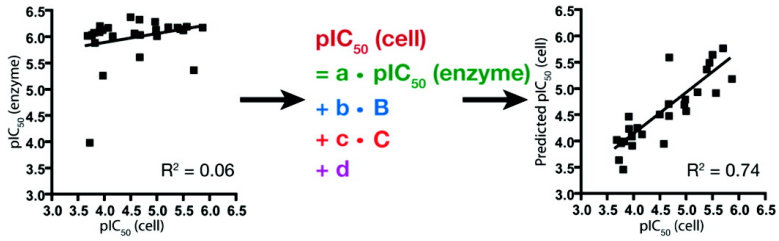
Journal of Medicinal Chemistry

Article

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

Dushyant Mukkamala, Joo Hwan No, Lauren M. Cass, Ting-Kai Chang, and Eric Oldfield

J. Med. Chem., **2008**, 51 (24), 7827-7833 • DOI: 10.1021/jm8009074 • Publication Date (Web): 21 November 2008



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

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

Dushyant Mukkamala,[†,‡] Joo Hwan No,[†,‡] Lauren M. Cass,[§] Ting-Kai Chang,[||] and Eric Oldfield*[,†,||]

Center for Biophysics and Computational Biology, University of Illinois at Urbana−Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, School of Molecular and Cellular Biology, University of Illinois at Urbana−Champaign, 505 South Goodwin Avenue, Urbana, Illinois 61801, and Department of Chemistry, University of Illinois at Urbana−Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801

Received July 21, 2008

We screened 26 bisphosphonates against a farnesyl diphosphate synthase from *Plasmodium vivax*, finding a poor correlation between enzyme and cell growth inhibition ($R^2 = 0.06$). To better predict cell activity data, we then used a combinatorial descriptor search in which pIC_{50}(cell) $= a$ pIC_{50}(enzyme) $+ bB + cC + d$, where B and C are descriptors (such as SlogP), and $a-d$ are coefficients. R^2 increased from 0.01 to 0.74 (for a leave-two-out test set of 26 predictions). The method was then further validated using data for nine other systems, including bacterial, viral, and mammalian cell systems. On average, experimental/predicted cell pIC_{50} correlations increased from $R^2 = 0.28$ (for an enzyme-only test set) to 0.70 (for enzyme plus two descriptor test set predictions), while predictions based on scrambled cell activity had no predictive value ($R^2 = 0.13$). These results are of interest since they represent a general way to predict cell from enzyme inhibition data, with in three cases, R^2 values increasing from ~0.02 to 0.72.

Introduction

Malaria causes $\sim 10^6$ deaths annually, so there is considerable interest in developing novel drugs to treat this disease. One pathway of interest is that involved in the isoprenylation of signaling proteins, where protein-farnesyl transferase inhibitors have shown promise,[1] since they act by preventing the post-translational modification of proteins such as Ras. A second potential target involves inhibition of the enzymes (such as farnesyl diphosphate synthase, FPPS[a]) that produce the isoprene diphosphates used in protein prenylation, and in recent work we showed that bisphosphonates, drugs that target FPPS and are widely used in treating bone resorption diseases,[2] had both in vitro and in vivo activity against *Plasmodium* parasites.[3,4] There were, however, some puzzling aspects to the results;[4] in particular, the most active species against *Plasmodium falciparum* had little activity against a variety of expressed FPPS enzymes, while other known potent FPPS inhibitors had low activity against the parasite.

There are several possible explanations for these observations. First, it might be that FPPS is not the actual (or only) target. Second, there might be surprising differences in the sensitivity of *Plasmodium* FPPS and human (or other) FPPS enzymes, since *Plasmodium* FPPS can also produce geranylgeranyl diphosphate (GGPP), due to a smaller barrier to chain elongation at the end of the binding site. Third, it might just be difficult to obtain good correlations between enzyme and cell inhibition data, due to neglect of "transport" issues. This latter point is an exceptionally important one[5] and is of very broad general interest for drug development, since while enzyme inhibition assays can generally be carried out very rapidly (and accurately), cell based assays are more expensive and take much longer. Indeed, in some cases, cell data may be essentially uncorrelated with enzyme inhibition. A good recent example of this is that of the inhibition of undecaprenyl diphosphate synthase from *Streptococcus pneumoniae* reported by Peukert et al.,[6] in which an $R^2 = 0.03$ is found from the reported cell/enzyme pIC_{50} ($= -\log_{10} IC_{50}$) results, obviously complicating lead optimization using enzyme based assays.

In this work, we report the inhibition of a *Plasmodium* FPPS (from *Plasmodium vivax*) by a library of 26 bisphosphonates whose activity against *P. falciparum* cell growth in vitro was reported previously.[4] The *P. falciparum* and *P. vivax* enzymes have very similar sequences (73% identity, 89% similarity) and catalytic site residues, but we find that there is essentially no correlation between the cell and enzyme pIC_{50} values ($R^2 = 0.06$). We find, however, that regression of cell, enzyme, and two other descriptors leads to a remarkable improvement in the correlation between experimental and predicted cell pIC_{50} (to $R^2 = 0.74$), suggesting that small descriptor sets might be used to predict cell activity results in many other systems, nine of which (including other antiparasitic, antibacterial, antiviral, antidiabetic, and anticancer drug leads) are investigated in this work. Overall, our results show that cell activity in 10 very varied systems can be well predicted by using enzyme inhibition data, when combined with the combinatorial descriptor search approach.

Results and Discussion

We show in Figure 1 the structures of the 26 bisphosphonates (**1**−**26**) investigated previously.[4] These compounds have IC_{50} values in *P. falciparum* growth inhibition varying from 1.43 to 222 μM[4] with the most active species also having in vivo activity in a *Plasmodium berghei* ANKA suppressive test (up to an 80% reduction in parasitemia). The IC_{50}/pIC_{50} values in FPPS inhibition are presented in Table 1, together with the

* To whom correspondence should be addressed. Phone: (217) 333-3374. Fax: (217) 244- 0997. E-mail: eo@chad.scs.uiuc.edu.
† Center for Biophysics and Computational Biology.
‡ These authors contributed equally to this work.
§ School of Molecular and Cellular Biology.
|| Department of Chemistry.

[a] Abbreviations: FPPS, farnesyl diphosphate synthase; GGPP, geranylgeranyl diphosphate; L2O, leave-two-out; UPPS, undecaprenyl diphosphate synthase; PAMPA, parallel artificial membrane permeability analysis; HIV, human immunodeficiency virus; HCV, hepatitis C virus; hlGP, human liver glycogen phosphorylase; QSAR, quantitative structure activity relationship; GPP, geranyl diphosphate; IPP, isopentenyl diphosphate.

10.1021/jm8009074 CCC: $40.75
Published on Web 11/21/2008

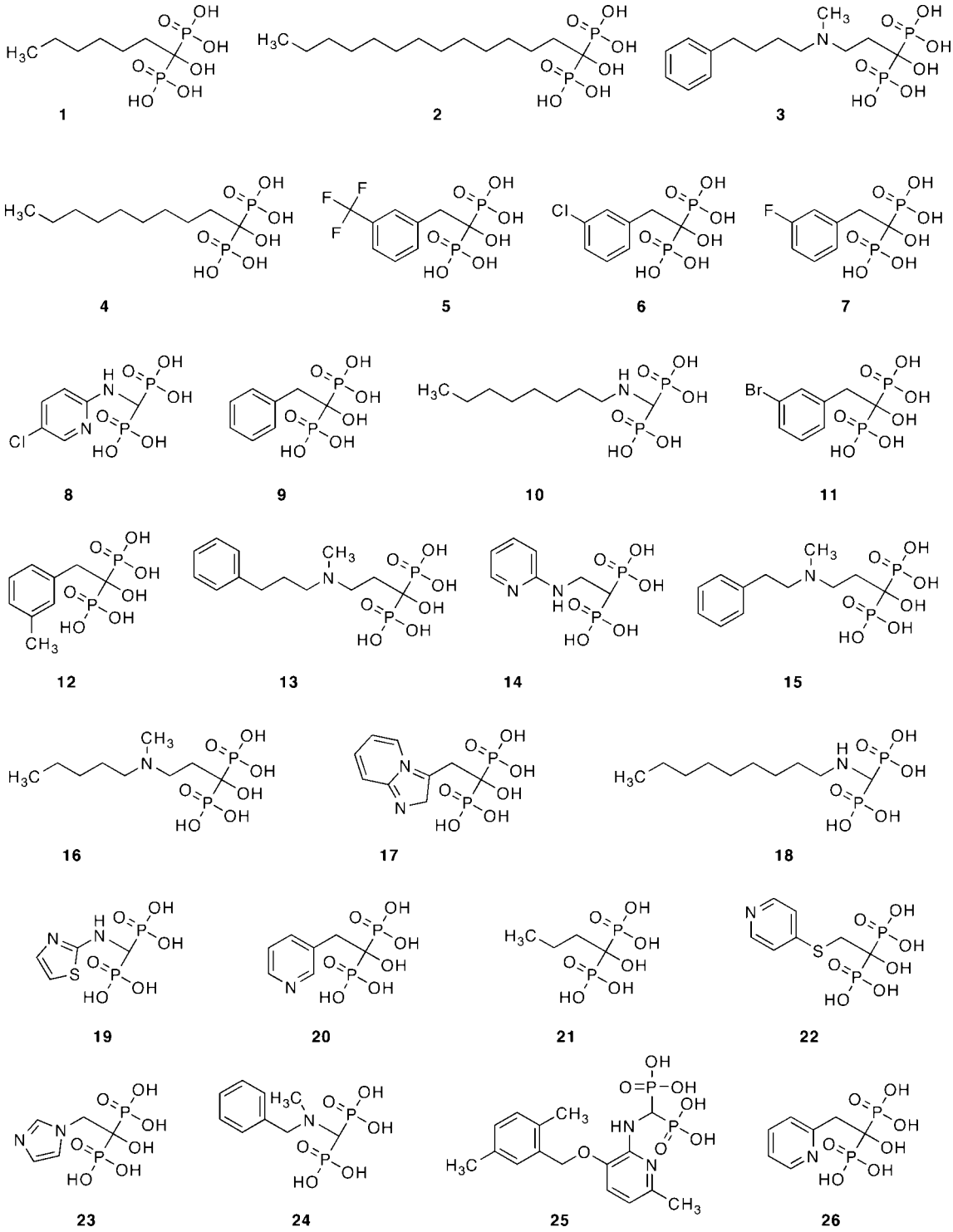


Figure 1. Structures of the 26 compounds investigated in cell (*Plasmodium falciparum*) and enzyme (*Plasmodium vivax* FPPS) assays.

previously published cell growth inhibition results.[4] When the cell pIC_{50} and enzyme pIC_{50} values are compared, we see that there is essentially no correlation between the two data sets ($R^2 = 0.06$), Figure 2a, resulting in a leave-2-out set of predictions having an $R^2 = 0.01$. For example, the most active species in FPPS inhibition is **15** ($IC_{50} = 400$ nM), but this compound has poor activity in vitro (33.6 μM, Table 1). It might therefore be thought that FPPS is not the actual (or only) target for these bisphosphonates or that there are small but significant differences in structure between the *P. falciparum* and *P. vivax* enzymes. However, on further inspection of Figure 2a, there appear to be several possible "clusters" of compounds (**A**, **B**, and **C** in Figure 2a), with the more active species in cells having more hydrophobic features. For example, in cluster **A**, known bisphosphonates such as risedronate (**20**), ibandronate (**16**), zoledronate (**23**), and minodronate (**17**) all have very poor (>70 μM) cell activity (Table 1), even though they have good activity in the enzyme assay. Since these compounds are among the most hydrophilic ones investigated (with an average SlogP = −5.00, for nine compounds in this cluster), it is possible that their poor cell based activity is due to poor transport. This idea receives support from the observation that a second class of bisphosphonates (such as **13**, **15**; cluster **B**, with average SlogP = −4.60, for five compounds) that have improved hydrophobic-

Table 1. Experimental IC_{50} and pIC_{50} Values for *P. falciparum* Growth Inhibition and *P. vivax* FPPS Inhibition

compound[a]	IC_{50} (enzyme, μM)	pIC_{50} (enzyme)[b]	IC_{50} (cell, μM)	pIC_{50} (cell)[c]
1	0.63	6.2	1.4	5.9
2	4.1	5.4	2.1	5.7
3	0.61	6.2	2.9	5.5
4	0.71	6.1	3.3	5.5
5	0.68	6.2	3.8	5.4
6	0.63	6.2	4.3	5.4
7	0.63	6.2	6.4	5.2
8	0.91	6.0	11	5.0
9	0.69	6.2	11	5.0
10	0.49	6.3	12	4.9
11	0.88	6.1	22	4.7
12	2.3	5.6	23	4.6
13	0.45	6.3	23	4.6
14	0.82	6.1	28	4.6
15	0.40	6.4	34	4.5
16	0.92	6.0	74	4.1
17	0.64	6.2	90	4.0
18	5.2	5.3	110	4.0
19	0.71	6.1	120	3.9
20	0.59	6.2	130	3.9
21	0.78	6.1	130	3.9
22	1.2	5.9	160	3.8
23	0.79	6.1	170	3.8
24	0.87	6.1	180	3.7
25	98	4.0	200	3.7
26	0.90	6.0	220	3.7

[a] See Figure 1 for chemical structures. [b] The enzyme assays were performed on FPPS from *Plasmodium vivax*. [c] The cell assay data is for *Plasmodium falciparum*.[4]

ity also exhibit improved cell activity, and compounds which have the highest cell-based activity (**1**−**6**, Figure 1, cluster **C**) also contain the most hydrophobic side chains (average SlogP = −3.60, for 10 compounds). How can we put these qualitative observations on a more quantitative footing? Can we predict cell activity from the enzyme inhibition results? Also, if this turns out to be possible in the *Plasmodium* system, can the method be generalized to other systems, something that would be of very broad, general interest?

The results shown in Table 1 and Figure 2a clearly indicate that there is essentially no correlation between the cell pIC_{50} and enzyme pIC_{50} values. Why is this? While, as noted above, there are several possibilities (other targets, poor *P. vivax*/*P. falciparum* similarity), the observation that cell based activity appears to increase as hydrophobicity increases suggests the possibility that cell permeability may be particularly important in governing overall cell activity. To test this hypothesis, we chose to represent the cell based activity (pIC_{50} (cell)) mathematically as

$$pIC_{50}(\text{cell}) = a\ pIC_{50}(\text{enzyme}) + b\ \text{SlogP} + c \qquad (1)$$

where SlogP is a computed oil/water partition coefficient.[7] We chose SlogP (rather than say clogP, etc.) without any particular basis, a point we return to shortly. Data were analyzed by using linear regression, from which we find $a = 1.26$, $b = 0.62$, and $c = -0.34$. Now, when the pIC_{50} (cell, experiment) results are correlated with the predicted pIC_{50}(cell) results, we find a good correlation between experiment and prediction, with an $R^2 = 0.66$. Given this promising result, we then sought to find other descriptors and descriptor combinations that might give improved predictivity, rather than restricting ourselves to SlogP. In some initial calculations we used enzyme pIC_{50} values together with three descriptors, or four random descriptors chosen from MOE,[7] that gave the best experimental versus predicted R^2 values. However, this approach was found to lead

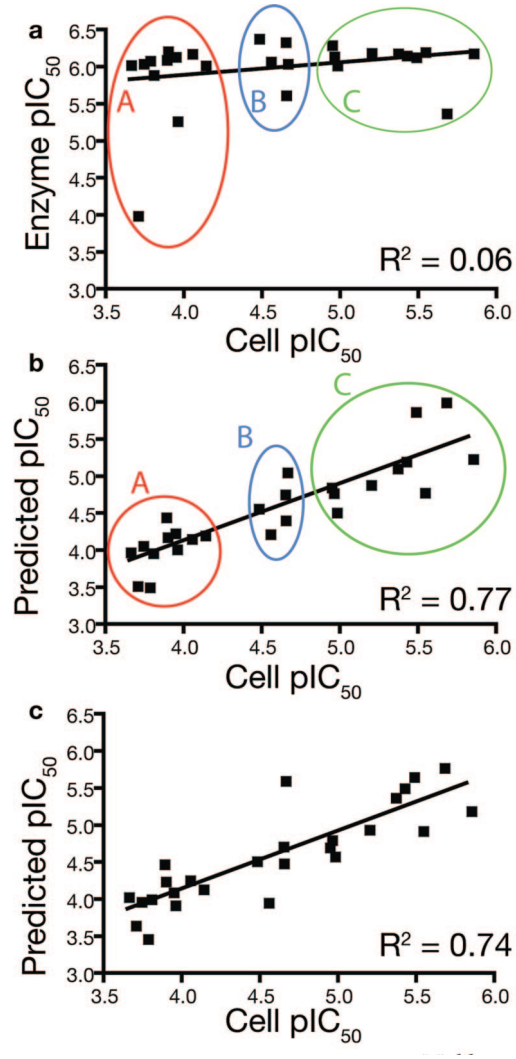


Figure 2. Correlation plots for cell (*Plasmodium falciparum*) and enzyme (*Plasmodium vivax* FPPS) assays and predicted cell activities from the training and test set data, obtained by using the combinatorial descriptor search method. (a) Plot showing correlation between cell pIC_{50} ($= -\log_{10}(IC_{50})$) and enzyme pIC_{50} values. (b) Best correlation between predicted cell pIC_{50} (enzyme plus two molecular descriptors) and experimental pIC_{50} values: training set results. (c) Test set pIC_{50} predictions (leave-two-out analysis) plotted against the experimental values. The R^2 value increases from ∼0 to 0.74, when adding the two molecular descriptors to the enzyme data. The colored circles (**A**−**C**) indicate bisphosphonates with side chains having different relative hydrophobicities and potencies in the cell assay (low, intermediate or high, respectively), as discussed in the text.

to overfitting, since some predictions using randomized cell activity data had R^2 values of ∼0.5. We thus eventually chose the following equation:

$$pIC_{50}(\text{cell}) = a\ pIC_{50}(\text{enzyme}) + bB + cC + d \qquad (2)$$

in which two additional descriptors (B and C) were chosen from a combinatorial search, performed on a large set of descriptors (Supporting Information, Table S1) in MOE.[7] We chose to fix pIC_{50} (enzyme) as one descriptor, since the objective is to predict cell activity from enzyme activity, where standard QSAR and structure-based methods can be employed to optimize enzyme inhibition. There are 230 descriptors available in MOE, and a full search and cross-validation of all (26 335) two-descriptor combinations for a data set containing 26 compounds is lengthy (∼30 h). However, we found that it was not necessary to use

Table 2. Top 10 "Enzyme Plus 2-Descriptor" Combinations with Coefficients and Relative Contributions for *P. falciparum* Growth Inhibition Predictions

rank	R^2	coefficient a^a (enzyme)	relative contribution (enzyme)	coefficient b^a	descriptor B^b	relative contribution (B)	coefficient c^a	descriptor C^b	relative contribution (C)	constant, d^a
1	0.77	1.0407	0.72	0.52375	logP(o/w)	1.00	−0.0028993	E_sol	0.47	−3.2113
2	0.76	1.3081	0.84	−0.5824	logS	1.00	4.6753	E_oop	0.42	−3.7232
3	0.76	1.0806	0.69	0.56454	logP(o/w)	1.00	4.9535	GCUT_SMR_0	0.45	1.4987
4	0.75	1.3317	1.00	0.42892	SlogP	0.75	0.011375	PEOE_VSA_NEG	0.50	−3.7753
5	0.75	0.98902	0.70	0.51324	logP(o/w)	1.00	−292.05	BCUT_PEOE_1	0.44	−203.38
6	0.75	1.2344	0.85	0.62848	SlogP	1.00	1.0862	E_strain	0.30	−0.23834
7	0.75	1.342	1.00	0.5162	SlogP	0.89	0.17337	chi1_C	0.40	−1.8939
8	0.75	1.2085	1.00	0.49972	SlogP	0.96	1.8143	GCUT_SLOGP_1	0.42	0.49546
9	0.75	1.044	0.74	0.51263	logP(o/w)	1.00	0.86654	BCUT_SLOGP_0	0.44	1.5301
10	0.74	1.0733	1.00	0.44935	SlogP	0.97	0.0098483	PEOE_VSA-1	0.49	−0.61185

[a] From the equation, pIC_{50} (cell) = a pIC_{50} (enzyme) + bB + cC + d. [b] Descriptors selected are described in Supporting Information, Table S1.

all the 230 descriptors, for two reasons: first, because some descriptors are linear combinations of other descriptors in the database, and second, because some descriptors are Boolean (with predominantly either zeros or ones) and do not contribute significantly in the linear regression. In the set of 26 *P. vivax* FPPS inhibitors, after eliminating redundant and Boolean descriptors, we obtained 150 descriptors, of which all combinations of two (D1, D2; D1, D3;. . . ; D1, D150; D2, D3; D2, D4;. . . ; D2, D150; D3, D4;D3, D5;. . . ; D3, D150;. . . ; D148, D149; D148, 150; 149, 150) were investigated, the coefficients (a. . . d) being determined via linear regression. The 10 descriptor combinations (pIC_{50}; B_i; C_j) giving the highest R^2 values (for the training set experimental versus predicted pIC_{50} correlation) are shown in Table 2.

As can be seen in Table 2, the best experimental/predicted pIC_{50} correlation has an $R^2 = 0.77$ (Figure 2b), a major increase over the cell/enzyme correlation ($R^2 = 0.06$) alone. For the top 10 predictions, $^9/_{10}$ contain a SlogP or logP(o/w) term, indicating the importance of hydrophobicity in cell based activity. Also, in $^6/_{10}$ cases, this SlogP or logP(o/w) term has the largest relative contribution to the correlation (Table 2). In the four cases where this term does not dominate, enzyme activity dominates, but as can be seen in Table 2, both enzyme pIC_{50} and logP-like terms are important in all cases, and the R^2 range is very small, 0.74−0.77. So, both enzyme and "logP" contribute to cell activity, with the cell activity results apparently being well predicted based on the enzyme inhibition and two-descriptor model. Of course, considerable care needs to be taken in order to not overfit data when using this approach. We tested this in several ways: First, we used a leave-two-out (L2O) method in which we systematically left out all possible pairs of data points ($^{26}C_2 = 26 \times 25/2 = 325$ calculations), re-evaluating the $a-d$ coefficients for all possible B and C descriptors (total time ~12 h). The results of this L2O test set of predictions are shown graphically in Figure 2c, where we find an $R^2 = 0.74$ for the experimental/predicted pIC_{50} test set correlation, close to the $R^2 = 0.77$ for the training set (Figure 2b). Second, we scrambled the cell pIC_{50} values and repeated the L2O test set prediction, 10 times. On average, the R^2 value was 0.16 for this scrambled test set, strongly supporting the idea that the approach, when applied to real cell growth inhibition data, is highly predictive. Using larger numbers of descriptors increased the R^2 value for both the unscrambled and scrambled L2O data sets and is not recommended, at least for relatively small data sets.

These results encouraged us to see to what extent this approach might enable us to improve cell activity predictions, based on enzyme activity and the combinatorial descriptor search, for a variety of other systems. We first chose to investigate two other systems we reported on previously: *Dictyostelium discoideum*[8] (a eukaryote used as a screen for bone resorption drugs that target FPPS) and *Leishmania donovani*,[3,9] a causative agent of visceral leishmaniasis. Both systems (or very close relatives) have been shown by knockouts or overexpression to require FPPS and that FPPS is the target for bisphosphonate drugs or inhibitors.

For *Dictyostelium discoideum* growth inhibition,[8] the cell/enzyme R^2 value is 0.49 (Figure 3a) or $R^2 = 0.46$ for a L2O enzyme only test set, and this increases to $R^2 = 0.70$ for the training set, using two additional descriptors (Figure 3b and Supporting Information, Table S2), with pIC_{50} (enzyme) being the most important contributor to the correlation. This is not unexpected since, unlike the *Plasmodium* results, we do see a correlation between cell and enzyme data, Figure 3a. The leave-two-out test set results (Table 3, Figure 3c) are likewise good ($R^2 = 0.70$) and as with the *Plasmodium* results, the scrambled data set has essentially no predictivity ($R^2 = 0.07$). Slightly improved results are obtained for *Leishmania donovani* cell growth inhibition. Here, the raw cell/enzyme correlation is $R^2 = 0.55$ (Table 3, Figure 3d), translating to a L2O enzyme-based test set $R^2 = 0.47$. The experimental /predicted cell pIC_{50} correlation increases to $R^2 = 0.87$ for the training set using two additional descriptors (Figure 3e and Supporting Information, Table S3) and 0.80 for the L2O test set results (Figure 3f), while using scrambled cell data yields $R^2 = 0.11$ (Table 3). We should also note here that in both *Dicytostelium discoideum* and *L. donovani*, FPPS inhibition results for the specific organisms were not available so we used a composite data set of *Trypanosoma brucei*,[10] *Trypanosoma cruzi*,[11] *Leishmania major*,[12] and human[13] FPPS results to model the *D. discoideum* and *L. donovani* data, just as we used *P. vivax* data to interpret the *P. falciparum* results. Naturally, it is reasonable to believe that using, e.g., *L. donovani* FPPS to model *L. donovani* growth inhibition will be preferred to using, e.g., *L. major* FPPS inhibition data. However, the predictions are still good, with for *P. falciparum*, *D. discoideum*, and *L. donovani*, the average R^2 value increasing from 0.37 (cell/enzyme L2O test set predictions) to 0.75, using pIC_{50} (enzyme) plus two descriptors, while the average R^2 for the scrambled cell activity data sets is $R^2 = 0.11$. However, given that all of the inhibitors studied here are bisphosphonates and the targets protozoa or a unicellular eukaryote, it is fair to ask, is this method robust? Does it apply to other types of cell and inhibitor? Is it, in short, a general method for probing cell activity based on enzyme inhibition?

Antibacterial and Antiviral Systems. To test the generality of this approach, we investigated two bacterial and two viral systems. In each case, the enzyme targeted in the enzyme assay was the same as that targeted in the cell based assay. In one recent study, workers at Novartis[6] reported the discovery of a series of novel inhibitors (tetramic, tetronic acids, and dihydropyridin-2-ones) of the enzyme undecaprenyl diphosphate

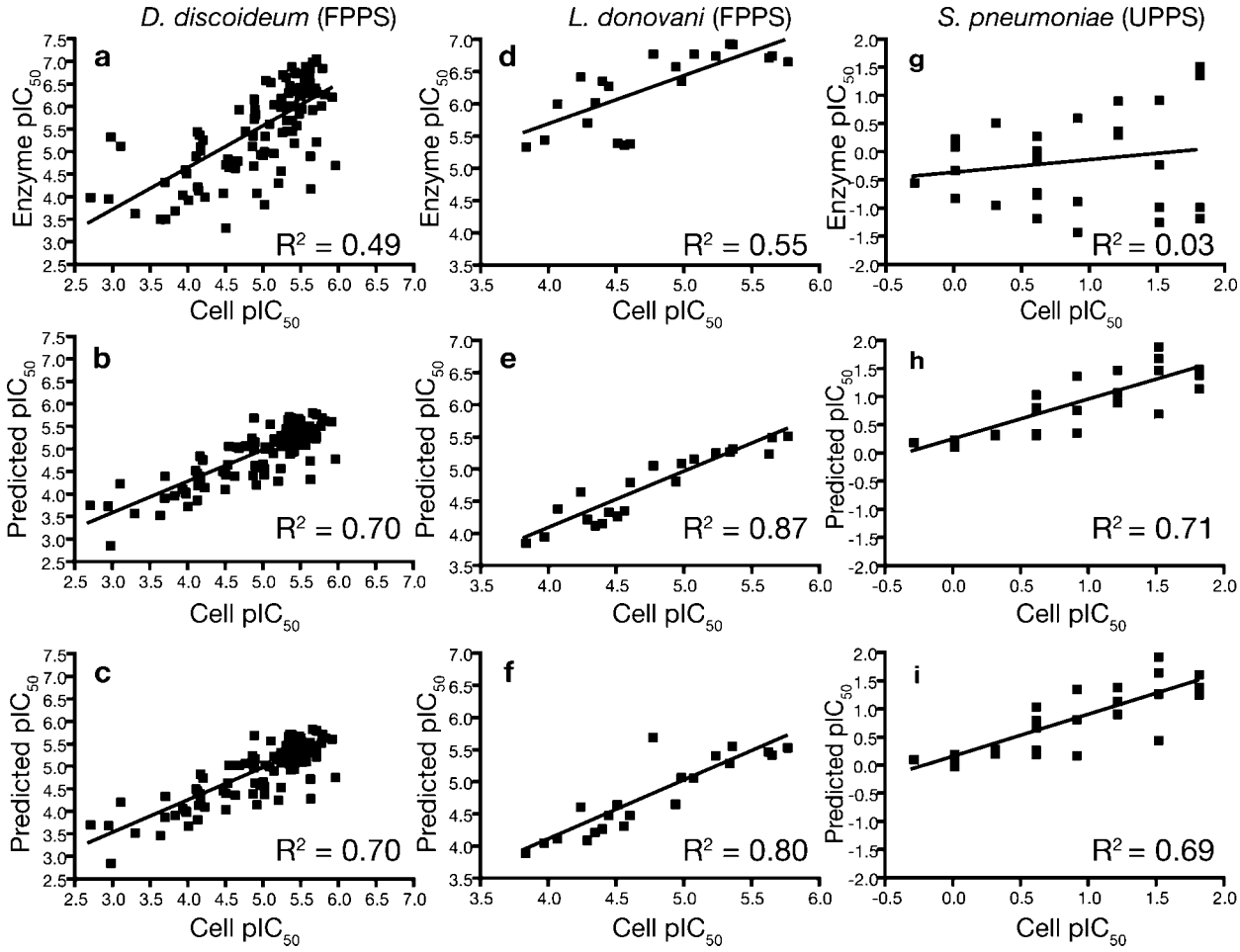


Figure 3. Correlation plots for the cell and enzyme assays, and predicted cell activities from the training and test set data in *Dictyostelium discoideum* (a−c), *Leishmania donovani* (d−f), and *Streptococcus pneumoniae* (g−i). (a) Plot showing the correlation between cell (*D. discoideum*) pIC_{50} and enzyme (FPPS) 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 (*L. donovani*) pIC_{50} and enzyme (FPPS) 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 (*S. pneumoniae*) pIC_{50} and enzyme (UPPS) 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.

Table 3. Results for 10 Different Systems Showing the Correlations between Enzyme and Cell Activities and the Correlations between Experimental and Predicted Cell Activities

target	cell	R^2, enzyme vs cell	R^2, enzyme test set[a]	R^2, training set[b]	R^2, test set[a]	R^2, scrambled[c]	descriptors tested	number of compounds
FPPS	*Plasmodium falciparum*	0.06	0.01	0.77	0.74	0.16	150	26
FPPS	*Dictyostelium discoideum*	0.49	0.46	0.70	0.70	0.07	123	102
FPPS	*Leishmania donovani*	0.55	0.47	0.87	0.80	0.11	138	21
UPPS	*Streptococcus pneumoniae*	0.03	0.10	0.71	0.69	0.14	109	27
MurI	*Streptococcus pneumoniae*	0.61	0.57	0.72	0.68	0.13	123	42
HCV NS3 Protease	Hu-7	0.03	0.02	0.78	0.77	0.12	142	34
HIV-1 integrase	MT4	0.12	0.04	0.69	0.64	0.13	126	30
HLG Phosphorylase	rat hepatocytes	0.55	0.48	0.71	0.74	0.19	132	35
KDR kinase	NIH 3T3	0.36	0.30	0.60	0.58	0.09	117	42
Akt Kinase	MiaPaCa-2 (pancreatic cancer)	0.41	0.35	0.68	0.68	0.12	80	48
average		0.32	0.28	0.72	0.70	0.13	124	42

[a] Correlation obtained from a leave-two-out analysis. [b] Correlations based on cell activity predictions using a combinatorial descriptor search. [c] Average correlation obtained from a leave-two-out analysis of 10 sets of scrambled data.

synthase (UPPS) that also had good cellular activity.[6] These inhibitors are of interest since UPPS is a potentially important target for anti-infective development, because undecaprenyl diphosphate is used in lipid A biosynthesis, and UPPS inhibition is expected to block bacterial cell wall biosynthesis. Also, the structures of some UPPS enzymes are now known.[14,15] *Streptococcus pneumoniae* enzyme and (approximate) cell growth inhibition results (inhibition at 1, 2, 4 μM, etc. inhibitor concentrations) were reported, but as with our *Plasmodium* results, there is little correlation between the cell growth and enzyme inhibition data sets ($R^2 = 0.03$; Figure 3g), even though in this case the enzyme target (*S. pneumoniae* UPPS) is thought to be exactly the same as that present in the bacterium, *Streptococcus pneumoniae*.[6] The small range in cell activity (128×) combined with the limited precision of the cell data (the IC_{50} values were rounded off by factors of 2) makes this a

challenging case. Even so, using the reported enzyme pIC_{50} results together with two descriptors yields a considerable improvement, with R^2 increasing from 0.10 for the L2O test set using enzyme data only, to $R^2 = 0.69$ (Figure 3i) for the L2O test set predictions using two extra descriptors, with R^2 for the L2O test set using scrambled data being 0.14, Table 3. As can be seen in the Supporting Information (Table S4), descriptors such as SlogP and logS (also found in the *Plasmodium* data sets) are again found here and, when combined with enzyme inhibition data, enable greatly improved predictions of the cell activity results. Another set of antibacterials are glutamic acid analogues that inhibit *S. pneumoniae* glutamate racemase (MurI).[16] In this case, there is a moderate correlation between cell growth and MurI inhibition ($R^2 = 0.61$), and this increases to 0.72 (training) or 0.68 (leave-two-out) test set predictions using the combinatorial descriptor search, Table 3 and Supporting Information, Figure S1 and Table S5, with the R^2 for scrambled cell activity test set data being 0.13.

In addition to these results with antibacterial systems, we find good predictivity in an antiviral assay using hepatitis C virus protease inhibitors.[17] Here, there is again a very poor correlation between enzyme and cell pIC_{50} values (obtained in a cell-based replicon assay) with $R^2 = 0.03$ (Supporting Information, Figure S1d). This increases to $R^2 = 0.78$ (training) and $R^2 = 0.77$ (leave-two-out) test set results (Supporting Information, parts e and f of Figure S1) with incorporation of two additional descriptors. In this case, the best results were obtained by using PAMPA (parallel artificial membrane permeability assay) as one of the descriptors (determined experimentally by Li et al.[17]). In these calculations, we added reported PAMPA, Caco-2, logP-neutral, logD-pH 7.4, clogP, and mlog P values to the 230 potential descriptor fields in MOE, with PAMPA being selected in the descriptor search as the most important descriptor, followed by the enzyme pK_i (Supporting Information, Table S6). So, the predictivity of the method can be improved when experimental data on permeability is available.

We also investigated a second series of antivirals, novel phthalimide-analogue inhibitors of HIV-1 integrase.[18] The results of enzyme pIC_{50}, HIV-1 replication in a cell-based assay (pEC_{50}), as well as cell-based toxicity (pCC_{50}) have been reported.[18] There is only a very poor correlation between the enzyme pIC_{50} and cell-based replication activity (pEC_{50}), with $R^2 = 0.12$ (Supporting Information, Figure S1g; $R^2 = 0.01$ for the L2O enzyme based test set). Using the combinatorial descriptor search, we find that pCC_{50} is the most important descriptor, enabling an experimental versus predicted pIC_{50} $R^2 = 0.69$ (Supporting Information, Figure S1h). That is, pCC_{50} makes the major contribution to the apparent cell (HIV) replication results (Supporting Information, Table S7), an observation of interest for further optimization of this class of compounds. Also on average, for the four antibacterial/antiviral systems, we see that R^2 increases from 0.20 to 0.70 (L2O test sets) on addition of two descriptors, about the same as that found for the protozoal/primitive eukaryote systems with scrambled cell data, $R^2 = 0.13$, about the same as the $R^2 = 0.11$ found with the protozoa/simple eukaryotes.

Mammalian Cell Lines: Antidiabetic and Anticancer Drugs. Finally, we wanted to test the method described above on three mammalian cell lines, in diabetes and cancer assays, as opposed to the primarily anti-infective drugs/targets described in the preceding sections. We investigated three systems: acyl urea inhibitors of human liver glycogen phosphorylase (hlGP)[19] having activity in rat hepatocytes; 1,4-dihydroindeno[1,2-c]pyrazoles targeting a KDR kinase in NIH3T3 cells;[20] and an indazole-pyridine series of protein kinase B/Akt inhibitors, active in pancreatic cancer cells.[21]

In the case of the acyl urea inhibitors of human liver glycogen phosphorylase, the enzyme/cell correlation was already evident ($R^2 = 0.55$), and this improved to $R^2 = 0.74$ for the L2O test set of 35 compounds, Table 3 and Supporting Information, Figure S2a−c and Table S8. The scrambled data was, as expected, not predictive, with an $R^2 = 0.19$. Results for the KDR kinase (Table 3 and Supporting Information, Figure S2d−f and Table S9) were slightly worse (L2O test set $R^2 = 0.58$), while results with the Akt kinase (Table 3 and Supporting Information, Figure S2g−i) were good, with an $R^2 = 0.68$ for the test set (Supporting Information, Table S10).

Conclusions

The results described above are of interest for several reasons. First, we screened a library of bisphosphonates against a *P. vivax* FPPS enzyme. The results were surprising since we found that inhibition of the *P. vivax* enzyme was essentially uncorrelated with growth inhibition by these same bisphosphonates of the malaria parasite, *P. falciparum*. However, when pIC_{50} results together with two additional descriptors were regressed against cell growth inhibition pIC_{50} results, we found good predictivity of the cell results, with an $R^2 = 0.77$ (0.74 for a leave-two-out test set). This observation then led to a general method for predicting the cell-based activity of a compound from its activity in an enzyme assay, in which we use a combinatorial descriptor search to choose two descriptors (*B* and *C*) that, in combination with the enzyme pIC_{50} results, enabled relatively good cell activity predictions. For the 10 diverse systems investigated here, which cover antiparasitic, antibacterial, antiviral, antidiabetes, and anticancer drug candidates, the average R^2 value increases from 0.28 for the leave-two-out test set predictions using solely enzyme inhibition data to 0.70 for the leave two-out test sets with two additional descriptors, to be compared with 0.13 for predictions based on randomized cell activity data. This approach should be particularly useful in optimizing cell activity when target structures are known, since structure-based (X-ray, QSAR) methods can be used to optimize enzyme inhibition, with the combinatorial descriptor search then being used to make good predictions of cell based activity. Also, when experimental data on permeability is available, this can be included in the descriptor field as another potential descriptor, to be used in cell activity predictions. A larger number of descriptors than the (on average) 124 we employed can also of course be used, at the expense of computational time, which can become large (~300 h for a 230 descriptor based series of calculations, including validation tests). The approach described above should thus enable the use of both crystallographic and enzyme QSAR data for enzyme inhibition optimization, with cell growth inhibition (or viral replication) activity being described using the combinatorial descriptor approach. The method should also be applicable in some cases to modeling toxicity and in vivo results as well, when sufficient data is available.

Experimental Section

***P. vivax* FPPS Expression and Inhibition.** A clone encoding *P. vivax* FPPS (PlasmoDB gene ID: Pv092040) with an N-terminally His_6-tagged fusion protein and a TEV protease site was expressed in *Escherichia coli* BL21-codon Plus (DE3) RIL (Stratagene) at 20 °C in baffled flasks. Cells were lysed in the presence of lysozyme (Sigma), Benzonase Nuclease (Novagen), and a protease inhibitor cocktail (Sigma), and protein purified chromatographically by using a Ni-NTA resin. The *P. vivax* FPPS assays were carried out by using 96-well plates with 150 μL of reaction mixture in each well. The condensation of geranyl diphosphate (GPP) with isopentenyl diphosphate (IPP) was monitored at room temperature by using a continuous spectrophotometric assay for

phosphate releasing enzymes in a reaction mixture containing 25 mM Hepes and 2.5 mM $MgCl_2$ at pH 7.4. The inhibitors were preincubated with the enzyme for 30 min at room temperature. The IC_{50} values were obtained from fitting the dose−response curve using Prism 4.0.[22]

Computational Aspects. We performed a complete combinatorial descriptor search in MATLAB[23] (running on a dual 2 GHz PowerPC G5 or a Dual-Core Intel Xeon Mac Pro), using linear regression of enzyme pIC_{50} plus all possible two-descriptor combinations ($^{150}C_2 = 11\,175$ combinations for a database containing 150 descriptors, for example, against the cell activities, pIC_{50}(cell). For the leave-two-out (L2O) cross-validation, the same exhaustive search was performed, using a training set obtained by leaving out two compounds from the initial data set. This process was repeated for all pairwise combinations of the compounds (e.g., for 26 compounds, 1,2; 1,3;. . . 0.1,26; 2,3; 2,4;. . . ; 24,25; 24,26; 25,26). So, if there are 26 compounds (as in *P. vivax*), then each compound is left out (and predicted) 25 times. In this way, for *P. vivax* FPPS we obtained 25 predicted values for each compound in the data set, and these were then averaged to give the overall leave-two-out predicted value, for each compound. To test whether the correlations obtained might occur by chance, we randomly scrambled the cell activities, then performed the leave-two-out validation on the scrambled data. The process was repeated 10 times, and the mean R^2 values are reported. For the 10 systems investigated, each scrambling analysis requires ∼10−90 h, depending upon the number of descriptors (80−150) and the number of compounds (21−102) in the data set. Linearly dependent descriptors were eliminated based on the occurrence of rank deficiency in the descriptor correlation matrix (in MATLAB[23]).

Acknowledgment. We thank Raymond Hui of the Structural Genomics Consortium (Toronto, Canada) for providing the *P. vivax* protein expression system. This work was supported by the United States Public Health Service (NIH Grant GM65307).

Supporting Information Available: List of descriptors, additional tables of enzyme plus 2-descriptor combinations, and correlation plots for the cell and enzyme assays. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Buckner, F. S.; Eastman, R. T.; Yokoyama, K.; Gelb, M. H.; Van Voorhis, W. C. Protein Farnesyl Transferase Inhibitors for the Treatment of Malaria and African trypanosomiasis. *Curr. Opin. Investig. Drugs* **2005**, *6* (8), 791–797.

(2) Dunford, J. E.; Kwaasi, A. A.; Rogers, M. J.; Barnett, B. L.; Ebetino, F. H.; Russell, R. G.; Oppermann, U.; Kavanagh, K. L. Structure-Activity Relationships Among the Nitrogen Containing Bisphosphonates in Clinical Use and Other Analogues: Time-Dependent Inhibition of Human Farnesyl Pyrophosphate Synthase. *J. Med. Chem.* **2008**, *51* (7), 2187–2195.

(3) Martin, M. B.; Grimley, J. S.; Lewis, J. C.; Heath, H. T., 3rd; Bailey, B. N.; Kendrick, H.; Yardley, V.; Caldera, A.; Lira, R.; Urbina, J. A.; Moreno, S. N.; Docampo, R.; Croft, S. L.; Oldfield, E. Bisphosphonates Inhibit the Growth of *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania donovani*, *Toxoplasma gondii*, and *Plasmodium falciparum*: A Potential Route to Chemotherapy. *J. Med. Chem.* **2001**, *44* (6), 909–916.

(4) Ghosh, S.; Chan, J. M. W.; Lea, C. R.; Meints, G. A.; Lewis, J. C.; Tovian, Z. S.; Flessner, R. M.; Loftus, T. C.; Bruchhaus, I.; Kendrick, H.; Croft, S. L.; Kemp, R. G.; Kobayashi, S.; Nozaki, T.; Oldfield, E. Effects of Bisphosphonates on the Growth of *Entamoeba histolytica* and *Plasmodium* Species in Vitro and in Vivo. *J. Med. Chem.* **2004**, *47* (1), 175–187.

(5) Dobson, P. D.; Kell, D. B. Carrier-Mediated Cellular Uptake of Pharmaceutical Drugs: An Exception or the Rule. *Nat. Rev. Drug Discov.* **2008**, *7* (3), 205–220.

(6) Peukert, S.; Sun, Y.; Zhang, R.; Hurley, B.; Sabio, M.; Shen, X.; Gray, C.; Dzink-Fox, J.; Tao, J.; Cebula, R.; Wattanasin, S. Design and Structure-Activity Relationships of Potent and Selective Inhibitors of Undecaprenyl Pyrophosphate Synthase (UPPS): Tetramic, Tetronic Acids and Dihydropyridin-2-ones. *Bioorg. Med. Chem. Lett.* **2008**, *18* (6), 1840–1844.

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

(8) Szabo, C. M.; Martin, M. B.; Oldfield, E. An Investigation of Bone Resorption and *Dictyostelium discoideum* Growth Inhibition by Bisphosphonate Drugs. *J. Med. Chem.* **2002**, *45* (14), 2894–2903.

(9) Yardley, V.; Khan, A. A.; Martin, M. B.; Slifer, T. R.; Araujo, F. G.; Moreno, S. N.; Docampo, R.; Croft, S. L.; Oldfield, E. *In Vivo* Activities of Farnesyl Pyrophosphate Synthase Inhibitors against *Leishmania donovani* and *Toxoplasma gondii*. *Antimicrob. Agents Chemother.* **2002**, *46* (3), 929–931.

(10) Montalvetti, A.; Fernandez, A.; Sanders, J. M.; Ghosh, S.; Van Brussel, E.; Oldfield, E.; Docampo, R. Farnesyl Pyrophosphate Synthase is an Essential Enzyme in *Trypanosoma brucei*. *In Vitro* RNA Interference and *in Vivo* Inhibition Studies. *J. Biol. Chem.* **2003**, *278* (19), 17075–17083.

(11) Montalvetti, A.; Bailey, B. N.; Martin, M. B.; Severin, G. W.; Oldfield, E.; Docampo, R. Bisphosphonates are Potent Inhibitors of *Trypanosoma cruzi* Farnesyl Pyrophosphate Synthase. *J. Biol. Chem.* **2001**, *276* (36), 33930–33937.

(12) Sanders, J. M.; Gomez, A. O.; Mao, J.; Meints, G. A.; Van Brussel, E. M.; Burzynska, A.; Kafarski, P.; Gonzalez-Pacanowska, D.; Oldfield, E. 3-D QSAR Investigations of the Inhibition of *Leishmania major* Farnesyl Pyrophosphate Synthase by Bisphosphonates. *J. Med. Chem.* **2003**, *46* (24), 5171–5183.

(13) Dunford, J. E.; Thompson, K.; Coxon, F. P.; Luckman, S. P.; Hahn, F. M.; Poulter, C. D.; Ebetino, F. H.; Rogers, M. J. Structure-Activity Relationships for Inhibition of Farnesyl Diphosphate Synthase *in Vitro* and Inhibition of Bone Resorption *in Vivo* by Nitrogen-Containing Bisphosphonates. *J. Pharmacol. Exp. Ther.* **2001**, *296* (2), 235–242.

(14) Fujihashi, M.; Zhang, Y. W.; Higuchi, Y.; Li, X. Y.; Koyama, T.; Miki, K. Crystal Structure of *cis*-Prenyl Chain Elongating Enzyme, Undecaprenyl Diphosphate Synthase. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (8), 4337–4342.

(15) Guo, R. T.; Cao, R.; Liang, P. H.; Ko, T. P.; Chang, T. H.; Hudock, M. P.; Jeng, W. Y.; Chen, C. K.; Zhang, Y.; Song, Y.; Kuo, C. J.; Yin, F.; Oldfield, E.; Wang, A. H. Bisphosphonates Target Multiple Sites in both *cis*- and *trans*-Prenyltransferases. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (24), 10022–10027.

(16) de Dios, A.; Prieto, L.; Martin, J. A.; Rubio, A.; Ezquerra, J.; Tebbe, M.; Lopez de Uralde, B.; Martin, J.; Sanchez, A.; LeTourneau, D. L.; McGee, J. E.; Boylan, C.; Parr, T. R., Jr.; Smith, M. C. 4-Substituted D-Glutamic Acid Analogues: The First Potent Inhibitors of Glutamate Racemase (MurI) Enzyme with Antibacterial Activity. *J. Med. Chem.* **2002**, *45* (20), 4559–4570.

(17) Li, C.; Nair, L.; Liu, T.; Li, F.; Pichardo, J.; Agrawal, S.; Chase, R.; Tong, X.; Uss, A. S.; Bogen, S.; Njoroge, F. G.; Morrison, R. A.; Cheng, K. C. Correlation between PAMPA Permeability and Cellular Activities of Hepatitis C Virus Protease Inhibitors. *Biochem. Pharmacol.* **2008**, *75* (5), 1186–1197.

(18) Verschueren, W. G.; Dierynck, I.; Amssoms, K. I.; Hu, L.; Boonants, P. M.; Pille, G. M.; Daeyaert, F. F.; Hertogs, K.; Surleraux, D. L.; Wigerinck, P. B. Design and Optimization of Tricyclic Phthalimide Analogues as Novel Inhibitors of HIV-1 Integrase. *J. Med. Chem.* **2005**, *48* (6), 1930–1940.

(19) Klabunde, T.; Wendt, K. U.; Kadereit, D.; Brachvogel, V.; Burger, H. J.; Herling, A. W.; Oikonomakos, N. G.; Kosmopoulou, M. N.; Schmoll, D.; Sarubbi, E.; von Roedern, E.; Schonafinger, K.; Defossa, E. Acyl Ureas as Human Liver Glycogen Phosphorylase Inhibitors for the Treatment of Type 2 Diabetes. *J. Med. Chem.* **2005**, *48* (20), 6178–6193.

(20) Dinges, J.; Albert, D. H.; Arnold, L. D.; Ashworth, K. L.; Akritopoulou-Zanze, I.; Bousquet, P. F.; Bouska, J. J.; Cunha, G. A.; Davidsen, S. K.; Diaz, G. J.; Djuric, S. W.; Gasiecki, A. F.; Gintant, G. A.; Gracias, V. J.; Harris, C. M.; Houseman, K. A.; Hutchins, C. W.; Johnson, E. F.; Li, H.; Marcotte, P. A.; Martin, R. L.; Michaelides, M. R.; Nyein, M.; Sowin, T. J.; Su, Z.; Tapang, P. H.; Xia, Z.; Zhang, H. Q. 1,4-Dihydroindeno[1,2-c]pyrazoles with Acetylenic Side Chains as Novel and Potent Multitargeted Receptor Tyrosine Kinase Inhibitors with Low Affinity for the hERG Ion Channel. *J. Med. Chem.* **2007**, *50* (9), 2011–2029.

(21) Zhu, G. D.; Gandhi, V. B.; Gong, J.; Thomas, S.; Woods, K. W.; Song, X.; Li, T.; Diebold, R. B.; Luo, Y.; Liu, X.; Guan, R.; Klinghofer, V.; Johnson, E. F.; Bouska, J.; Olson, A.; Marsh, K. C.; Stoll, V. S.; Mamo, M.; Polakowski, J.; Campbell, T. J.; Martin, R. L.; Gintant, G. A.; Penning, T. D.; Li, Q.; Rosenberg, S. H.; Giranda, V. L. Syntheses of Potent, Selective, and Orally Bioavailable Indazole-Pyridine Series of Protein Kinase B/Akt Inhibitors with Reduced Hypotension. *J. Med. Chem.* **2007**, *50* (13), 2990–3003.

(22) *GraphPad Prism*, version 4.0c; GraphPad Software, Inc.: La Jolla, CA, 2005.

(23) *MATLAB*, version 7.4; The Mathworks, Inc.:, Natick, MA, 2007.

JM8009074