 10 Cycles
-------------------------------------------------
   q^2            RMS
-------------------------------------------------
   0.5974        0.5906
-------------------------------------------------

Randomization over 610 Cycles
-------------------------------------------------
  r^2          S.D.           p
-------------------------------------------------
   0.0506        0.9358     0.000e+00
-------------------------------------------------

Predicted Results: (Full Results shown in Table S1, Main Text)
------------------------------

RMS        0.5494
Q^2        0.6499


Job completed successfully
User time (sec):       0.7400
System time (sec):     0.0100
Real time (sec):       0.7600
```

Table S5: CoMSIA Feature Alignment Full Test Set Results for GGPPS Inhibition

	Experimental			Predicted pIC_{50} (M)						
Compound	IC_{50} (μM)	pIC_{50} (M)	Training	1	2	3	4	5	Predicted	Residual
9	0.10	7.0	6.5	6.4	6.6	6.1	6.1	6.2	6.3	0.7
14	0.28	6.6	6.3	**5.5**	6.3	6.3	6.7	6.8	5.5	1.0
15	0.28	6.6	6.0	6.1	6.0	5.9	5.9	**5.7**	5.7	0.8
16	0.35	6.5	6.2	6.0	6.2	5.9	**5.8**	5.9	5.8	0.7
17	0.40	6.4	6.0	5.9	**6.0**	5.9	5.8	5.8	6.0	0.4
18	0.51	6.3	6.3	6.3	6.3	**5.8**	6.2	6.1	5.8	0.5
19	0.59	6.2	6.0	6.2	6.1	5.9	6.0	5.9	6.0	0.2
20	0.66	6.2	5.6	**5.6**	5.7	5.5	5.5	5.5	5.6	0.5
4	0.69	6.2	5.5	5.6	5.6	5.5	5.4	5.4	5.5	0.7
21	0.72	6.1	6.0	6.1	**5.6**	6.0	5.8	6.0	5.6	0.5
13	0.76	6.1	6.4	6.3	6.4	6.5	6.5	**6.2**	6.2	-0.1
22	0.89	6.1	6.0	5.8	6.1	5.7	**5.5**	5.7	5.5	0.5
23	0.89	6.1	5.6	**5.3**	5.5	5.7	5.7	5.5	5.3	0.8
5	0.98	6.0	6.2	6.1	6.2	6.0	6.1	6.1	6.1	-0.1
24	1.15	5.9	5.7	5.6	5.7	5.6	5.6	**5.5**	5.5	0.5
25	1.23	5.9	6.3	6.1	6.3	**6.0**	5.9	6.0	6.0	-0.1
26	1.26	5.9	6.0	5.9	**6.0**	5.9	5.8	5.8	6.0	-0.1
27	1.38	5.9	5.9	5.8	5.9	5.7	**5.5**	5.6	5.5	0.3
28	1.48	5.8	6.3	6.1	6.2	6.0	5.9	**6.0**	6.0	-0.2
29	1.74	5.8	5.7	**5.5**	5.7	5.7	5.5	5.5	5.5	0.3
12	1.86	5.7	6.0	5.6	6.1	6.2	6.0	6.1	6.0	-0.3
30	2.14	5.7	5.6	5.6	5.6	**5.2**	5.3	5.3	5.2	0.5
31	2.14	5.7	5.8	5.8	5.9	**5.8**	5.7	5.8	5.8	-0.1
32	2.34	5.6	5.7	5.4	5.7	6.0	**5.9**	5.9	5.9	-0.3
33	2.51	5.6	6.1	5.9	**6.4**	6.4	6.5	6.4	6.4	-0.8
34	2.51	5.6	5.6	5.8	5.7	5.5	5.5	5.3	5.5	0.1
35	2.51	5.6	5.2	**5.1**	5.2	5.2	5.2	5.2	5.1	0.5
6	2.69	5.6	5.4	**6.8**	5.4	5.8	5.9	6.1	6.8	-1.3
11	2.69	5.6	5.9	5.7	5.8	5.9	5.9	**5.7**	5.7	-0.1
36	3.02	5.5	5.8	5.9	5.9	5.7	**5.9**	5.7	5.9	-0.4
37	3.24	5.5	5.8	5.6	**5.9**	5.6	5.5	5.6	5.9	-0.4
8	3.98	5.4	5.4	5.4	**4.9**	5.5	5.4	5.5	4.9	0.5
38	4.07	5.4	5.5	5.4	5.6	5.6	5.6	**5.6**	5.6	-0.2
39	4.17	5.4	5.5	5.4	5.4	**5.4**	5.5	5.3	5.4	0.0

40	4.57	5.3	5.0	4.9	5.2	5.3	5.3	5.1	5.2	0.2
41	4.68	5.3	5.5	**5.7**	5.5	5.5	5.6	5.5	5.7	-0.3
42	6.76	5.2	5.4	5.1	5.4	5.4	**5.2**	5.3	5.2	-0.1
43	8.71	5.1	4.6	4.6	4.6	**4.5**	4.6	4.7	4.5	0.6
44	8.91	5.1	5.5	5.4	**5.5**	5.5	5.3	5.3	5.5	-0.5
45	10.00	5.0	4.8	4.6	4.8	4.8	4.7	**4.8**	4.8	0.2
7	11.22	5.0	5.3	5.2	5.3	**5.5**	5.4	5.4	5.5	-0.5
10	11.22	5.0	4.9	4.7	4.9	4.9	5.0	**5.0**	5.0	0.0
46	14.13	4.9	4.8	**4.6**	4.8	4.9	4.8	4.8	4.6	0.2
47	27.54	4.6	4.9	4.7	4.9	4.9	4.9	5.0	4.9	-0.3
48	53.70	4.3	4.3	4.2	4.3	4.2	**4.2**	4.2	4.2	0.1
49	66.07	4.2	4.3	4.3	4.2	4.2	4.3	**4.4**	4.4	-0.2
50	66.07	4.2	4.0	4.0	**4.1**	4.1	4.1	4.1	4.1	0.0
51	74.13	4.1	4.3	4.1	4.3	**4.8**	3.8	3.8	4.8	-0.6
52	79.43	4.1	6.3	5.9	**6.3**	5.8	5.7	5.8	6.3	-2.2
53	83.18	4.1	4.7	4.4	4.6	4.7	**5.0**	4.7	5.0	-0.9
54	93.33	4.0	4.1	**5.1**	4.1	4.3	4.4	4.3	5.1	-1.1
55	100.00	4.0	3.9	3.9	3.9	3.9	**3.9**	3.8	3.9	0.1
56	107.15	4.0	4.1	4.1	4.1	4.1	4.1	**4.2**	4.2	-0.2
57	107.15	4.0	3.9	3.9	3.9	**4.4**	4.3	4.2	4.4	-0.4
58	117.49	3.9	4.1	4.0	**4.3**	4.1	4.2	4.2	4.3	-0.4
59	141.25	3.9	3.6	3.8	3.7	3.7	3.9	3.8	3.8	0.1
60	169.82	3.8	3.9	3.8	4.0	**4.1**	4.0	4.0	4.1	-0.3
62	181.97	3.7	3.8	4.1	3.8	3.6	3.4	3.4	3.7	0.1
63	316.23	3.5	3.7	3.7	3.6	3.9	3.9	4.0	3.8	-0.3
64	331.13	3.5	3.3	**4.6**	3.3	3.4	3.7	3.8	4.6	-1.1
65	436.52	3.4	3.4	3.4	3.3	3.3	3.4	3.5	3.4	0.0
		q2	0.59	0.68	0.72	0.54	0.56	0.52		
		r2	0.79	0.85	0.92	0.81	0.80	0.79		
		N	4	4	4	3	3	3		
		F	73	65	139	67	63	59		
		n	61	51	51	51	51	51		
		%Steric	0.19	0.22	0.19	0.18	0.21	0.19		
		%Hydrophobic	0.35	0.36	0.36	0.38	0.38	0.38		
		%Donor	0.21	0.18	0.20	0.24	0.23	0.23		
		%Acceptor	0.25	0.24	0.26	0.20	0.18	0.20		

N = number of components; n = Number of training set compounds

Table S6: CoMSIA Feature Alignment Results

```
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 (2280 vars)       0.556     0.205
 2 COMSIA_HY (2280 vars)       1.030     0.379
 3 COMSIA_DO (2280 vars)       0.602     0.222
 4 COMSIA_AC (2280 vars)       0.526     0.194

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: 3
  Scaling: COMFA_STANDARD    Intercept forced through 0.0: NO
  NIPALS: (Max Iter = 100; EPS = 0.000100)
  Target (Y) Variable(s):
   1: GGPPS_pIC50
  Explanatory (X) Variable(s): (Actual terms: 9120 requested, 8785 used)
   2: COMSIA_ST             4: COMSIA_HY             5: COMSIA_DO
   6: COMSIA_AC
  Rows in Analysis: (61 rows)
   1: 1                     2: 2                     3: 210
   4: 23                    5: 24                    6: 252
   7: 261                   8: 278                   9: 28
  10: 30                   11: 300                  12: 364
  13: 470                  14: 472                  15: 474
  16: 483                  17: 5                    18: 527
  19: 536                  20: 579                  21: 598
  22: 601                  23: 604                  24: 608
  25: 628                  26: 629                  27: 637
  28: 638                  29: 653                  30: 675
  31: 676                  32: 678                  33: 679
  34: 683                  35: 684                  36: 685
  37: 688                  38: 693                  39: 694
  40: 715                  41: 716                  42: 717
  43: 722                  44: 727                  45: 728
  46: 733                  47: 734                  48: 742
  49: 745                  50: 749                  51: 754
  52: 798                  53: 799                  54: 804
  55: 805                  56: 806                  57: 811
  58: 812                  59: 815                  60: 816
  61: 91

Summary output
Standard Error of Estimate           0.437
R squared                          0.793
F values     ( n1= 3, n2=57 )     72.964
Prob.of R2=0 ( n1= 3, n2=57 )      0.000
```

Table S7: CoMSIA Receptor-Guided Alignment Results

```
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 (2028 vars)                      0.924     0.217
 2 COMSIA_HY (2028 vars)                      1.531     0.360
 3 COMSIA_DO (2028 vars)                      0.965     0.227
 4 COMSIA_AC (2028 vars)                      0.833     0.196

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: 5
  Scaling: COMFA_STANDARD    Intercept forced through 0.0: NO
  NIPALS: (Max Iter = 100; EPS = 0.000100)
  Target (Y) Variable(s):
   1: HumanGGPPS
  Explanatory (X) Variable(s): (Actual terms: 10140 requested, 8111 used)
   2: COMSIA_ST             4: COMSIA_HY             5: COMSIA_DO
   6: COMSIA_AC
  Rows in Analysis: (61 rows)
   1: 210                   2: 23                    3: 261
   4: 278                   5: 30                    6: 300
   7: 364                   8: 470                   9: 472
  10: 474                  11: 483                  12: 527
  13: 536                  14: 601                  15: 604
  16: 608                  17: 628                  18: 629
  19: 637                  20: 638                  21: 653
  22: 675                  23: 676                  24: 678
  25: 679                  26: 683                  27: 684
  28: 685                  29: 688                  30: 693
  31: 694                  32: 715                  33: 716
  34: 717                  35: 722                  36: 727
  37: 728                  38: 733                  39: 734
  40: 742                  41: 745                  42: 749
  43: 754                  44: 798                  45: 799
  46: 805                  47: 806                  48: 811
  49: 815                  50: 816                  51: 91
  52: 252                  53: 28                   54: 579
  55: 812                  56: 598                  57: 2
  58: 804                  59: 24                   60: 1
  61: 5

Summary output
Standard Error of Estimate           0.244
R squared                        0.938
F values     ( n1= 5, n2=55 )   165.558
Prob.of R2=0 ( n1= 5, n2=55 )     0.000
```

Table S8: Microanalysis Data for New Compounds

Compound	Required %C	Found % C	Required %H	Found %H	Required %N	Found %N
6	45.8	45.74	3.2	3.29	2.23	2.2
9	41.56	41.31	7.63	7.91	0	0
12	45.9	45.91	7.36	7.28	0	0
13	50.74	50.87	6.29	6.12	2.57	2.7
14	32.78	32.67	5.2	5.59	2.12	2.02
15	47.22	47.58	7.46	7.35	3.24	3.38
16	33.72	34.15	6.23	6.58	2.62	2.63
17	49.04	49.33	7.75	7.65	3.36	3.25
18	37.65	37.48	5.76	5.66	2.58	2.6
19	46.06	45.87	6.87	6.88	0	0
20	45.86	45.71	7.31	7.09	3.57	3.64
22	39.78	39.45	8.01	8.12	0	0
24	46.74	46.37	7.5	7.1	3.21	3.43
26	46.84	46.78	7.75	7.55	6.43	6.39
27	39.4	39.79	5.95	6.32	3.06	3.09
29	46.88	46.88	6.29	5.98	3.64	3.65
30	43.08	43.21	6.71	6.48	3.59	3.64
32	40.06	40.12	6.47	6.67	0	0
34	39.3	39.32	6.16	6.15	6.11	6.08
35	49.16	49.09	5.58	5.53	3.37	3.46
36	49.56	49.5	7.88	7.75	3.04	3.15
38	47.73	47.26	3.81	3.69	0	0
40	46.26	46.33	7.54	7.46	3.17	3.18
41	48.61	48.57	7.94	7.81	2.98	2.93
42	40.85	40.73	6.07	6.09	3.66	3.83
44	49.52	49.43	6.84	6.68	3.4	3.45
51	13.01	13.21	3.43	3.03	13.01	13.04
54	40.1	40.43	5.77	6.02	3.34	3.51