A Quantitative Structure−Activity Relationship and Pharmacophore Modeling Investigation of Aryl-X and Heterocyclic Bisphosphonates as Bone Resorption Agents

Evangelia Kotsikorou and Eric Oldfield*

Department of Chemistry, University of Illinois at Urbana−Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801

Received February 3, 2003

We have used quantitative structure−activity relationship (QSAR) techniques, together with pharmacophore modeling, to investigate the relationships between the structures of a wide variety of geminal bisphosphonates and their activity in inhibiting osteoclastic bone resorption. For aryl-X (X = alkyl, oxyalkyl, and sulfanylalkyl) derivatives of pamidronate and one alendronate, a molecular field analysis (MFA) yielded an R^2 value of 0.900 and an F-test of 54 for a training set of 29 compounds. Using reduced training sets, the activities of 20 such compounds were predicted with an average error of 2.1 over a 4000× range in activity. Such good results were only obtained when using the X-ray crystallographic structure of farnesyl pyrophosphate (FPP) bound to the target enzyme, farnesyl pyrophosphate synthase (FPP synthase), to guide the initial molecular alignment. For a series of heterocyclic bisphosphonates, use of the MFA method yielded an R^2 of 0.873 and an F-test of 36 for a training set of 26 compounds. Using a reduced training set, the activities of 20 compounds were predicted with an average error of 2.5 over a 2000× range in activity. With the heterocyclic compounds, test calculations indicated the importance of correct choice of protonation of the heterocyclic rings. For example, thiazoles, pyrazoles, and triazoles have low (~2−3) pK_a values and the derived bisphosphonates are inactive in bone resorption since they cannot readily be side chain protonated and are thus poor carbocation reactive intermediate analogues. On the other hand, aminothiazoles, imidazoles, pyridyl, and aminopyridyl species typically have pK_a values in the range ~5−9 and, in the absence of unfavorable steric interactions, the corresponding bisphosphonates are generally good inhibitors. However, aminoimidazole bisphosphonates are generally less active, since their pK_as (~11) are so high, due to guanidinium-like resonance, that they cannot readily be deprotonated, which we propose results in poor cellular uptake. The results of pharmacophore modeling using the Catalyst program revealed the importance of two negative ionizable and one positive charge feature for both aryl-X and heterocyclic pharmacophores, together with the presence of a distal hydrophobic feature in the aryl bisphosphonate and a more proximal aromatic feature in the heterocyclic bisphosphonate pharmacophores. When taken together, these results show that it is now possible to predict the activity, within a factor of about 2.3, of a wide range of aryl-X and heterocyclic bisphosphonates. The results emphasize the importance of utilizing crystallographic structural information to guide the initial alignment of extended bisphosphonates, and in the case of heterocyclic bisphosphonates, the importance of side chain protonation state. These simple ideas may facilitate the design of other, novel bisphosphonates, of use in bone resorption therapy, and as antiparasitic and immunotherapeutic agents.

Introduction

There is currently considerable interest in the development of geminal bisphosphonates for the treatment of tumor-induced hypercalcemia, Paget's disease of bone, osteolytic metastases, and postmenopausal osteoporosis,[1] with bisphosphonates making an ~$4−5 billion contribution to the global pharmaceutical market.[2] In addition to treating these bone related disorders, bisphosphonates have also been found to have considerable antiparasitic activity[3] and to be immunomodulators.[4−6] For example, bisphosphonates inhibit the growth of many trypanosomatid and apicomplexan parasites,[3] and parasitological cures of cutaneous leishmaniasis (caused by the protozoan *Leishmania mexicana*) and visceral leishmaniasis (caused by *Leishmania donovani*) have

been obtained in mice using the bisphosphonates pamidronate (Aredia) and risedronate (Actonel), respectively.[7,8] The bisphosphonates pamidronate, alendronate, ibandronate, risedronate, and zoledronate have also been shown to be potent stimulators of $\gamma\delta$ T cells,[4−6,9] and it has been proposed that this T cell activation may represent a new approach to the development of antibiotic compounds[9] and potentially the development of novel anticancer drugs.[6,10,11] There is, therefore, considerable interest in understanding the mode of action of these compounds.

In early work, it was thought that bisphosphonates acted by directly modulating bone mineral dissolution.[12] However, more recently, it has been shown that bisphosphonates act in the mevalonate pathway,[13] and very recently it has been demonstrated that the most potent, nitrogen-containing bisphosphonates (such as alendronate, risedronate, and zoledronate) act by in-

* To whom correspondence should be addressed. Phone: (217) 333−3374. Fax: (217) 244−0997. E-mail: eo@chad.scs.uiuc.edu.

hibiting the mevalonate/isoprene pathway enzyme farnesyl pyrophosphate synthase.[14-18] In so doing, they prevent the post-translation modification (prenylation) of proteins such as rho, in osteoclasts, which leads to decreases in osteoclastic bone resorption.[19] Of particular importance here has been the observation that there is an excellent correlation between the activity of a given bisphosphonate in inhibiting bone resorption and its activity in inhibiting the farnesyl pyrophosphate synthase enzyme.[18] On the basis of the likely mode of action of farnesyl pyrophosphate synthase, it appears that the most potent, nitrogen-containing bisphosphonates act as aza-analogues of one or more putative isoprene pyrophosphate carbocation reactive intermediates,[20] a view supported by several recent molecular modeling studies.[21,22]

In our group, we have been interested in using quantitative techniques to investigate the relationships between the three-dimensional structures of bisphosphonates (and related systems) and their activities in inhibiting farnesyl pyrophosphate synthases,[23] in inhibiting the growth of parasitic protozoa,[24] in $\gamma\delta$ T cell stimulation,[25] and in bone resorption.[26] In this paper, we focus on predicting the activities of a wide variety of novel aryl-X (X = alkyl, oxyalkyl, and sulfanylalkyl) substituted pamidronates, one alendronate, and a series of five-membered heterocyclic aromatic ring-containing bisphosphonates, whose activities in inhibiting bone resorption have recently been reported by the Novartis group.[1]

Results and Discussion

In recent work, Widler et al. reported the activities of a large number of aryl-X and heterocyclic bisphosphonates.[1] However, there have been no published, quantitative analyses of these results. In earlier work, we used QSAR techniques to investigate the activities of variety of bisphosphonate bone resorption drugs developed, primarily, by the Procter and Gamble Company, and we found that it was possible to predict their activities within about a factor of 3.[26] Here, we report an analysis of the bone resorption activity of a larger range of compounds, consisting of arylalkyl, aryloxyalkyl, arylsulfanylalkyl, and heterocyclic (imidazole, thiazole, triazole, and pyrazole) bisphosphonates, which display a broad range of structural diversity and activity.

Initially, we carried out a quantitative structure–activity relationship analysis of 86 compounds reported by Widler et al.[1] to have specific (i.e., not greater than) activities, using the molecular field analysis (MFA) method.[27] For the alignment of the side chains of these compounds, we used the crystal structure of geranyl pyrophosphate (GPP) bound to farnesyl pyrophosphate synthase,[28] and for the positions of the nitrogens we used the crystal structure of risedronate, aligned to geranyl pyrophosphate. After some adjustments to achieve the best alignment for all compounds, we obtained an R^2 of 0.558, a cross-validation R^2_{cv} of 0.491 and an F-test of 25.6. These results indicated a poor correlation between the experimental activity and that predicted using the MFA method. Consequently, we decided to focus our study on the two major groups of compounds investigated by Widler et al.: the aryl-X
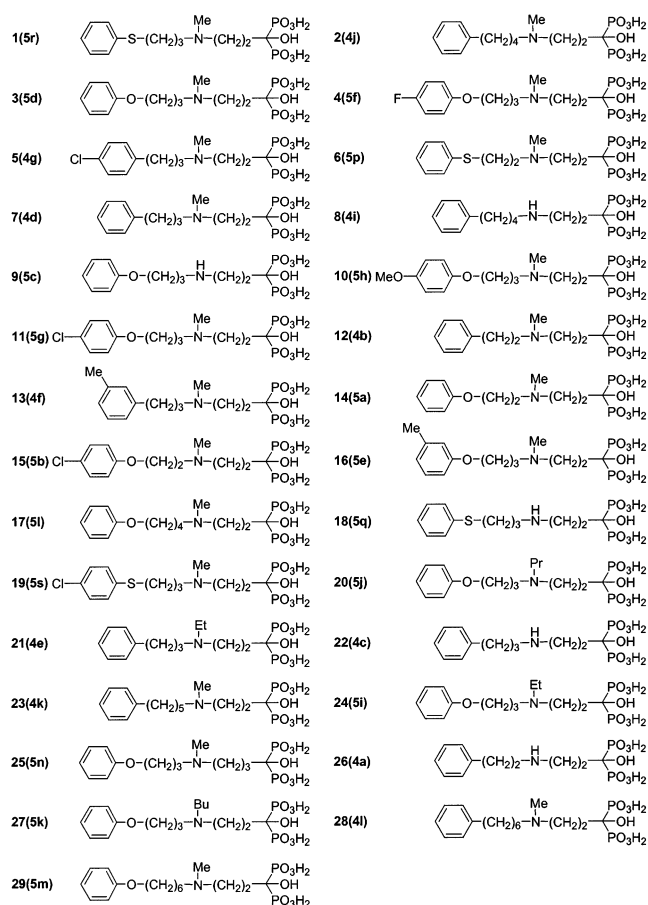

Figure 1. Structures of the 29 aryl-X bisphosphonates investigated. The designations in parentheses are those used by Widler et al. (ref 1).

derivatives of pamidronate (and one derivative of alendronate) and the heterocyclic bisphosphonates, to try to isolate possible deficiencies in our initial approach.

Aryl-X Bisphosphonates. We show in Figure 1 the structures of the 29 aryl bisphosphonates whose activities in inhibiting bone resorption were reported by Widler et al.[1] The structures shown in Figure 1 are listed (left to right and top to bottom) in order of decreasing activity. Only compounds having discrete activities are presented. To facilitate comparison with the published results, the code designations used by Widler et al. are shown in parentheses in Figure 1. All of these compounds are analogues of pamidronate or alendronate.
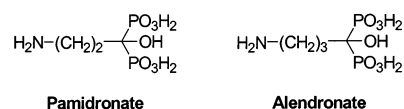
The experimentally determined activities of compounds **1**–**29** are shown in Table 1 and cover a very wide range in activity (ED$_{50}$ values from 0.33 to 7500 μg/kg). The ED$_{50}$ values are taken directly from Widler et al. and are expressed in μg/kg, while the pIC$_{50}$ values are those we have estimated assuming a rat density ρ = 1, using the relation pIC$_{50}$ = $-$log (IC$_{50}$,M). This approximation is necessary for accurate QSAR investigations since the molecular weights of the compounds investigated differ considerably. The compounds of interest contain a broad range of nitrogen substitutions,

Table 1. Experimental (ED_{50} and pIC_{50}) and MFA Predicted (pIC_{50}) Values for Bisphosphonates against Bone Resorption and Statistical Results of 3D-QSAR MFA Models

compound[a]	experimental activity ED_{50} (μg/kg)	pIC_{50}	predicted pIC_{50}[b] training set	four-compound test sets				
1 (5r)	0.33	9.209	8.635	8.632	8.669	8.691	8.624	8.620
2 (4j)	0.4	8.979	8.494	**8.491**	8.512	8.550	8.487	8.473
3 (5d)	0.5	8.885	8.716	8.714	**8.754**	8.773	8.708	8.700
4 (5f)	0.6	8.835	8.306	8.298	8.333	8.354	**8.292**	8.309
5 (4 g)	0.7	8.778	8.489	8.482	8.525	**8.539**	8.475	8.480
6 (5p)	0.7	8.741	8.566	8.559	8.601	8.618	8.553	8.541
7 (4d)	1	8.586	8.637	8.633	8.671	8.692	**8.627**	8.616
8 (4i)	1.0	8.565	8.368	8.360	8.393	8.417	8.355	8.356
9 (5c)	1.2	8.488	8.383	8.378	8.401	8.435	8.374	8.373
10 (5h)	1.2	8.546	8.626	8.622	8.667	8.680	8.616	8.621
11 (5 g)	1.3	8.506	8.472	8.466	8.503	8.524	8.461	8.467
12 (4b)	1.4	8.407	8.395	8.387	8.417	**8.446**	8.382	8.370
13 (4f)	1.5	8.405	8.651	8.649	8.686	8.708	**8.643**	8.638
14 (5a)	1.5	8.402	8.450	8.443	8.470	8.502	8.438	8.416
15 (5b)	1.7	8.385	8.303	8.293	8.326	8.350	8.287	**8.281**
16 (5e)	1.7	8.369	8.731	**8.729**	8.771	8.789	8.723	8.715
17 (5l)	4	7.997	8.432	8.428	8.444	**8.487**	8.425	8.408
18 (5q)	7	7.741	8.439	8.434	**8.465**	8.491	8.429	8.434
19 (5s)	7.8	7.745	8.425	8.419	8.453	8.476	8.413	8.420
20 (5j)	10	7.624	7.416	7.365	7.511	7.412	**7.348**	7.423
21 (4e)	15	7.405	7.293	**7.241**	7.376	7.287	7.225	7.299
22 (4c)	20	7.253	6.580	6.581	6.574	6.580	6.581	6.586
23 (4k)	20	7.306	7.418	7.367	7.513	7.413	7.350	**7.427**
24 (5i)	20	7.142	7.416	7.365	**7.511**	7.411	7.348	7.423
25 (5n)	100	6.609	6.544	**6.499**	6.570	6.536	6.490	6.637
26 (4a)	300	6.053	6.721	6.721	6.728	6.721	6.720	6.716
27 (5k)	500	5.930	5.943	5.937	5.956	5.948	5.936	**6.442**
28 (4l)	1500	5.436	5.188	5.166	**4.946**	5.209	5.188	5.173
29 (5m)	7500	4.754	5.047	5.021	4.808	5.063	5.043	5.044
R^2[c]			0.900	0.902	0.901	0.919	0.906	0.886
F_{test}[d]			54.275	46.272	45.278	56.734	48.277	39.946
R^2_{cv}[e]			0.799	0.758	0.713	0.809	0.802	0.766
R^2_{bs}[f]			0.900	0.902	0.901	0.919	0.906	0.886
N[g]			5	5	5	5	5	5
N[h]			29	25	25	25	25	25

[a] See Figure 1 for compound structures. Structures in parentheses are from ref 1. [b] Bold values represent predicted activities of compounds that were not included in the training set. [c] Correlation coefficient. [d] Ratio of R^2 explained to unexplained $= R^2/(1 - R^2)$. [e] Cross-validated correlation coefficient after leave-one-out procedure. [f] Average squared correlation coefficient calculated during the validation procedure. [g] Optimal number of principal components. [h] Number of compounds.

as shown in Figure 1. The most active compounds (**1**–**6**) are all analogues of pamidronate containing one N-Me group, but the aryl-X side chains differ in overall length and chemical composition. It is also clear from even a cursory inspection of Figure 1 that those compounds having the longest side chains (**28**, **29**) are the ones that are the least active.

We used the molecular field analysis (MFA) approach[27] to develop QSAR models for the compounds shown in Figure 1 using basically the techniques described in detail previously.[23,24,26] To guide the molecular alignment, we utilized the X-ray crystallographic structure of farnesyl pyrophosphate bound to an avian farnesyl pyrophosphate synthase.[28] This facilitated positioning of the extended side chains and the alignment so obtained is shown in Figure 2. We also utilized the crystallographic structure of risedronate,[29] docked onto geranyl pyrophosphate, to localize the positions of the nitrogen atoms in the aryl bisphosphonates. A second and less successful approach, in which the alignment was guided by the crystallographic structure of geranyl pyrophosphate bound to FPP synthase, was also investigated, and is described in more detail below. In all cases, the nitrogens in the bisphosphonate side chains were protonated, while the bisphosphonate phos-

phonate groups each carried a single negative charge. This is the pattern of protonation which has been found in several single-crystal X-ray diffraction studies of bisphosphonates crystallized in the pH range ~4–10,[23,26,29] and is also what we used previously.[23,24,26] Bisphosphonate structures were generated and geometry optimized by using a three-step protocol consisting of steepest descents,[30] conjugate gradients,[31] and Newton–Raphson minimization,[32] with no constraints on the internal geometries of the molecules, using the minimizer function of the OFF methods module in Cerius[2] 4.6.[33] Each molecule was aligned to the shape reference compound, farnesyl pyrophosphate, using the shape reference alignment function of the QSAR module of Cerius[2] 4.6. A second set of alignments was made by using a set of fully extended aryl side chains.

MFA 3-D QSAR methodology was then applied to the bone resorption inhibition data, to help identify the structural features required for optimum bone resorption activity. We applied a regression analysis to the data, leading to eq 1:

$$pIC_{50} = 5.137 - 0.021(\text{"H}^+/326\text{"}) + 0.052(\text{"CH}_3/335\text{"}) - 0.073(\text{"H}^+/514\text{"}) - 0.041(\text{"H}^+/564\text{"}) \quad (1)$$
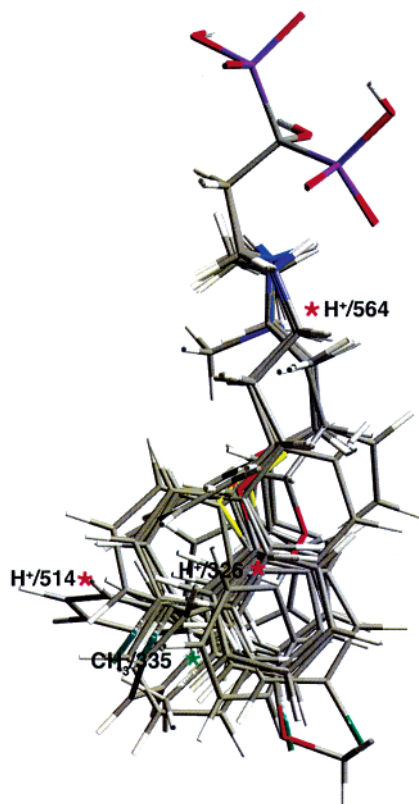
Figure 2. Maximum common subgroup alignment of the 29 aryl-X bisphosphonates that had quantitatively measurable activities, together with the locations of the four descriptors given in eq 1.

where the H^+ descriptors define the corresponding interaction energies between a proton probe and the molecule at the specified (numbered) grid points; see Figure 2 for their locations. The CH_3 descriptor defines the corresponding interaction energy between a hydrophobic probe and the molecule at the specified (numbered) grid point, as described by a Lennard-Jones 6−12 potential. The descriptor positions and coefficients in the 3D-QSAR/MFA equation were chosen by using a genetic function approximation (GFA) algorithm.[34] The genetic function approximation involves starting with an initial population of random equations, which are then allowed to evolve by application of crossover operations such that terms from models with the best fit will most likely occur in the "offspring" models. Best-fit models are retained, while poor-fit models are replaced with subsequent generations. The optimal number of components in the final GFA model was determined by cross validated R^2 values (R^2_{cv}), as obtained from the leave-one-out cross validation technique,[35] standard error prediction values, and also by F-test values.[36] An MFA model containing the five components shown above afforded a correlation coefficient $R^2 = 0.900$ and a cross-validation R^2_{cv} of 0.80. These statistical parameters are given in Table 1, together with the experimental and predicted pIC_{50} values. Training set results for the 29 compounds are shown graphically in Figure 3A, and a confidence limit of 95% was established by using the data randomization procedure in Cerius[2] 4.6.

   Next, we investigated to what extent it might be possible to predict the activities of such aryl-X bispho-
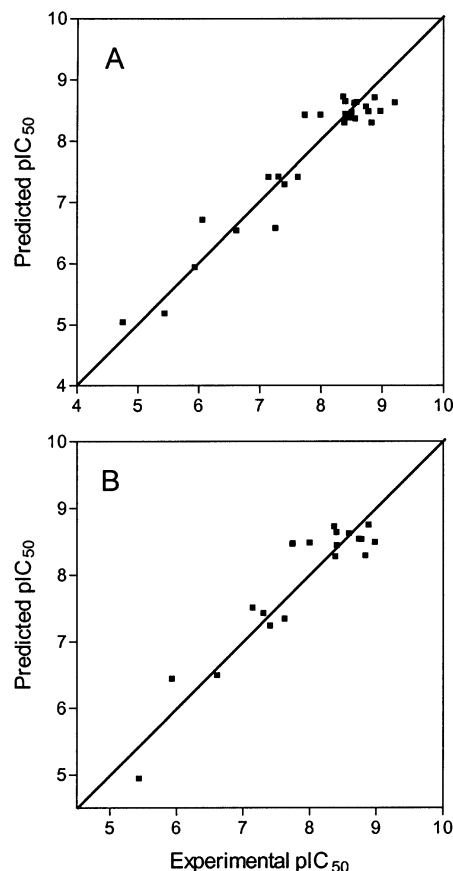

Figure 3. Graphical comparisons between experimental and predicted pIC_{50} values for bone resorption inhibition. A, 29-compound training set. B, superposition of five, four-compound test sets. For the test set, $R^2 = 0.854$.

sphonates by using the MFA approach. We followed basically the same approach as that reported previously,[24,26] in which we deleted four compounds from the overall training set of 29 compounds, then we used the reduced training set of 25 compounds to compute new MFA equations, which were then used to predict the activities of the four test set compounds. This procedure was repeated four more times using different training set compounds, resulting in five sets of four compound predictions ($N = 20$). The five sets of results are shown in Table 1, with the predicted activities shown in bold. The five new QSAR equations utilized were very similar, and were as follows:

$$pIC_{50} = 5.123 - 0.022(\text{"}H^+/326\text{"}) + 0.052(\text{"}CH_3/335\text{"}) - 0.073(\text{"}H^+/514\text{"}) - 0.040(\text{"}H^+/564\text{"}) \quad (2)$$

$$pIC_{50} = 5.074 - 0.021(\text{"}H^+/326\text{"}) + 0.053(\text{"}CH_3/335\text{"}) - 0.079(\text{"}H^+/514\text{"}) - 0.043(\text{"}H^+/564\text{"}) \quad (3)$$

$$pIC_{50} = 5.109 - 0.023(\text{"}H^+/326\text{"}) + 0.054(\text{"}CH_3/335\text{"}) - 0.074(\text{"}H^+/514\text{"}) - 0.041(\text{"}H^+/564\text{"}) \quad (4)$$

$$pIC_{50} = 5.129 - 0.023(\text{"}H^+/326\text{"}) + 0.052(\text{"}CH_3/335\text{"}) - 0.073(\text{"}H^+/514\text{"}) - 0.039(\text{"}H^+/564\text{"}) \quad (5)$$

$$pIC_{50} = 5.224 - 0.021(\text{"}H^+/326\text{"}) + 0.052(\text{"}CH_3/335\text{"}) - 0.073(\text{"}H^+/514\text{"}) - 0.027(\text{"}H^+/564\text{"}) \quad (6)$$

Table 2. Comparison of Statistics of Bent Aryl-X versus Linear Bisphosphonates

statistics	bent (FPP-like)	linear
$R^{2\ a}$	0.900	0.853
$F_{\text{test}}{}^{b}$	54.275	34.904
$R^2_{\text{cv}}{}^{c}$	0.799	0.757
$R^2_{\text{bs}}{}^{d}$	0.900	0.853
N^e	5	5
N^f	29	29

a Correlation coefficient. *b* Ratio of R^2 explained to unexplained = $R^2/(1 - R^2)$. *c* Cross-validated correlation coefficient after leave-one-out procedure. *d* Average squared correlation coefficient calculated during the validation procedure. *e* Optimal number of principal components. *f* Number of compounds.

The predicted activities for the 20 test set compounds are shown graphically in Figure 3B, plotted as a function of the experimental pIC_{50} values. For these 20 compounds, the average absolute error in pIC_{50} prediction is 0.319, which corresponds to about a 2.1× error, in this case covering a 4000× range in experimental activity. This is similar to the 2.8× error found previously (over a similar range in activity) for a series of aminomethylene and 1-hydroxybisphosphonates which lack the long aryl-X side chains.[26]

We also investigated a second alignment of molecules **1**−**29**, in which the alignments of the side chains of the compounds were guided by the X-ray crystal structure of geranyl pyrophosphate bound to FPP synthase.[28] This alignment performed well for the shorter chain compounds, but not for the longest compounds (**28** and **29**). Their side chains adopted an essentially all-trans conformation, since this is the expected lowest energy conformation. With this alignment there was a noticeable degradation in each of the statistical parameters computed, as shown in Table 2. Specifically, the R^2 value decreased from 0.900 to 0.853, the F-test value decreased from 54.3 to 34.9, and the R^2_{cv} decreased from 0.799 to 0.757. These results strongly suggest that the longer arylalkyl and related side chains bind in the "bent" or "curved" manner exhibited by farnesyl pyrophosphate when bound in the FPP synthase active site.[28] We show in Figure 4 an illustration of the conformations of geranyl and farnesyl pyrophosphate in the active site of farnesyl pyrophosphate synthase (Figures 4A, 4C), together with both extended (Figure 4B) and bent (Figure 4D) structures of **29**. The activities of the least active compounds are quite well described using this bent structure (4D), as shown in Table 1, and as noted above, all the training set statistics are improved with this alignment. These results imply that fully extended long aryl-X side chains will encounter strong repulsive interactions with the protein. The longer chain inhibitors thus conform to the "bent" FPP structure in the FPP synthase binding site. However, even in this bent conformation, **28** and **29** are still very poor inhibitors of bone resorption, which strongly suggests that there is still a repulsive or steric interaction with the protein, resulting in relatively weak binding, rather than an enhanced hydrophobic stabilization, due to the increased chain length. This can, however, be readily understood based on the results of several mutagenesis studies. For example, it has been shown that there are two phenyl groups in FPP synthase which when converted to the smaller Ala and
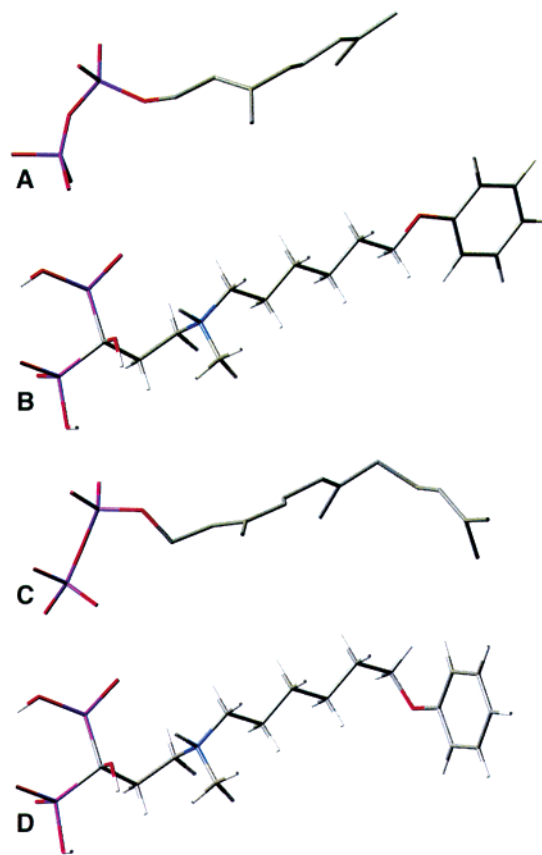

Figure 4. Conformations of A, GPP (PDB File 1UBW); B, **29**, extended alignment; C, FPP (PDB File 1UBX); and D, **29**, alignment based on C.

Ser groups, permit chain elongation from C_{15} to C_{20}.[28] These two Phe groups appear to provide a steric block to chain elongation, and it seems likely that they may also have a repulsive interaction with the long arylalkyl side chains of **28** and **29**, resulting in the high (1500, 7500 µg/kg) ED_{50} values, Table 1. A related effect also appears to be operative in geranylgeranyl pyrophosphate synthase (GGPP synthase), where engineering an MS → YF mutation in the same region results in conversion of GGPP to FPP synthase activity.[37]

As can be seen in the data of Widler et al.,[1] the shorter chain aryl-X compounds having 5−6 heavy (i.e., carbon, nitrogen, oxygen, or sulfur) atoms in the side chain linker (and containing an N-Me substituent) typically have somewhat less activity than do slightly longer chain compounds, and maximal activity (0.33−0.5 µg/kg) is seen with **1**−**3**, which all have seven heavy atoms (plus a phenyl) in their side chain. This corresponds to a chain length longer than that which would be found in the GPP substrate or reactive intermediate. That is, optimum activity is found with species having chain lengths close to that of FPP, but not much longer, where repulsive interactions occur. Shorter chains are less active, presumably due to decreased hydrophobic stabilization in the active site.

Heterocyclic Bisphosphonates. Next, we describe our QSAR results for a series of heterocyclic bisphosphonates. The structures of the 26 compounds considered (**30**−**55**) are shown in Figure 5, together with the designations (in parentheses) used by Widler at al.[1] The activities of the compounds are, once again, taken
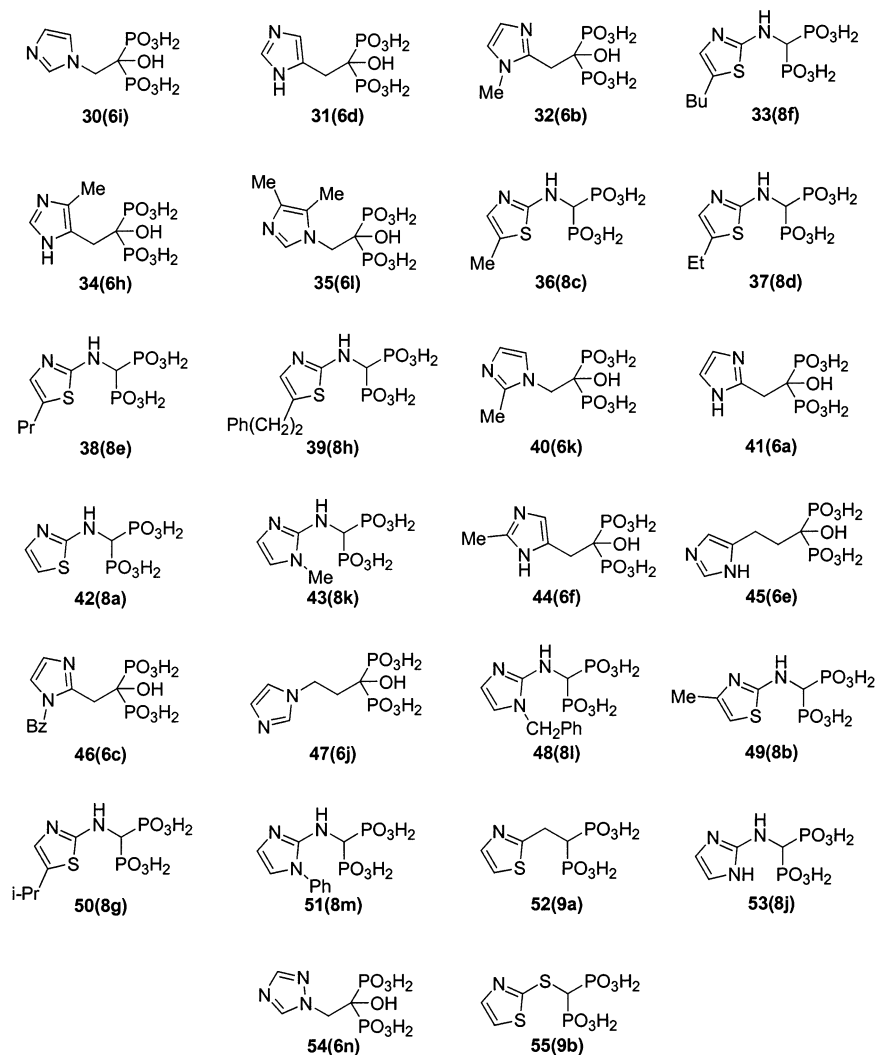
Figure 5. Structures of the 26 heterocyclic bisphosphonates investigated. The designations in parentheses are those used by Widler et al. (ref 1).

directly from the paper of Widler et al. and are reproduced as ED_{50} (μg/kg) values and as transformed pIC_{50} values, in Table 3. The compounds investigated have quite broad structural diversity in that activities for both aminomethylene bisphosphonate and 1-hydroxy-1,1-bisphosphonate backbones, together with imidazole, thiazole, triazole, and pyrazole side chains have been reported, with activities covering a 10 000× range (or >100 000× if the "inactive" benzannulated species are considered).

We began our investigation of these compounds using the structures of geranyl pyrophosphate and risedronate to guide the alignment. The phosphonate moieties each contained one protonated and two nonprotonated oxygens (the same as with the aryl-X bisphosphonates), and the ring nitrogens were all protonated. This basic motif has been used successfully by us in the past[26] to predict the activities of a small set of, primarily, pyridyl bisphosphonates, but with this more diverse set of compounds, the initial training set results were poor. While it might be that our alignment procedure was inadequate, such problems had not been observed previously, so we hypothesized that this poor agreement between experiment and prediction might in fact be due to an inappropriate choice of protonation states

for these new heterocycles. For example, it is known that thiazoles[38] and triazoles[39] are very weak bases (pK_a ~3), unlike imidazoles (pK_a ~7.2), so it seemed likely that the low activity of compounds such as **52**, **54**, and **55** might be related to their weak basicity. However, some thiazoles, for example **33**, **36**, **37**, **38**, and **39**, have quite pronounced activity. The difference between the active thiazoles and the most inactive species, at least to first order, would thus appear to be due to the presence of either α-NH protonation (or amidinium-like structures) in the most active thiazoles, which would be absent in the least active compounds, although other factors, such as the presence of an unbranched 4′-substituent, also appear to contribute to high activity. In any case, to test the protonation state hypothesis, we next computed QSAR equations using structures in which imidazolium groups were ring-protonated, aminothiazoles were protonated on the α-N, while the triazole and non-amino-substituted thiazoles were not protonated.
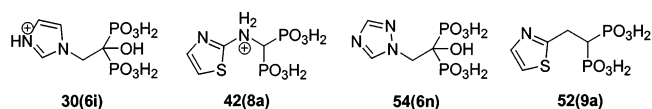
Table 3. Experimental (ED_{50} and pIC_{50}) and MFA Predicted (pIC_{50}) Values for Bisphosphonates against Bone Resorption and Statistical Results of 3D-QSAR MFA Models

compound[a]	experimental activity		predicted pIC_{50}[b]					
	ED_{50} (μg/kg)	pIC_{50}	training set	four-compound test sets				
30 (6i)	0.07	9.617	9.162	9.135	9.238	9.060	9.129	9.277
31 (6d)	0.3	8.958	9.281	9.253	9.359	**9.177**	9.242	9.399
32 (6b)	0.6	8.705	8.134	8.080	8.211	**8.026**	8.156	8.226
33 (8f)	0.9	8.565	8.378	8.452	8.365	8.384	8.355	**8.424**
34 (6h)	1.5	8.287	7.958	7.977	7.915	8.105	**7.835**	7.984
35 (6l)	1.5	8.301	7.746	7.799	7.735	7.744	7.758	7.780
36 (8c)	1.5	8.290	8.370	8.446	**8.355**	8.376	8.347	8.415
37 (8d)	1.5	8.317	8.376	8.450	8.362	8.382	**8.353**	8.421
38 (8e)	2	8.211	8.377	8.451	8.363	8.383	8.354	**8.422**
39 (8h)	2.7	8.146	8.378	8.452	8.364	8.384	8.355	**8.424**
40 (6k)	3	7.979	7.982	7.995	7.944	8.123	7.860	8.012
41 (6a)	5	7.750	8.266	8.208	8.349	8.151	8.283	**8.365**
42 (8a)	5	7.739	7.126	**7.054**	7.197	7.019	7.200	7.193
43 (8k)	5	7.734	7.210	7.342	7.245	**7.077**	7.234	7.235
44 (6f)	15	7.294	7.828	**7.885**	7.817	7.827	7.835	7.863
45 (6e)	20	7.176	6.940	6.895	**6.925**	7.046	6.881	6.959
46 (6c)	25	7.187	7.150	7.299	7.171	7.033	**7.172**	7.165
47 (6j)	45	6.803	6.844	**6.797**	6.830	6.949	6.791	6.862
48 (8l)	75	6.732	7.198	**7.329**	7.234	7.064	7.223	7.223
49 (8b)	100	6.460	7.112	7.041	**7.182**	7.006	7.187	7.178
50 (8 g)	200	6.199	5.931	5.925	**5.931**	5.907	6.043	5.930
51 (8m)	200	6.222	5.938	6.099	5.865	6.039	5.880	5.879
52 (9a)	200	6.135	6.320	6.242	6.381	6.222	**6.434**	6.363
53 (8j)	~500	5.711	6.505	6.558	6.462	**6.610**	6.445	6.488
54 (6n)	600	5.658	5.583	5.506	5.566	5.683	5.596	5.571
55 (9b)	700	5.619	5.705	5.622	5.697	5.797	5.715	5.701
R^{2}[c]			0.873	0.902	0.879	0.899	0.865	0.876
F_{test}[d]			35.991	39.158	30.814	37.803	27.267	30.034
R^{2}_{cv}[e]			0.809	0.846	0.802	0.827	0.775	0.799
R^{2}_{bs}[f]			0.873	0.902	0.879	0.899	0.865	0.876
N[g]			5	5	5	5	5	5
N[h]			26	26	26	26	26	26

[a] See Figure 5 for compound structures. Structures in parentheses are from ref 1. [b] Bold values represent predicted activities of compounds that were not included in the training set. [c] Correlation coefficient. [d] Ratio of R^2 explained to unexplained = $R^2/(1 - R^2)$. [e] Cross-validated correlation coefficient after leave-one-out procedure. [f] Average squared correlation coefficient calculated during the validation procedure. [g] Optimal number of principal components. [h] Number of compounds.

We then computed QSAR equations using the MFA technique and the alignment shown in Figure 6. We obtained the following MFA QSAR equation:

$$pIC_{50} = 7.781 - 0.052(\text{``}CH_3/156\text{''}) + 0.050(\text{``}H^+/227\text{''}) + 0.021(\text{``}H^+/294\text{''}) - 0.032(\text{``}CH_3/365\text{''}) \quad (7)$$

The spatial locations of the grid points are shown in Figure 6 and the detailed training set prediction results are shown in Table 3. The statistical results so obtained are compared with those obtained by using uniformly protonated rings in Table 4, where it can be seen that there is a marked improvement using the revised protonation scheme. For example, the R^2 value increases from 0.710 to 0.873, the cross-validated R^2_{cv} increases from 0.629 to 0.809, and the F-test result increases from 18 to 36.

The new training set predictions are shown plotted as a function of the experimental pIC_{50} values in Figure 7A. These results are somewhat worse than those obtained with the aryl bisphosphonates, but they nevertheless do suggest predictive utility. To test this idea further, we again carried out five training ($N = 22$)/test-($N = 4$) set calculations for a total of 20 activity predictions, shown in bold in Table 3 and graphically

in Figure 7B. The MFA equations so derived were the following:

$$pIC_{50} = 7.737 - 0.049(\text{``}CH_3/156\text{''}) + 0.052(\text{``}H^+/227\text{''}) + 0.020(\text{``}H^+/294\text{''}) - 0.025(\text{``}CH_3/365\text{''}) \quad (8)$$

$$pIC_{50} = 7.815 - 0.055(\text{``}CH_3/156\text{''}) + 0.050(\text{``}H^+/227\text{''}) + 0.023(\text{``}H^+/294\text{''}) - 0.034(\text{``}CH_3/365\text{''}) \quad (9)$$

$$pIC_{50} = 7.776 - 0.048(\text{``}CH_3/156\text{''}) + 0.051(\text{``}H^+/227\text{''}) + 0.017(\text{``}H^+/294\text{''}) - 0.032(\text{``}CH_3/365\text{''}) \quad (10)$$

$$pIC_{50} = 7.750 - 0.050(\text{``}CH_3/156\text{''}) + 0.047(\text{``}H^+/227\text{''}) + 0.022(\text{``}H^+/294\text{''}) - 0.032(\text{``}CH_3/365\text{''}) \quad (11)$$

$$pIC_{50} = 7.844 - 0.055(\text{``}CH_3/156\text{''}) + 0.051(\text{``}H^+/227\text{''}) + 0.022(\text{``}H^+/294\text{''}) - 0.034(\text{``}CH_3/365\text{''}) \quad (12)$$

and show considerable similarity. For the $N = 20$ test set prediction, we found using these MFA equations an average absolute error of 0.39 in the pIC_{50} predictions, which corresponds to about a factor of $2.5\times$ average error over the $2000\times$ range in predicted activity. This, once again, indicates that the MFA technique appears to have utility in predicting ED_{50} values even in as
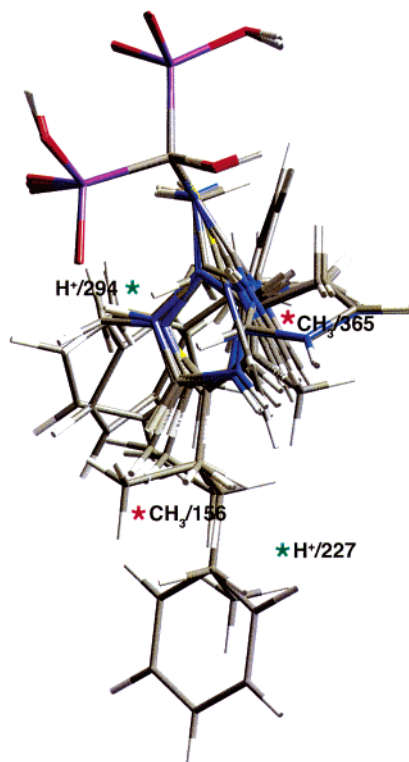
Figure 6. Maximum common subgroup alignment of the 26 heterocyclic bisphosphonates that had quantitatively measurable activities, together with the locations of the four descriptors given in eq 7.

Table 4. Comparison of Ring Protonated versus Selectively Protonated Heterocyclic Bisphosphonate MFA Statistics

statistics	ring protonated	selectively protonated
$R^{2 \ a}$	0.710	0.873
$F_{test}{}^b$	17.991	35.991
$R^2{}_{cv}{}^c$	0.629	0.809
$R^2{}_{bs}{}^d$	0.710	0.873
N^e	5	5
N^f	26	26

^{a} Correlation coefficient. ^{b} Ratio of R^2 explained to unexplained = $R^2/(1 - R^2)$. ^{c} Cross-validated correlation coefficient after leave-one-out procedure. ^{d} Average squared correlation coefficient calculated during the validation procedure. ^{e} Optimal number of principal components. ^{f} Number of compounds.

complex a system as bone resorption, although it is equally clear that additional information on protein crystal structures will be desirable to clarify more precisely how these inhibitors function. For example, while it seems very likely that the long aryl-X compounds are GPP mimics, the small heterocycles might also inhibit at the isopentenyl diphosphate (IPP) site (or both GPP and IPP sites), as suggested recently for risedronate.[21,22] Nevertheless, when both the aryl-X and heterocyclic predictions are combined, Figure 7C, it is clear that quite accurate activity predictions ($2.3\times$ average error) can now be made.

pK_a-Activity Correlations. The results described above suggest an important relationship between the pK_a of the heterocycle ring and activity. While this is of course not wholly unexpected, the relationship appears not to have been previously reported with bisphosphonate drugs and in fact only becomes apparent when data on a wide range of compounds, such as the thiazoles, pyrazoles, etc., is available, as exemplified in the recent
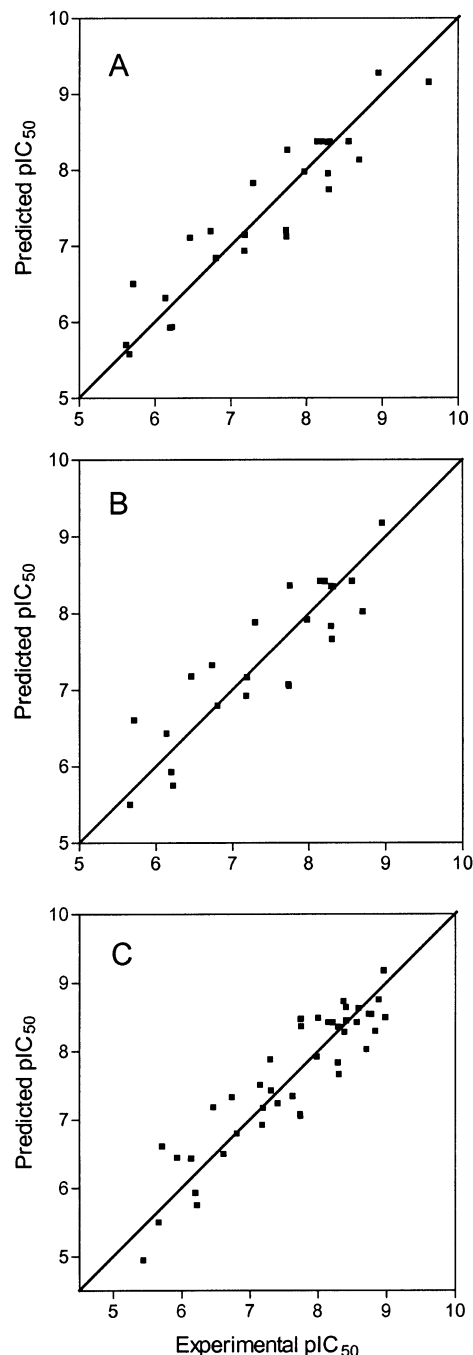

Figure 7. Graphical comparisons between experimental and predicted pIC$_{50}$ values for bone resorption inhibition. A, 26-compound heterocyclic training set. B, Superposition of five, four-compound heterocyclic test set predictions. C, Superposition of the five, four-compound test sets for both aryl-X and heterocyclic bisphosphonates ($N = 40$). For the heterocyclic test set B, $R^2 = 0.784$. For the combined test sets, $R^2 = 0.824$.

work of Widler et al. Here, then, we explore this topic in more detail and try to put on a somewhat more firm foundation some of the very interesting (and at first sight, puzzling) recent results. Moreover, we extend the investigation to bisphosphonates investigated by other groups to further test some of these ideas, which if correct, may have important implications for bisphosphonate drug design.

We show in Figure 8 the structures and pK_a values computed by using SciFinder Scholar/ACD software[40] of nine heterocycles which form the side chains of

56 (pK_a=2.44)   **57** (pK_a=2.46)   **58** (pK_a=2.55)   Inactive, pK_a <3

59 (pK_a=5.25)   **60** (pK_a=5.36)   Active

pK_a ~5-9

61 (pK_a=6.82)   **62** (pK_a=6.95)   **63** (pK_a=9.11)   Very active
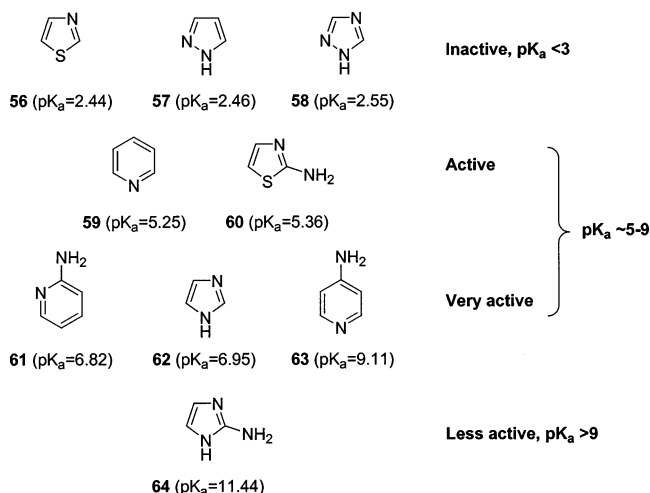
64 (pK_a=11.44)   Less active, pK_a >9

Figure 8. Structures of nine heterocycles with their computed pK_a values and an indication of the activities of the derived bisphosphonates.

bisphosphonates having a wide range of activity in bone resorption, and in some cases as herbicides.[41] The free base compounds are ordered by increasing basicity. Comparison of these computed heterocycle pK_a values with the corresponding bisphosphonate activity results reported by Widler et al.[1] immediately helps explain many of the rather puzzling activity results obtained. For example, Widler et al. note that the 2-aminothiazolyl derivative **42** (their **8a**) has a potency comparable to that of alendronate, or at least 10 times that of pamidronate, but unexpectedly, replacement of the nitrogen linker atom of the thiazole by a methylene or sulfur bridge led to species (**52**, **55**) which were much less potent. We propose this is related to the basicity of the heterocycle rings. Specifically, the pK_a of free thiazole (**56**) is 2.44,[38,40] so S- and CH$_2$-linked bisphosphonates will be unprotonated, lacking the key positive charge feature required for carbocation-mimicry. On the other hand, the bisphosphonate (**42**) derived from 2-aminothiazole (**60**), whose parent base has a pK_a of 5.36, would likely be protonated on the α-N, with some resonance stabilization, and would consequently be active. Similarly, low pK_a values are found with the free pyrazole **57** (basic pK_a = 2.46) and the 1,2,4-triazole **58** (basic pK_a = 2.55), and the corresponding bisphosphonates are inactive.[1] Specifically, the ED$_{50}$ for a pyrazole bisphosphonate (**6m** in ref 1) is >300 μg/kg while that for the triazole bisphosphonate **54** is 600 μg/kg, since in the absence of a linker nitrogen there is unlikely to be any positive charge feature. The low basicity of such triazoles is also confirmed in studies of phospholipase A$_2$ and in lysozyme in which 1,2,4-triazole alanine was substituted for histidine.[42,43] On the other hand, **59**−**62** all have pK_a values in the range ~5−7, and the derived bisphosphonates are generally all very active and include species such as risedronate and zoledronate (**30**). These results strongly support the idea that a pK_a for the heterocyclic ring in the range ~5−7 is likely to support high activity, due to carbocation formation, although certainly it is necessary to have the protonatable group in a suitable location. For example, the para pyridyl analogue of risedronate has generally low activity.[44] However, the presence of a para nitrogen

in a phenyl ring, as opposed to an ortho or meta nitrogen, does not in and of itself guarantee low activity. For example, **63** (Figure 8) has a para nitrogen but is very basic, with a pK_a of 9.11, and the corresponding aminomethylene bisphosphonate was found by Nissan[41] to be one of the most potent herbicidal bisphosphonates investigated, many of which are known low nM inhibitors of FPP synthase (such as analogues of **61**). This activity is likely to be due to the large charge delocalization which can be expected in the quinonoid-like structure, shown on the left below, in much the same way that charge delocalization can occur in the amidinium-like species which can be formed for **61** with the
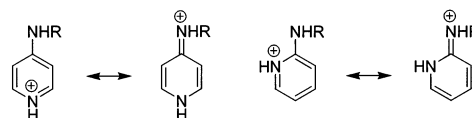
charge being highly delocalized over the entire ring in these quinonoid-like species, as shown above.

Paradoxically, having an exceptionally basic nitrogen may contribute to decreased activity. Specifically, Widler et al.[1] note that most of the aminomethylene bisphosphonates containing imidazolyl groups are "surprisingly less interesting" with **53** (the unsubstituted species) having an activity of ~500 μg/kg, to be compared with the 0.07 μg/kg of zoledronate (**30**). One possible explanation for this extremely low activity may lie in the fact that the predicted pK_a of the parent compound (**64**) is ~11.4. In this case, the side chain will readily protonate and form a guanidinium-like species. While the charge
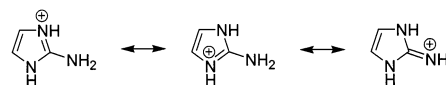
here will be delocalized, it may result in bisphosphonate species which are poorly transported into the cell, since they may be incapable of deprotonation. Addition of an N-Me group increases activity, as found in all of the N-Me pamidronate aryl-X species, and may contribute to hydrophobic stabilization, although the *N*-benzyl and *N*-phenyl species (**48**, **51**) are less active, due perhaps to steric repulsion in the active site.

While admittedly qualitative, we nevertheless believe that the major patterns of activity seen by Widler et al. for heterocyclic bisphosphonates can now be correlated with well-known pK_a values. The poor bases (pK_a < 3): thiazoles, pyrazoles, and triazoles, are all inactive. However, species with pK_a values in the range ≈ 5−9 can be very active. This is well illustrated by the conversion **52** → **42**, where a 40× increase in activity can be correlated with a pK_a change from 2.44 (**56**) to 5.36 (**60**), for the parent bases. However, species which cannot deprotonate (**43**, **48**, **51**, and **53**) are clearly in the less active category, although it is likely that if transported, they will be good inhibitors of FPP synthase, something which can be tested in future work. With benzannulated thiophenes and imidazoles, activity is very low (>3000 and 10 000 μg/kg, respectively). While activity might be expected with the benzthiazol-2-yl aminomethylene bisphosphonate, the C2-substitution most likely causes a steric repulsion with the

protein, as is found in the isosteric minodronate derivative.[45] In the benzimidazol-2-yl aminomethylene bisphosphonate (**8o** in ref 1), this steric effect is compounded with the proposed guanidinium fixed-charge effect, resulting in an ED_{50} of >10 000 μg/kg.

Pharmacophore Modeling. In addition to these molecular field analysis investigations, we used pharmacophore modeling to investigate structure–activity relationships for bisphosphonates in inhibiting bone resorption. We used the Catalyst approach,[46] in which up to 255 conformations of each molecule can be investigated, to develop a series of pharmacophore hypotheses. We used the structure/protonation files developed for the MFA analysis to generate a series of Catalyst pharmacophore hypotheses for both the aryl-X and heterocyclic bisphosphonates. The top scoring Catalyst hypotheses all contained two negative ionizable groups and a positive charge feature. For the aryl bisphosphonates, we added a custom feature, an aromatic ring attached to an alkyl chain via a sulfur or oxygen atom, while for the heterocyclic bisphosphonates, we added a custom ring feature, generating the pharmacophore hypotheses shown in Figures 9A and 9B, superimposed on **12**(A) and **30**(B). The two pharmacophore hypotheses are shown superimposed in Figure 9C. The two pharmacophores are clearly quite similar with respect to the location of the negative ionizable and positive charge features but, of course, the locations of the ring features (shown as green spheres) necessarily differ in the two set of compounds.

The training set activity predictions for both sets of molecules (**1–29**; **30–55**) are shown in Tables 5 and 6 and are somewhat worse than those obtained by using the MFA method, with R^2 values in the range 0.83–0.85, to be compared with 0.87–0.90 (Tables 1, 3). Likewise, the test set results (on the same two sets of 20 compounds as used in the molecular field analysis, Tables 1 and 3) shown in Tables 5 and 6 were less well correlated with experiment, with error factors of ~2.9–3.4 being found.

And finally, in an effort to make these basic QSAR approaches more synthetically useful, we have begun to explore the possible utility of a comparative molecular similarity indices analysis (CoMSIA)[47] of the results since it has been argued that CoMSIA contour maps may provide a useful graphical approach to drug design. We followed the MFA alignment protocol described above but used the Tripos Sybyl 6.8 program[48] for the CoMSIA analysis. The training set results were relatively predictive (average errors ~4.3–5.9) but more importantly yielded the CoMSIA contour maps shown in Figure 10 for the steric (A,D), electrostatic (B,E) and hydrophobic (C,F) interactions for the aryl-X (A,B,C) and heterocyclic (D,E,F) compounds. For the steric interaction, green represents a favorable interaction, yellow unfavorable. For the electrostatic interaction, blue is favorable for more positive charge density in the ligand while red is favorable for more negative charge density in the ligand. For the hydrophobic interaction, orange represents a favorable hydrophobic interaction; white denotes an area of unfavorable interaction.

For the aryl-X compounds, the hydrophobic interaction (C) dominates (59%) followed (A) by the steric contribution (31%), and the location of these features clearly correlate with the (favorable) presence of the
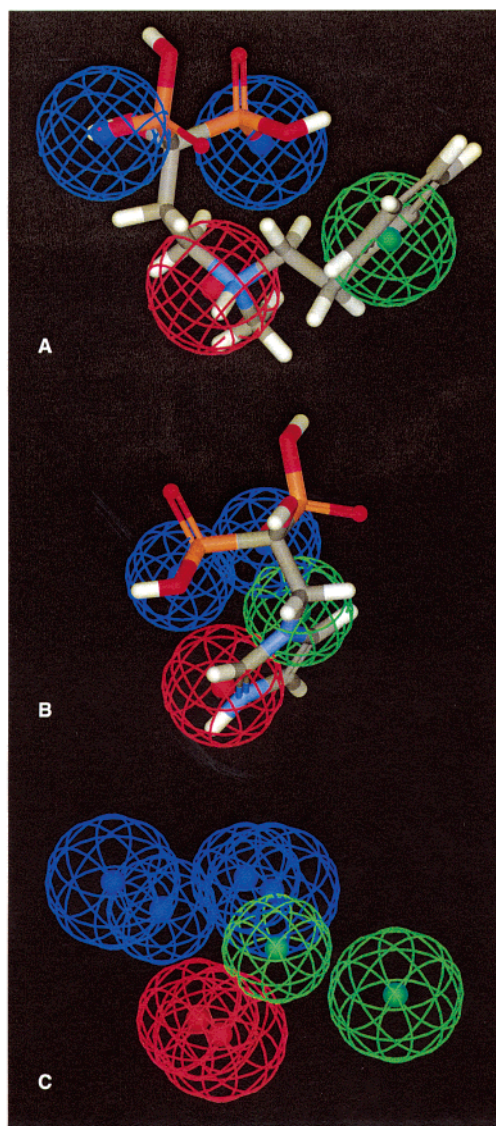

Figure 9. Catalyst pharmacophore hypotheses for bone resorption inhibition by aryl-X and heterocyclic bisphosphonates. The blue spheres represent negative ionizable features, the red spheres positive charge features, the green spheres custom hydrophobic features. In A, the green sphere represents an aromatic ring attached to a hydrophobic chain, in B, it represents an aromatic ring carbon. The pharmacophores are superimposed in A on compound **12(4b)**, in B, on compound **30(6i)**, zoledronate. C is the superposition of the two pharmacophores.

aromatic residues in the medium chain length compounds, Figure 10A,C. The red feature in Figure 10B indicates a favorable negative electrostatic feature in the region where the sulfur or oxygen atoms of the compounds are located, consistent with the observation that the thio analogue **1** is the most active species seen in this series.

For the heterocyclic bisphosphonates, the hydrophobic interaction (F) is still large (41%) but the electrostatic interaction has increased considerably, to 32%. The corresponding CoMSIA contour maps are shown in Figures 10E,F. Quite clearly, the hydrophobic (E) and steric (D) contour maps suggest that ring substitution might be beneficial. The yellow feature in Figure 10D represents a repulsive interaction found with some side chains (e.g., in **46**, **48**).

Table 5. Experimental (ED$_{50}$ and pIC$_{50}$) and Catalyst Predicted (pIC$_{50}$) Values for Bisphosphonates against Bone Resorption and Statistical Results

compound	experimental activity		predicted pIC$_{50}$[a]					
	ED$_{50}$ (µg/kg)	pIC$_{50}$	training set	four-compound test sets				
1 (5r)	0.33	9.209	8.489	8.763	8.497	8.648	8.502	8.686
2 (4j)	0.4	8.979	8.474	**8.581**	8.326	8.502	8.542	8.540
3 (5d)	0.5	8.885	8.192	8.584	8.121	8.542	8.291	8.407
4 (5f)	0.6	8.835	8.355	8.358	8.291	**8.291**	8.477	8.409
5 (4 g)	0.7	8.778	8.351	8.368	**8.301**	8.368	8.410	8.393
6 (5p)	0.7	8.741	8.183	8.307	8.355	8.264	**8.264**	8.188
7 (4d)	1	8.586	8.322	8.331	8.331	**8.206**	8.243	8.285
8 (4i)	1.0	8.565	7.963	8.286	8.060	8.310	8.363	**8.310**
9 (5c)	1.2	8.488	7.908	8.225	**8.266**	8.120	8.370	8.337
10 (5h)	1.2	8.546	8.280	8.395	8.450	8.395	8.541	**8.480**
11 (5 g)	1.3	8.506	8.382	8.240	8.341	8.319	8.377	8.444
12 (4b)	1.4	8.407	7.944	7.702	7.650	7.512	7.540	7.696
13 (4f)	1.5	8.405	8.293	8.326	8.183	**8.326**	8.277	8.259
14 (5a)	1.5	8.402	8.036	8.499	**8.256**	8.374	8.502	8.536
15 (5b)	1.7	8.385	8.241	8.273	8.293	8.315	8.257	8.360
16 (5e)	1.7	8.369	8.485	**8.237**	8.520	8.184	8.298	8.423
17 (5l)	4	7.997	8.334	8.081	8.201	8.108	8.243	**8.081**
18 (5q)	7	7.741	8.115	8.171	8.307	8.081	8.132	8.109
19 (5s)	7.8	7.745	7.935	8.407	8.523	8.359	8.407	**8.491**
20 (5j)	10	7.624	7.719	8.624	8.544	**8.544**	8.581	8.679
21 (4e)	15	7.405	8.022	**8.351**	8.303	8.201	8.323	8.377
22 (4c)	20	7.253	7.843	8.173	**8.211**	8.106	8.351	8139
23 (4k)	20	7.306	7.873	8.227	8.528	8.115	**8.239**	8.493
24 (5i)	20	7.142	8.214	8.239	8.188	8.267	8.433	8.297
25 (5n)	100	6.609	7.247	7.752	7.495	7.640	7.451	7.405
26 (4a)	300	6.053	6.705	7.489	7.557	7.300	7.166	7.384
27 (5k)	500	5.930	5.694	5.689	5.633	5.656	**5.653**	5.656
28 (4l)	1500	5.436	5.798	5.639	5.621	5.625	5.665	5.639
29 (5m)	7500	4.754	5.683	5.670	5.638	5.638	5.660	5.629
$R^{2\,b}$			0.85	0.73	0.66	0.74	0.72	0.71
N^c			29	25	25	25	25	25

[a] Bold values represent predicted activities of compounds that were not included in the training set. [b] Correlation coefficient for training set compounds. [c] Number of compounds.

Table 6. Experimental (ED$_{50}$ and pIC$_{50}$) and Catalyst Predicted (pIC$_{50}$) Values for Bisphosphonates against Bone Resorption and Statistical Results

compound	experimental activity		predicted pIC$_{50}$[a]					
	ED$_{50}$ (µg/kg)	pIC$_{50}$	training set	four-compound test sets				
30 (6i)	0.07	9.617	8.533	8.538	8.630	8.421	8.349	8.489
31 (6d)	0.3	8.958	8.393	8.393	**7.972**	8.745	8.443	8.571
32 (6b)	0.6	8.705	8.161	8.161	**7.759**	8.121	7.965	8.253
33 (8f)	0.9	8.565	8.343	8.440	8.000	8.565	**8.288**	8.477
34 (6h)	1.5	8.287	8.477	8.463	8.162	**8.872**	8.500	8.644
35 (6l)	1.5	8.301	8.155	8.398	7.864	8.199	8.199	**8.155**
36 (8c)	1.5	8.290	7.922	8.019	7.891	8.144	7.935	8.019
37 (8d)	1.5	8.317	7.949	8.002	8.477	**7.962**	7.962	8.031
38 (8e)	2	8.211	8.007	8.211	7.968	8.282	**8.211**	8.336
39 (8h)	2.7	8.146	7.814	8.347	7.965	8.323	**8.323**	8.464
40 (6k)	3	7.979	8.507	8.512	8.415	8.377	8.252	**8.457**
41 (6a)	5	7.750	7.795	7.858	7.836	7.592	**7.716**	7.733
42 (8a)	5	7.739	7.905	**7.991**	7.610	8.116	7.906	7.991
43 (8k)	5	7.734	6.902	6.820	**6.685**	6.800	6.865	6.717
44 (6f)	15	7.294	7.993	**8.128**	7.857	7.926	7.868	8.039
45 (6e)	20	7.176	7.221	7.508	7.744	7.272	7.513	7.435
46 (6c)	25	7.187	7.635	8.284	8.041	**7.895**	7.922	8.017
47 (6j)	45	6.803	7.226	**7.377**	7.156	7.201	7.415	7.280
48 (8l)	75	6.732	6.984	**7.566**	7.227	7.566	7.643	7.781
49 (8b)	100	6.460	6.778	6.380	6.590	6.608	6.735	6.515
50 (8 g)	200	6.199	6.737	7.199	5.746	6.857	6.819	6.995
51 (8m)	200	6.222	6.807	7.018	6.867	6.703	6.869	**6.979**
52 (9a)	200	6.135	5.783	5.931	7.018	**6.005**	5.704	6.021
53 (8j)	~500	5.711	6.437	6.632	**6.854**	6.471	6.618	6.578
54 (6n)	600	5.658	5.591	5.729	5.755	5.696	5.585	**5.746**
55 (9b)	700	5.619	5.774	5.884	5.732	5.831	5.740	5.908
$R^{2\,b}$			0.83	0.77	0.80	0.81	0.74	0.73
N^c			26	26	26	26	26	26

[a] Bold values represent predicted activities of compounds that were not included in the training set. [b] Correlation coefficient for training set compounds. [c] Number of compounds.

The electrostatic field result, Figure 10E, shows an interesting feature (in addition to the expected favored positive charge feature, on the imidazole ring shown), a negative charge favored region (red) encompassing the bisphosphonate 1-OH group. This feature is not seen in the aryl-X CoMSIA contour maps, Figure 10B, since
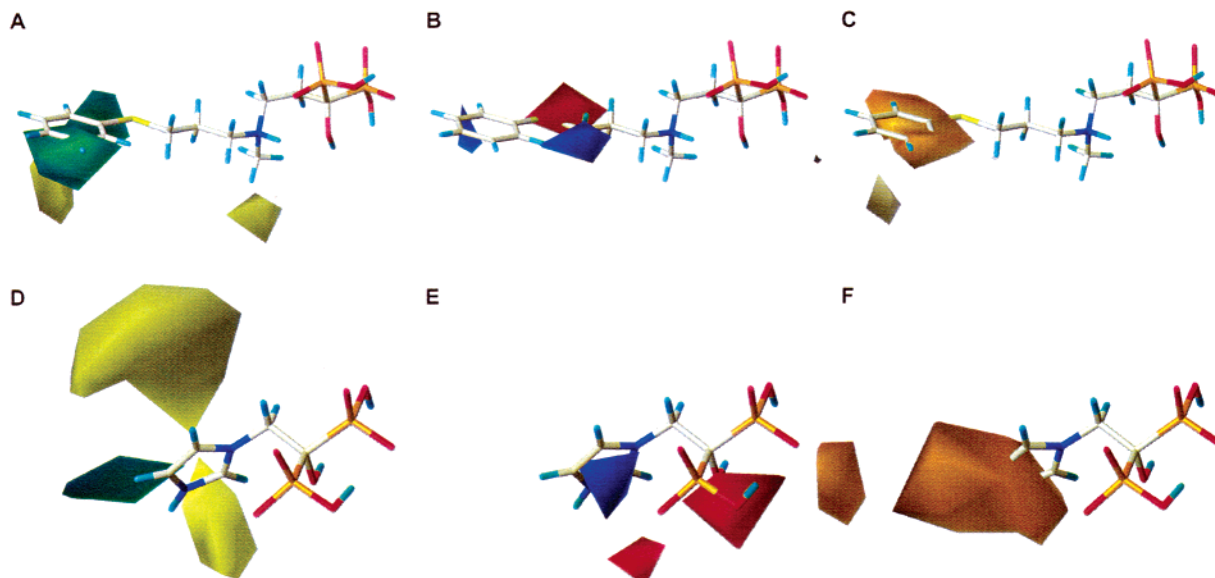
Figure 10. CoMSIA field analysis contours for **A**−**C**: aryl-X bisphosphonates and **D**−**F**, heterocyclic bisphosphonates. **A**, **D**, steric field (green, favored; yellow, disfavored). **B**, **E**, electrostatic field (blue, favored; red, disfavored). **C**, **F**, hydrophobic interactions (orange, favored; white, disfavored). The CoMSIA field contours are superimposed on **1(5r)** in **A**−**C** and zoledronate **30(6i)** in **D**−**F**.

all compounds in the aryl-X series have this feature in common. Since this feature is not seen in the CoMSIA field for FPP synthase itself (unpublished results) it seems likely that this group is indeed important in bisphosphonate binding to bone (the "bone hook"). These CoMSIA contour maps thus augment our pK_a and other QSAR data and may help guide the design of additional novel species.

Conclusions

The results we have obtained above are of interest for a number of reasons. First, we have used 3D-QSAR/MFA techniques to predict the activities of a wide range of aryl-X (X = alkyl, oxyalkyl, and sulfanylalkyl) substituted bisphosphonates (pamidronates and one alendronate). Activities can be predicted with an average error of $2.1\times$ over an overall range in activity of $4000\times$, when using crystallographic structural information to guide the overall molecular alignment. Second, we have used MFA techniques to predict the activities of a wide range of heterocyclic bisphosphonates. Activities can be predicted within a factor of $2.5\times$ over a $2000\times$ range in activity. Third, we propose that, in general, the activities of heterocyclic bisphosphonates are strongly related to the pK_a values of the (parent) heterocycle. Thiazole, triazole, and pyrazole-based bisphosphonates have low activity since the pK_a values of the parent bases are all low, typically <3. On the other hand, pyridyl-, imidazolyl-, and aminothiazole species are all active since they have (free base) pK_a values in the range ~5−9 and can thus be protonated and therefore act as carbocation reactive-intermediate analogue FPP synthase inhibitors. For aminoimidazole bisphosphonates, activity is surprisingly low, but this may be correlated with an extremely high pK_a (~11) which could result in diminished transport. Fourth, we have used the Catalyst pharmacophore modeling approach to develop pharmacophore hypotheses for bone resorption by bisphosphonates. The hypotheses consist of two negative ionizable groups, a positive charge feature and an aromatic

feature, but with different spatial dispositions of the aromatic feature in the two classes of compounds. Activity predictions using this approach are slightly worse than when using the MFA approach, but are still quite promising. Fifth, we used a CoMSIA analysis to investigate the various contributions to activity in both the aryl-X and heterocyclic bisphosphonates, finding at least initial evidence for the importance of the 1-OH "bone hook".

These results represent the most comprehensive quantitative structure−activity relationship investigation of bisphosphonates as bone resorption agents to date and demonstrate that activities can be predicted within about a factor of 2.3 (over an overall range in activity of ~4000×). In addition, for heterocyclic bisphosphonates, the results are of considerable interest since they provide the first insights into structure−activity relationships at the level of protonation state or pK_a values. This can be expected to facilitate the further development of novel bisphosphonates and related systems for a wide variety of therapeutic uses.

References

(1) Widler, L.; Jaeggi, K. A.; Glatt, M.; Muller, K.; Bachmann, R.; Bisping, M.; Born, A. R.; Cortesi, R.; Guiglia, G.; Jeker, H.; Klein, R.; Ramseier, U.; Schmid, J.; Schreiber, G.; Seltenmeyer Y.; Green J. R. Highly potent geminal bisphosphonates. From pamidronate disodium (Aredia) to zoledronic acid (Zometa). *J. Med. Chem.* **2002**, *45*, 3721−3738.
(2) imshealth.com. Osteoporosis: the bare bones. http://www.ims-global.com/insight/news_story/0211/news_story_021106.htm.
(3) Martin, M. B.; Grimley, J. S.; Lewis, J. C.; Heath, H. T., III; 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*, 909−916.
(4) Kunzmann, V.; Bauer, E.; Wilhelm, M. $\gamma\delta$ T-cell stimulation by pamidronate. *N. Engl. J. Med.* **1999**, *340*, 737−738.

(5) Espinosa, E.; Belmant, C.; Sicard, H.; Poupot, R.; Bonneville, M.; Fournié, J.-J. Y2K+1 state-of-the-art on non-peptide phosphoantigens, a novel category of immunostimulatory molecules. *Microbes Infect.* **2001**, *3*, 645−654.

(6) Das, H.; Wang, L.; Kamath, A.; Bukowski, J. F. Vγ2Vδ2 T-cell receptor-mediated recognition of aminobisphosphonates. *Blood* **2001**, *98*, 1616−1618.

(7) Rodriguez, N.; Bailey, B. N.; Martin, M. B.; Oldfield, E.; Urbina, J. A.; Docampo, R. Radical cure of experimental cutaneous leishmaniasis by the bisphosphonate pamidronate. *J. Infect. Diseases* **2002**, *186*, 138−140.

(8) 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*, 929−931.

(9) Wang, L.; Kamath, A.; Das, H.; Li, L.; Bukowski, J. F. Antibacterial effect of human Vγ2Vδ2T cells *in vivo*. *J. Clin. Invest.* **2001**, *108*, 1349−1357.

(10) Kato, Y.; Tanaka, Y.; Miyagawa, F.; Yamashita, S.; Minato, N. Targeting of tumor cells for human γδ T cells by nonpeptide antigens. *J. Immunol.* **2001**, *167*, 5092−5098.

(11) Shipman, C. M.; Rogers, M. J.; Apperley, J. F.; Russell, R. G.; Croucher, P. I. Bisphosphonates induce apoptosis in human myeloma cell lines: a novel anti-tumour activity. *Br. J. Haematol.* **2001**, *113*, 841−842.

(12) Fleisch, H.; Russell, R. G.; Simpson, G.; Buhlbauer, R. C. Prevention by a diphosphonate of immobilization "osteoporosis" in rats. *Nature* **1969**, *223*, 211−212; Fleisch, H. Development of bisphosphonates. *Breast Cancer Res.* **2002**, *4*, 30−34.

(13) Luckman, S. P.; Hughes, D. E.; Coxon, F. P.; Graham, R.; Russell, G.; Rogers, M. J. Nitrogen-containing bisphosphonates inhibit the mevalonate pathway and prevent post-translation prenylation of GTP-binding proteins, including Ras. *J. Bone Miner. Res.* **1998**, *13*, 581−589.

(14) Cromartie, T. H.; Fisher, K. J.; Grossman, J. N. The discovery of a novel site of action for herbicidal bisphosphonates. *Pesticide Biochem. Physiol.* **1999**, *63*, 114−126.

(15) van Beek, E.; Pieterman, E.; Cohen, L.; Lowik, C.; Papapoulos, S. Farnesyl pyrophosphate synthase is the molecular target of nitrogen-containing bisphosphonates. *Biochem. Biophys. Res. Commun.* **1999**, *264*, 108−111.

(16) Keller, R. K.; Fliesler, S. J. Mechanism of aminobisphosphonate action: characterization of alendronate inhibition of the isoprenoid pathway. *Biochem. Biophys. Res. Commun.* **1999**, *266*, 560−563.

(17) Grove, J. E.; Brown, R. J.; Watts, D. J. The intracellular target for the antiresorptive aminobisphosphonate drugs in *Dictyostelium discoideum* is the enzyme farnesyl diphosphate synthase. *J. Bone Miner. Res.* **2000**, *15*, 971−981.

(18) 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*, 235−242.

(19) Rodan, G. A.; Reszka, A. A. Bisphosphonate mechanism of action. *Curr. Mol. Med.* **2002**, *2*, 571−577.

(20) Martin, M. B.; Arnold, W.; Heath, H. T., III; Urbina, J. A.; Oldfield, E. Nitrogen-containing bisphosphonates as carbocation transition state analogues for isoprenoid biosynthesis. *Biochem. Biophys. Res. Commun.* **1999**, *263*, 754−758.

(21) Dunford, J. E.; Ebetino, F. H.; Rogers, M. J. The mechanism of inhibition of farnesyl diphosphate synthase by nitrogen-containing bisphosphonates. http://www.ibmsonline.orgt/Bisphosphonates%20Abstracts%20(oral)%201.doc.

(22) Ebetino, F. H.; Rogers, M.; Dunford, J.; Lou, X.; Phipps, R.; Russell, R. G. G.; Barnett, B.; Correa, P.; Mieling, G. Modeling of bisphosphonate binding to farnesyl diphosphate synthase. http://www.ibmsonline.org/bisphosphonates %20Abstracts%20-(oral)%201.doc.

(23) Montalvetti, A.; Bailey, B. N.; Martin, M.; Severin, G.; Oldfield, E.; Docampo, R. Bisphosphonates are potent inhibitors of *Trypanosoma cruzi* farnesyl pyrophosphate synthase. *J. Biol. Chem.* **2001**, *276*, 33930−33937.

(24) Martin, M. B.; Sanders, J. M.; Kendrick, H.; de Luca-Fradley, K.; Lewis, J. C.; Grimley, J. S.; Van Brussel, E. M.; Olsen, J. R.; Meints, G. A.; Burzynska, A.; Kafarski, P.; Croft, S. L.; Oldfield, E. Activity of bisphosphonates against *Trypanosoma brucei rhodesiense*. *J. Med. Chem.* **2002**, *45*, 2904−2914.

(25) Ghosh, S.; Sanders, J. M.; Meints, G.; Gossman, W.; Morita, C. T.; Oldfield, E. Unpublished results.

(26) 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*, 2894−2903.

(27) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(28) Tarshis, L. C.; Proteau, P. J.; Kellogg, B. A.; Sacchettini, J. C.; Poulter, C. D. Regulation of product chain length by isoprenyl diphosphate synthases. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 15018−15023.

(29) Gossman, W. L.; Wilson, S. R.; Oldfield, E. Three hydrates of 1-hydroxy-2-(3-pyridinyl)ethylidene-1,1-bisphosphonate (Risedronate) consisting of one molecular and two ionic structures. *Acta Crystallogr.* **2003**, *C59*, m33−m35.

(30) Arfken, G. B.; Weber, H. J. *Mathematical Methods for Physicists*, 4th ed.; Academic Press: New York, 1995, p 446.

(31) Fletcher, R.; Reeves, C. M. Functional minimization by conjugate gradients. *Comput. J.* **1964**, *7*, 149−154.

(32) Acton, F. S. *Numerical Methods That Work*; Harper and Row: New York, 1970; Chapter 2.

(33) *Cerius² Modeling Environment*, version 4.6; Accelrys, Inc.: San Diego, CA, 2001.

(34) Rogers, D.; Hopfinger, A. J. Application of genetic function approximation to quantitative structure activity relationships and quantitative structure property relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854−866.

(35) Devijver, P. A.; Kittler, J. *Pattern Recognition: A Statistical Approach*; Prentice/Hall Int., London, 1982.

(36) Mason, R. L.; Gunst, R. F.; Hess, J. L. *Statistical Design and Analysis of Experiments;* John Wiley and Sons: New York, 1989; pp 243−245.

(37) Sitthithaworn, W.; Kojima, N.; Viroonchatapan, E.; Suh, D.-Y.; Iwanami, N.; Hayashi, T.; Noji, M.; Saito, K.; Niwa, Y.; Sankawa, U. Geranylgeranyl diphosphate synthase from *Scoparia dulcis* and *Croton sublyratus*. Plastid localization and conversion to a farnesyl diphosphate synthase by mutagenesis. *Chem. Pharm. Bull.* **2001**, *49*, 197.

(38) Dissociation Constants of Organic Acids and Bases. http://www.zirchrom.com/organic.htm.

(39) Hansen, L. D.; Baca, E. J.; Scheiner, P. Thermodynamics of proton ionization from some substituted, unsaturated, five-membered nitrogen heterocycles. *J. Heterocycl. Chem.* **1970**, *7*, 991−995.

(40) Advanced Chemistry Development (ACD) Software Solaris V4.67.

(41) Suzuki, F.; Yoshihiro, F.; Mizutani, H.; Funabashi, C.; Ikai, T. et al. N−Pyridylaminomethylenediphosphonsaureverbindungen. Ger. Offen. 2,831,578, 1979.

(42) Beiboer, S. H. W.; van den Berg, B.; Dekker: N.; Cos, R. C.; Verheij, H. M. Incorporation of an unnatural amino acid in the active site of porcine pancreatic phospholipase A(2). Substitution of histidine by 1,2,4-triazole-3-alanine yields an enzyme with high activity at acidic pH. *Prot. Eng.* **1996**, *9*, 345−352.

(43) Soumillion, P.; Fastrez, J. Incorporation of 1,2,4-triazole-3-alanine into a mutant of phage lambda lysozyme containing a single histidine. *Prot. Eng.* **1998**, *11*, 213-217.

(44) Ebetino, F. H.; Bayless, A. V.; Amburgey, J.; Ibbotson, K. J.; Dansereau, S.; Ebrahimpour, A. Elucidation of a pharmacophore for the bisphosphonate mechanism of bone antiresorptive activity. *Phosphorus, Sulfur Silicon Relat. Elem.* **1996**, *110*, 217−220.

(45) Takeuchi, M.; Sakamoto, S.; Kawamuki, K.; Kurihara, H.; Nakahara, H.; Isomura, Y. Studies on novel bone resorption inhibitors. II. Synthesis and pharmacological activities of fused aza-heteroarylbisphosphonate derivatives. *Chem. Pharm. Bull.* **1998**, *46*, 1703−1709.

(46) Catalyst, version 3.1; Molecular Simulations Inc.: San Diego, CA 1997.

(47) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130−4146.

(48) SYBYL version 6.8; Tripos Inc.: St. Louis, MO, 2001.