Internet Electronic Journal of Molecular Design **2004**, *3*, 622–650 ISSN 1538–6414
BioChem Press http://www.biochempress.com

Inter*net* Electronic Journal of Molecular Design

October 2004, Volume 3, Number 10, Pages 622–650

Editor: Ovidiu Ivanciuc

Proceedings of the Internet Electronic Conference of Molecular Design, IECMD 2003
November 23 – December 6, 2003
Part 5

Quantitative Structure–Activity Relationship Study of Bisphosphonates

Aihua Xie,[1,2] Chenzhong Liao,[1,2] Zhibin Li,[1] Zhiqiang Ning,[1] Weiming Hu,[1] Xianping Lu,[1] Leming Shi,[1] and Jiaju Zhou[2]

[1] Chipscreen Biosciences, Ltd., Research Institute of Tsinghua University, Suite C301, Shenzhen, Guangdong 518057, China
[2] Institute of Process Engineering, Chinese Academy of Sciences, P.O. Box 353, Beijing 100080, China

Received: October 10, 2003; Revised: April 5, 2004; Accepted: May 27, 2004; Published: October 31, 2004

Citation of the article:
A. Xie, C. Liao, Z. Li, Z. Ning, W. Hu, X. Lu, L. Shi, and J. Zhou, Quantitative Structure–Activity Relationship Study of Bisphosphonates, *Internet Electron. J. Mol. Des.* **2004**, *3*, 622–650, http://www.biochempress.com.

Internet Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

Quantitative Structure–Activity Relationship Study of Bisphosphonates [#]

Aihua Xie,[1,2,*] Chenzhong Liao,[1,2] Zhibin Li,[1] Zhiqiang Ning,[1] Weiming Hu,[1] Xianping Lu,[1] Leming Shi,[1] and Jiaju Zhou [2]

[1] Chipscreen Biosciences, Ltd., Research Institute of Tsinghua University, Suite C301, Shenzhen, Guangdong 518057, China

[2] Institute of Process Engineering, Chinese Academy of Sciences, P.O. Box 353, Beijing 100080, China

Received: October 10, 2003; Revised: April 5, 2004; Accepted: May 27, 2004; Published: October 31, 2004

Internet Electron. J. Mol. Des. **2004**, *3* (10), 622–650

Abstract

Motivation. Bisphosphonates (BPs) are most widely used as agents for treating osteoporosis. They have also been used for other purposes such as herbicides, anticancer agents, and antiparasitics. Here we report QSAR models of four BPs datasets based on the 118 structural and biological data we have collected from various literature sources.

Method. The structures were energy minimized with the MMFF94 force field, then 181 2D and 3D descriptors were calculated with MOE (Molecular Operating Environment). The step by step multiple regression and principle analysis were used to generate QSAR models. The predictive power of each QSAR model was estimated with the leave–one–out (LOO) cross–validation method.

Results. The QSAR model for a dataset of 28 GGPPSase inhibitors (r^2 = 0.86, r^2_{LOO} = 0.82, s = 0.45, s_{LOO} = 0.51, F = 77.56) is made up of two descriptors. Another dataset of 28 compounds with bioactivities against the growth of *T. Brucei rhodesiense* was studied using PCA and reached a model (r^2 = 0.85, r^2_{LOO} = 0.79, s = 0.30, s_{LOO} = 0.35, F = 32.03) with four principle components (PCs). Both the above models have comparable predictive ability with CoMFA model reported by Szabo *et al*. The 86 BPs provided by Novartis with *in vivo* bio–data of TPTX rats were divided into two datasets. A six PCs model (r^2 = 0.80, r^2_{LOO} = 0.72, s = 0.44, s_{LOO} = 0.53, F = 24.83) elucidated the dataset of 44 compounds in which containing aliphatic linked nitrogen atoms. The other dataset includes 42 BPs containing a heterocyclic moiety with at least one nitrogen atom. Its PCA model (r^2 = 0.80, r^2_{LOO} = 0.71, s = 0.46, s_{LOO} = 0.57, F = 19.99) consists of seven PCs.

Conclusions. A leave–four–out test procedure shows that though the QSAR models based on *in vivo* bone resorption pED_{50} values cannot provide explicit indications for drug design, their predictive ability for related compounds is quite good.

Keywords. Bisphosphonate; principal component analysis (PCA); GGPPSase inhibitor; quantitative structure–activity relationships (QSAR); IC_{50}; ED_{50}.

Abbreviations and notations

Bps, bisphosphonates

ED_{50}, the dose of compound administratered sc, which results in a 50% reduction of the hypercalcemia

r^2_{LOO}, leave–one–out cross validated correlation

[#] Presented in part at the Internet Electronic Conference of Molecular Design, IECMD 2003.

* Correspondence author; phone: 86–10–82612204; fax: 86–10–62561822; E–mail: ahxie@home.ipe.ac.cn.

induced in TPTX rats by 1,25–dihydroxyvitamin D_3
FPPSase, farnesyl pyrophosphate synthase
GGPPSase, geranylgeranyl diphosphate synthase
IC_{50}, experimental concentration required to reduce activity/proliferation of enzymes/cells/parasites by 50%
PCA, principal component analysis
PRED, leave–one–out cross validated prediction
QSAR, quantitative structure–activity relationships
r^2, correlation coefficient. $r^2 = 1 - \frac{\sum (y_{calc} - y_{act})^2}{\sum (y_{act} - y_{mean})^2}$
r_{LOO}^2, leave–one–out cross validated correlation coefficient. $r_{LOO}^2 = 1 - \frac{\sum (y_{pred} - y_{act})^2}{\sum (y_{act} - y_{mean})^2}$
s, root mean square error. $s = \sqrt{\frac{\sum (y_{\mathrm{calc}} - y_{\mathrm{act}})^2}{n-k-1}}$, n, number of observations; k, number of descriptors
TPTX, thyroparathyroidectomy
s_{LOO}, leave–one–out cross validated correlation root mean square error. $s_{LOO} = \sqrt{\frac{\sum (y_{pred} - y_{\mathrm{act}})^2}{n-k-1}}$, n, number of objects; k, number of descriptors

1 INTRODUCTION

Bisphosphonates (BPs) are the most widely used inhibitors of bone resorption. They all contain two phosphonate groups attached to a single carbon atom, forming a P–C–P structure. Bisphosphonates are stable analogs of naturally occurring pyrophosphate–containing compounds, which now helps to explain their intracellular as well as their extracelluar modes of action. Several bisphosphonates, *e.g.*, etidronate, clodronate, pamidronate, alendronate, tiludronate, risedronate, and ibandronate, have been established as effective treatments in clinical disorders such as Paget's disease of bone, tumour–associated bone disease, and osteoporosis [1]. Bisphosphonates have also been repeated for uses as herbcides [2], anticancer agents [3], and antiparasitics [4,5].

Recent studies suggest that bisphosphonates inhibit bone resorption by cellular effects on bone–resorbing osteoclasts, rather than by purely physicochemical mechanisms. It is likely that BPs are internalized by osteoclasts and interfere with specific biochemical process and induce apoptosis [6]. In recent work, the site of action has been narrowed down to the mevalonate pathway and the isoprene pathway.

The exact enzymes of the mevalonate pathway that are inhibited by BPs have not yet been fully identified. However, incadronate and ibandronate are known inhibitors of squalene synthase, an enzyme in the mevalonate pathway required for cholesterol biosynthesis [6]. Alendronate and pamidronate are less potent inhibitors of squalene synthase but can also inhibit sterol biosynthesis, suggesting that these bisphosphonates may inhibit up stream enzymes of the mevalonate pathway other than squalene synthase [7].

Several enzymes of the mevalonate pathway such as isoprenoid diphosphate isomerase (IPP isomerase), farnesyl diphosphate synthase (FPPSase), geranylgeranyl diphosphate synthase (GGPPSase), and squalene synthase, utilize an isoprenoid diphosphate as a substrate and thus are likely to have similar substrate binding sites. Thus if nitrogen–containing BPs act as substrate analogs of an isoprenoid diphosphate, it is likely that these BPs actually inhibit several enzymes of the mevalonate pathway. FPPSase are the most reported target for many BPs. For example,

Cromartie and Fisher demonstrated that herbicidal bisphosphonates were potent, low–nanomolar inhibitors of a daffodil FPPSase [2,8], and Grove *et al*. reported that BPs were growth and FPPSase inhibitors of the primitive eukaryote *Dictyostelium discoideum* [9].

Several groups [1,3,10,11] have reported that FPPSase was the target of the nitrogen–containing bisphosphonates in bone, leading to the apoptosis of osteoclasts. The group of Eric Oldfield, which did a lot of jobs on chemotherapy of parasitic protozoa diseases, reported that bisphosphonates were in vitro inhibitors of the growth of the causative agents of Chagas' disease, human East African trypanosomiasis, visceral leishmaniasis, toxoplasmosis, malaria, and cryptosporidiosis, *T. Cruzi*, *Trypanosoma brucei rhodesiense*, *Leishmania donovani*, *Toxoplasma gondii*, *Plasmodium falciparum*, *and Cryptosporidium parvum* [4,5]. They also showed that in some of the parasites, such as *T. b. rhodesiense* and *D. discoideum*, the molecular target of some bisphosphonates such as risedronate is FPPSase and reported 3D–QSAR/CoMFA investigation of bisphosphonate drugs, in the inhibition of bone resorption as well as the growth of *D. discoideum* and the bloodstream form of *T. b. rhodesiense* [5,12]. Though there are fewer reports, other enzymes, *e.g*. IPP isomerase, GGPP synthase, and squalene synthase of the mevalonate pathway, may be also potential targets for different bisphosphonates.

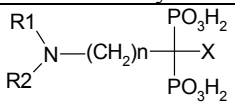
Eric Oldfield group has investigated the inhibition of a human recombinant GGPPSase by 23 bisphosphonates and six azaptenyl diphosphates. In addition to CoMFA analysis of structure–activity relationship, the pharmacophore of these GGPPSase inhibitors obtained from Catalyst was also provided [13].

Though the actual conformations of the bisphosphonates in the FPPSase and GGPPSase active sites are not yet known, good predictive CoMFA models were obtained using the molecular mechanics–derived lowest–energy conformers [5,12,13].

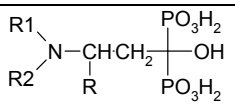
Widler *et al*. reported an extensive structure–activity relationship (SAR) study of bisphosphonates [14]. Small changes of the structure of pamidronate (compound **2**) lead to marked improvements of the inhibition of osteoclastic resorption potency. Alendronate (compound **3** in Table 1), with an extra methylene group in the *N*–alkyl chain, and olpadronate (compound **7**), the *N*,*N*–dimethyl analogue, are about 10 times more potent than pamidronate (compound **2**).

Extending one of the *N*–methyl groups of olpadronate to a pentyl substituent leads to ibandronate (compound **10**), which is the most potent close analogue of pamidronate. Even slightly better antiresorptive potency is achieved with derivatives having a phenyl group linked via a short aliphatic tether of three to four atoms to nitrogen, the second substituent being preferentially a methyl group. The most potent bisphosphonate, zoledronate (compound **65**), is found in the series containing a heteroaromatic moiety with at least one nitrogen atom, which is linked via a single methylene group to the geminal bisphosphonate unit [14].

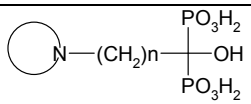
Table 1. Stucture and bioactivity of bisphosphonates



No	Cmpd code	R_1	R_2	X	n	ED_{50} (μg/kg)[a]	IC_{50}(μM)[b]	IC_{50}(μM)[c]
1	Novartis 1a	H	H	OH	1	150		
2	Pamidronate	H	H	OH	2	61	177	180
3	Alendronate	H	H	OH	3	8		440
4	Novartis 1d	H	H	OH	4	20		
5	Neridronate	H	H	OH	5	60	31.7	690
6	Novartis 1g	Me	H	OH	2	15		
7	Olpadronate	Me	Me	OH	2	12	5.4	
8	T.B. 009	propyl	Me	OH	2	3	7.8	330
9	Novartis 1j	Et	Et	OH	2	3		
10	Ibandronate	pentyl	Me	OH	2	1.1	0.96	83
11	Novartis 1l	Me	Me	H	2	100		



No	Cmpd code	R_1	R_2	R	ED_{50} (μg /kg)[a]	IC_{50}(μM)[b]	IC_{50}(μM)[c]
12	Novartis 1n	H	H	Me	3.4		
13	Novartis 1o	Me	Me	Me	18		
14	Novartis 1p	pentyl	Me	Me	65		



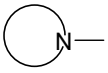
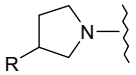
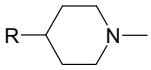
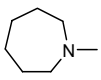
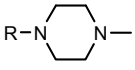
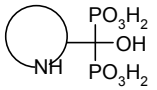
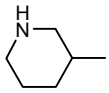
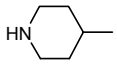
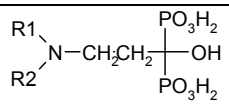
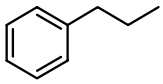
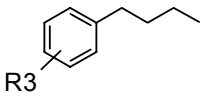
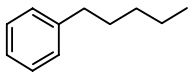
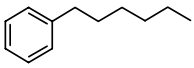
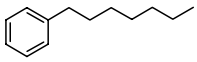
No	Cmpd code		R	n	ED_{50} (μg /kg)[a]
15	Novartis 2a		H	2	10
16	Novartis 2b		H	3	25
17	Novartis 2c		H	5	250
18	Novartis 2d		Ph	2	70
19	Novartis 2e		4–Cl–Ph	2	3.5
20	Novartis 2f		H	2	5.6
21	Novartis 2g		Ph	2	11
22	Novartis 2h		Ph	3	100
23	Novartis 2j		3–F–Ph	2	30
24	Novartis 2k			2	25
25	Novartis 2m		Me	2	400

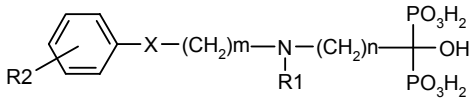
Table 1. (Continued)



Serial	Cmpd code		ED$_{50}$ (μg /kg)[a]
26	Novartis 3a		50
27	Novartis 3b		250
28	Novartis 3c		2500



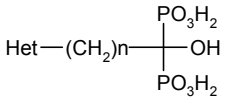
Serial	Cmpd code	R$_1$	R$_2$	R$_3$	ED$_{50}$ (μg /kg)[a]
29	Novartis 4a		H		300
30	Novartis 4b		Me		1.4
31	Novartis 4c		H	H	20
32	Novartis 4d		Me	H	1
33	Novartis 4e		Et	H	15
34	Novartis 4f		Me	3–Me	1.5
35	Novartis 4g		Me	4–Cl	0.7
36	Novartis 4i		H		1.0
37	Novartis 4j		Me		0.4
38	Novartis 4k		Me		20
39	Novartis 4l		Me		1500

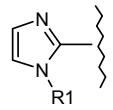
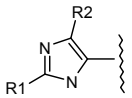
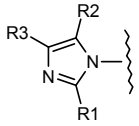
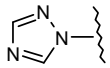


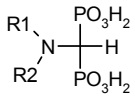
Serial	Cmpd code	X	R$_1$	R$_2$	m	n	ED$_{50}$ (μg /kg)[a]
40	Novartis 5a	O	Me	H	2	2	1.5
41	Novartis 5b	O	Me	4–Cl	2	2	1.7
42	Novartis 5c	O	H	H	3	2	1.2
43	Novartis 5d	O	Me	H	3	2	0.5
44	Novartis 5e	O	Me	3–Me	3	2	1.7
45	Novartis 5f	O	Me	4–F	3	2	0.6
46	Novartis 5g	O	Me	4–Cl	3	2	1.3
47	Novartis 5h	O	Et	4–MeO	3	2	1.2
48	Novartis 5i	O	propyl	H	3	2	20
49	Novartis 5j	O	butyl	H	3	2	10
50	Novartis 5k	O	Me	H	4	2	500

Table 1. (Continued)

Serial	Cmpd code	X	R_1	R_2	m	n	ED_{50} (μg /kg)[a]
51	Novartis 5l	O	Me	H	6	2	4
52	Novartis5m	O	Me	H	3	2	7500
53	Novartis 5n	O	Me	H	2	3	100
54	Novartis 5p	S	Me	H	2	2	0.7
55	Novartis 5q	S	H	H	3	2	7
56	Novartis 5r	S	Me	H	3	2	0.33
57	Novartis 5s	S	Me	4–Cl	3	2	7.8



Serial	Cmpd code	Het	R1	R2	R3	n	ED50 (μg /kg)[a]
58	Novartis 6a		H			1	5
59	Novartis 6b		Me			1	0.6
60	Novartis 6c		Bz			1	25
61	Novartis 6d		H	H		1	0.3
62	Novartis 6e		H	H		2	20
63	Novartis 6f		Me	H		1	15
64	Novartis 6h		H	Me		1	1.5
65	Zoledronate		H	H	H	1	0.07
66	Novartis 6j		H	H	H	2	45
67	Novartis 6k		Me	H	H	1	3
68	Novartis 6l		H	Me	Me	1	1.5
69	Novartis 6n					1	600



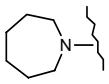
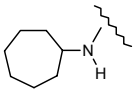
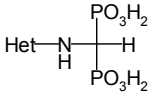
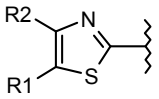
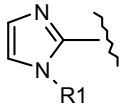
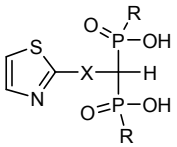
Serial	Cmpd code	R_1R_2N	ED_{50} (μg /kg)[a]
70	Novartis 7c		800
71	Novartis 7d		40
72	Novartis 7e		7

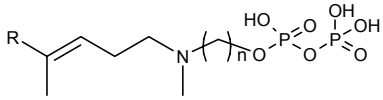
Table 1. (Continued)

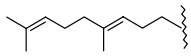
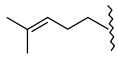
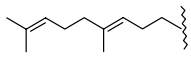
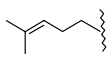
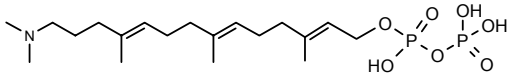


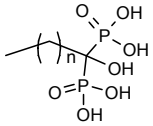
Serial	Cmpd code	Het	R_1	R_2	ED_{50} (μg /kg)[a]
73	Novartis 8a		H	H	5
74	Novartis 8b		H	Me	100
75	Novartis 8c		Me	H	1.5
76	Novartis 8d		Et	H	1.5
77	Novartis 8e		Pr	H	2
78	Novartis 8f		Bu	H	0.9
79	Novartis 8g		Pr	H	200
80	Novartis 8h		$PhCH_2CH_2$	H	2.7
81	Novartis 8j		H		500
82	Novartis 8k		Me		5
83	Novartis 8l		$PhCH_2$		75
84	Novartis 8m		Ph		200



Serial	Cmpd code	X	R	ED_{50} (μg /kg)[a]
85	Novartis 9a	CH_2	OH	200
86	Novartis 9b	S	OH	700



Serial	Cmpd code	R	n	IC_{50} (μM)[c]
87	3–azaGGPP		2	0.14
88	3–azaFPP		2	0.74
89	3–azaGPP	Me	2	240
90	3–azahomoGGPP		3	0.37
91	3–azahomoFPP		3	0.31
92	15–azaGGPP			>>100



Serial	Cmpd code	n	IC_{50} (μM)[b]	IC_{50} (μM)[c]
93	T.B. 024	2	92.0	620
94	T.B. 025	3	99.8	200
95	T.B. 023	4	62.4	53
96	GGPP018	5		11.0

Table 1. (Continued)

Serial	Cmpd code	n	IC_{50} (μM)[b]	IC_{50} (μM)[c]
97	GGPP017	6		4.3
98	T.B. 014	8	20.5	0.72
99	T.B. 010	9	8.0	1.4
100	T.B. 007	10	2.0	0.92

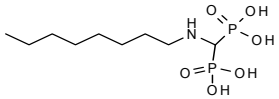
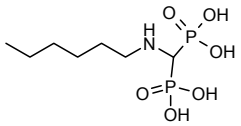
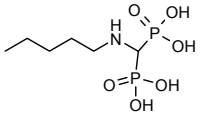
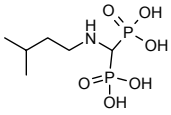
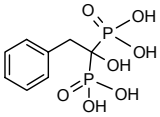
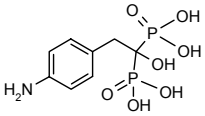
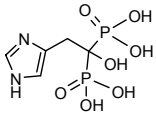
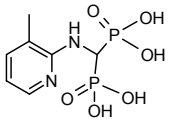
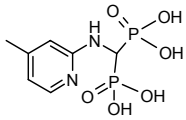
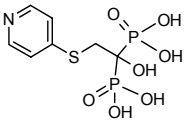
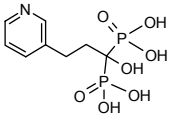
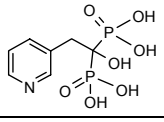
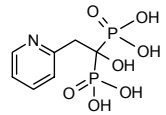
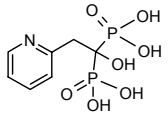
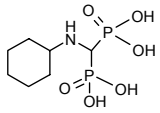
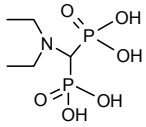
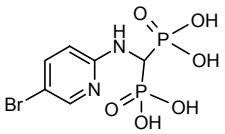
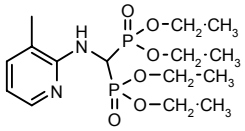
Serial	Cmpd code	Structure	IC_{50} (μM)[b]	IC_{50} (μM)[c]
101	T.B. 006		1.7	2.2
102	GGPP031			19.0
103	T.B. 021		50.6	
104	T. B. 026		102	
105	T.B. 016		21.3	220.0
106	T.B. 020		40.0	180.0
107	T.B. 012		8.6	220.0
108	NE97220		0.7	220.0
109	N–(2–(4–picolyl))ADMP		0.61	260.0
110	T.B. 013		19.8	550.0
111	Homorisedronate		1.7	410.0
112	Risedronate		8.6	350.0

Table 1. (Continued)

Serial	Cmpd code	Structure	IC_{50} (μM)[b]	IC_{50} (μM)[c]
113	NE58018		0.22	
114	N–(2–(5–chloro)–pyridyl)AMDP		53.3	
115	T.B. 015		20.9	
116	T.B. 018		34.4	
117	T.B. 019		39.5	
118	T.B. 2–13		27.9	

[a] the dose of compound administratered sc, which results in a 50% reduction of the hypercalcemia induced in TPTX rats by 1,25–dihydroxyvitamin D_3 [14]
[b] experimental concentration required to reduce proliferation of *T. Brucei rhodesiense* by 50% [4,5]
[c] experimental concentration required to reduce activity of GGPPSase by 50% [13].

The comprehension of BPs mechanism gives us indications to investigate the quantitative structure–activity relationship of the bisphosphonates provided by Novartis [14] with *in vivo* ED_{50} against hypercalcemia induced in thyroparathyroidectomy (TPTX) rats. The total of 86 compounds were divided into two datasets, one for the series containing a heterocyclic moiety, which contains at least one nitrogen atom; the other for bisphosphonates that contains a nitrogen atom in aliphatic link and do not possess a heterocyclic substitute. The two datasets were analyzed using QSAR module of MOE and achieved two predictive models through principal component analysis.

We also investigated the BPs with IC_{50} against *T. Brucei Trypomastigotes* [4,5] and BPs with IC_{50} for GGPPSase inhibition [13] using the molecular modeling package MOE respectively, and achieved more simple and lightening models.

2 MATERIALS AND METHODS

2.1 Chemical Data

The structures we have collected here are listed in the Table 1 along with compound code and bioactivities. Some of the compound codes were assigned following their traditional name such as

Alendronate and Pamidronate, or codes from original references such as NE58018 and NE97220. The others were assigned according original activity source such as T.B.006 and GGPP031, or data provider such as Novartis1a and Novartis1d.

Dataset1 are made up of 28 BPs with IC_{50} values against GGPPSase. The library covers many diverse structural features: ionic bisphosphonate and diphosphate groups; alkyl, alkenyl (prenyl), aryl, and heteroaryl side chains; 1–OH– and 1–H–bearing bisphosphonates; and nitrogen–containing or nitrogen–free side chains, together with different location of the side chain nitrogens. The pIC_{50} values of this dataset vary from 3.16 to 6.85, with a mean value of 4.49 and a SD of 1.23. The distribution of activity of this dataset is shown in Figure 1.

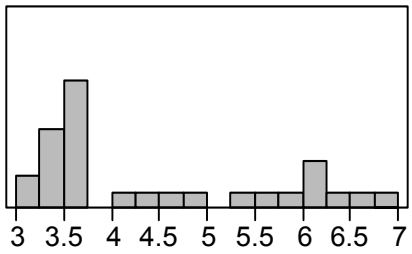


Figure 1. pIC_{50} distribution of dataset1.

Dataset2 include 28 bisphosphonates and their IC_{50} values against the growth of *T. Brucei rhodesiense* that is one of the causative agents of human African trypanosomiasis (sleeping sickness) [5]. The FPPSase is considered at least the main target of nitrogen–containing bisphosphonates in *T. Brucei rhodesiense* [4,5]. The pIC_{50} values of this dataset vary from 3.75 to 6.66; with a mean value of 4.91 and a SD of 0.79. The distribution of activity of this dataset is shown in Figure 2.

The structures and activity data of dataset3 and dataset4 are both from Novartis pharma research [14]. The ED_{50} values in the two datasets are the doses of compound administratered sc, which results in a 50% reduction of the hypercalcemia induced in TPTX rats by 1,25–dihydroxyvitamin D_3.

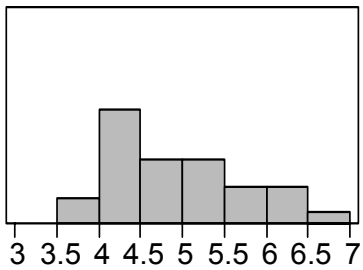


Figure 2. pIC_{50} distribution of dataset2.

We have known from the Introduction that the *in vivo* effect of bisphosphonates involves several enzymes of the mevalonate pathway e.g. IPP isomerase, FPP synthase, GGPP synthase, and squalene synthase. Therefore, the total of 86 compounds from Novartis were divided into two datasets according their structural features and rough speculation on their mode of action.

Dataset3 includes 44 bisphosphonates that contain a nitrogen atom in aliphatic link and do not possess nitrogen–containing heterocyclic substitutes. These compounds are less potent inhibitors of FPPSase and are speculated to act mainly with GGPPSase. The pED_{50} values of this dataset vary from 5.12 to 9.48, with a mean value of 8.13 and a SD of 0.98. The distribution of activity of this dataset is shown in Figure 3.

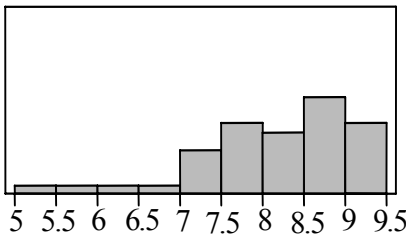


Figure 3. pED_{50} distributions of dataset3.

Dataset4 includes 42 bisphosphonates containing a heterocyclic moiety, which contains at least one nitrogen atom. Some of these BPs are more potent antiresorptive agents in the *in vivo* experiment and more potent FPPSase inhibitors *in vitro*. The pED_{50} values of this dataset vary from 5.60 to 10.16; with a mean value of 7.66 and a SD of 1.06. The distribution of activity of this dataset is shown in Figure 4.

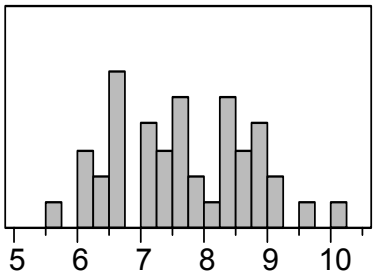


Figure 4. pED_{50} distributions of dataset4

The IC_{50} or ED_{50} values and the respective negative logarithm (pIC_{50} or pED_{50}) for all compounds are listed in the tables of supplementary materials along with model predictions. The stronger inhibitor a compound is, the greater the pIC_{50} or pED_{50} is.

2.2 Previous QSAR Models

Using the comparative molecular field analysis (CoMFA), Szabo *et al*. [13] obtained a fairly good model for dataset1:

$$pIC_{50} = 3.55 + 0.44\times"CH_3/473"-0.37\times"CH_3/330"-0.01\times"H^+/318"$$
$$n = 28, \quad r^2 = 0.938, \quad r^2{}_{LOO} = 0.900, \quad F = 86.6 \tag{1}$$

Martin *et al*. [5] obtained another CoMFA model for dataset2:

$$pIC_{50} = 4.60 - 0.06"H^+/187"+0.02"H^+/237"+0.05"H^+/186"+0.01"H^+/185"$$
$$n = 26, \quad r^2 = 0.87, \quad r^2_{LOO} = 0.79, \quad F = 34.80 \tag{2}$$

where H^+/i represents the interaction energy between a proton probe and the molecule at the gridpoint *i*, and CH_3/j represents the interaction energy between a methyl probe and the molecule at the grid point *j*.

2.3 Molecular Modeling

The structures and biological activity data were stored in an ISIS/Base database from which an SD file was exported. The SD file was imported into a molecular modeling package (MOE) for subsequent calculations. The molecular structures were optimized using MMFF94 force field. All the 181 2D and inner 3D descriptors available in MOE [15] were calculated for every molecule. The QuaSAR–Contingency module was used to prune the descriptors in order to select an optimum subset for QSAR. The Qua–cluster module of MOE was used to evaluate the diversity of the collection of our molecules based on the table of selected molecular descriptors and assigned weights to molecules if necessary. JMP4.5 (SAS Institute) [16] was used to perform most of the statistical analyses reported in this study.

2.4 Structure Descriptors

The 181 descriptors calculated in MOE include 2D and internal 3D descriptors.

2D descriptors only use atom and connection information for the calculation, and no 3D coordinates or individual conformations are needed. The 2D descriptors include physical properties such as atom counts and bond counts, mr, logP and vdw_area etc.; subdivided surface areas that are based on an approximate accessible van der Waals surface area calculation for each atom, v_i, along with some other atomic property, p_i (the v_i are calculated using a connection table approximation) [17]; Kier and Hall connectivity and Kappa Shape indices [18]; adjacency and distance matrix descriptors [19–21]; pharmacophore features (*e.g*. donor, acceptor, polar, positive, negative, hydrophobic.) descriptors; and two sets of partial charges; one set was calculated from Partial Equalization of Orbital Electronegativities method [22] and the other set was previously stored forcefield (MMFF94) partial charges.

Internal 3D descriptors use 3D coordinate information but independent on the rotations and translations of the conformation. The internal 3D descriptors include potential energy descriptors, surface area, volume and shape descriptors, and conformation–dependent charge descriptors. The potential energy descriptors such as value of the potential energy (E), electrostatic component of potential energy (E_ele), and solvation energy (E_sol) etc. were calculated using MMFF94 forcefield and corresponding default potential set up in MOE. Surface area, volume, and shape descriptors such as water accessible surface area (ASA); Van der Waals volume (vol) and Van der Waals surface area (VSA) etc. depend on the structure connection and conformation. Conformation–dependent charge descriptors such as water accessible surface area of all atoms with positive charges (ASA+), water accessible surface area of all atoms with negative charges (ASA–), and water accessible surface area of all hydrophobic atoms (ASA_H) etc. depend upon partial charges and conformations. Water accessible surface area was calculated using a radius of 1.4 angstroms for the water molecule.

2.5 Computational Methods

The QSAR modeling process consists of the following steps: structure optimization using MMFF94 force field; evaluation of chemical structure descriptors; descriptor pruning through QSAR–contingency, correlation analysis of descriptors, step–forward and step–backward selection of descriptors; structural diversity analysis of the dataset based on pruned descriptor set and assigned weight to molecules if necessary; multiple regression analysis between pIC_{50} and selected descriptors; evaluation of the significance level of the model and each determined descriptor; validation and cross–validation (leave–one–out procedure) of the model; detection of outliers and modification of QSAR–model; interpretation of the model equation.

MOE detects outliers with Grubbs test. The first step is to quantify how far away the experimental pIC_{50} is from the model value, by calculating the ratio Z–SCORE, defined as the difference between the pIC_{50} and model value divided by the SD of the whole dataset. MOE provides Z–SCORE values for all molecules and considers molecules with a Z–SCORE of 2.5 or more to be possible outliers.

Grubbs and others have tabulated critical values for Z–SCORE which are tabulated below for p = 0.05/0.02 (two tails) [24]. The critical value increases with sample size. Thus instead of simply taking the MOE criteria of outlier detection, we consulted the Grubbs table of Z–SCORE for different sample sizes for detecting outliers, and considered the complex influence of the PCA method, take the values of p = 0.02 as criteria.

Model adequacy was measured as the square of correlation coefficient (r^2), root mean square error (s), cross–validated r^2 (r^2_{LOO}) and cross–validated s (s_{LOO}).

3 RESULTS AND DISCUSSION

3.1 QSAR Model for dataset1

After structure optimization, 181 descriptors were selected and evaluated from MOE descriptor selection panel. After descriptor pruning procedures, two descriptors were selected to build the final QSAR model for the data set. ASA denotes the water accessible area calculated using a radius of 1.4 angstroms for the water molecule, while PEOE_VSA–1 denotes the sum of van der Waals surface areas of the atoms whose PEOE partial charge is in the range of [–0.10, –0.05]. PEOE (Partial Equalization of Orbital Electronegativities) [22] method of calculating atomic partial charges is a method in which charge is transferred between bonded atoms until equilibrium. Diversity analysis based on the two descriptors showed that there was no need to assign weight to the molecules. The two–descriptor linear model is shown in Eq. (3):

$$\begin{gathered} pIC_{50} = 0.51396 + 0.00675 \times (ASA) + 0.01742 \times (PEOE\_VSA - 1) \\ r^2 = 0.86,\ \ s = 0.45,\ \ r^2_{LOO} = 0.82,\ \ s_{LOO} = 0.51,\ \ n = 28,\ \ F = 77.56,\ \ k = 2 \end{gathered} \tag{3}$$

ASA and PEOE_VSA–1 are all positively correlated with pIC_{50} values, thus increasing ASA and PEOE_VSA–1 will lead to the improvement of pIC_{50}. The parameter effect tests for the model show that ASA is the determined descriptor in the model (Table 2). The 3D–QSAR/CoMFA analysis carried out by Szabo *et al*. [13] indicates that van der Waals interactions are very important in GGPPSase inhibition. Our model revealed the importance of water accessible surface area, which is mainly responsible for the van der Waals interactions between BPs and GGPPSase enzyme. Though our model did not provide 3D information like the CoMFA model, it offers a much simple equation and fast method to gain insight into the GGPPSase inhibitor system.

Table 2. Effect tests of the descriptors for Eq. (3)

Descriptor	Correlation to pIC_{50} (r^2)	Sum of Squares	F Ratio	Prob > F
ASA	0.76	14.61	64.78	<0.0001
PEOE_VSA–1	0.50	4.18	18.53	0.0002

The leave–one–out cross–validated predictive pIC_{50} values (PRED) were listed in Table 1 of supplementary material and plotted in Figure 5.

To test the predictive ability of our model, we also removed three compounds from the training set and performed the whole QSAR procedure on the reduced training set; then using the resulting model to predict the activities of the three excluded compounds. This procedure was repeated three times using different test sets, and the predicted pIC_{50} values are listed in bold in Table 1 of supplementary material along with individual training sets and all statistical data for QSAR equations.

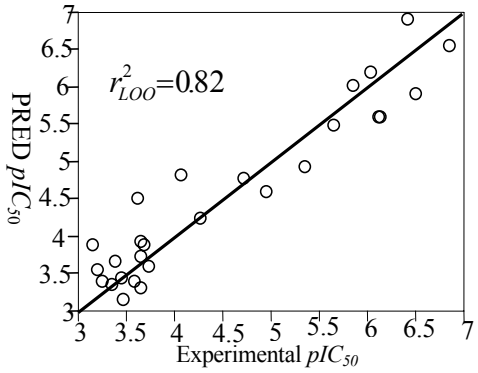


Figure 5. Leave–one–out cross–validated prediction versus experimental pIC_{50} values for dataset1.

The three compounds in each test set were chosen following the Ref. [13] in order to compare the model predictive ability with that of the CoMFA model performed by Szabo *et al*.. The graphical result of the total nine compounds test set is shown in Figure 6. The rms error in predicted pIC_{50} of the test set compounds is 0.44, the correlation coefficient between experimental and predicted values is $r^2 = 0.80$.

The QSAR equations for the three training sets with reduced size are as follows:

$$pIC_{50} = 0.5489 + 0.006720 \times (ASA) + 0.01747 \times (PEOE\_VSA - 1)$$
$$r^2 = 0.85, \quad s = 0.47, \quad r^2_{LOO} = 0.81, \quad s_{LOO} = 0.54, \quad n = 25, \quad F = 64.53, \quad k = 2 \tag{4}$$

$$pIC_{50} = 0.8793 + 0.005812 \times (ASA) + 0.02083 \times (PEOE\_VSA - 1)$$
$$r^2 = 0.88, \quad s = 0.40, \quad r^2_{LOO} = 0.85, \quad n = 25, \quad F = 83.83, \quad k = 2 \tag{5}$$

$$pIC_{50} = 0.4741 + 0.006827 \times (ASA) + 0.01726 \times (PEOE\_VSA - 1)$$
$$r^2 = 0.85, \quad s = 0.46, \quad r^2_{LOO} = 0.81, \quad s_{LOO} = 0.54, \quad n = 25, \quad F = 64.81, \quad k = 2 \tag{6}$$

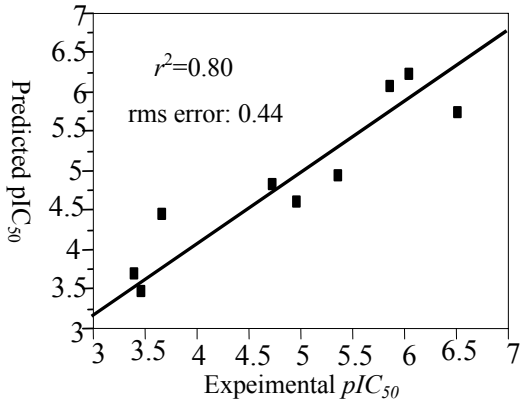


Figure 6. Predicted pIC_{50} values versus experimental pIC_{50} values for 9 GGPPSase inhibitors test set.

The comparison of Eq. (3) and the CoMFA model (Eq. (1)) reported by Szabo *et al*. [13] of the dataset1 is listed in Table 3. The rms error value between predicted and experimental values of the

test set is 0.39 for the CoMFA model and 0.44 for Eq. (3). Then, to compare the predictive ability of the two models, we can calculate the $F_{9,9} = 1.27$ from the rms error values and look up the $F_{0.05;\,9,9} = 3.18$ from the F distribution. The result of the F test tells us the predictive ability of the two models has no significant difference at $\alpha = 0.05$.

Table 3. Statistical comparison of model (3) from the current study and model (1) reported by Szabo *et al.* [13]

Model	r^2	r^2_{LOO}	n^a	k^b	F	rms error	Test r^2
Model (1)	0.938	0.90	28	3	86.8	0.39	0.88
Model (3)	0.86	0.82	28	2	77.56	0.44	0.80

[a] number of observations. [b] number of descriptors for certain model.

3.2 QSAR Model for Dataset2

Firstly, the QSAR–contingency, correlation analysis, step–forward and step–backward selection procedures recommended 11 descriptors for the model of dataset2. Some of the descriptors such as a_nH, apol, and KierFlex are correlated with (coefficient > 0.8) and irreplaceable by each other in the model. So many descriptors make the model complicated and difficult to interpret. And a model of 11 descriptors for a 28–observation dataset is sure over–fitting. In order to obtain a more robust and concise model, we performed principal components analysis (PCA) to reduce the dimensions of the descriptor subset, but failed.

We tried to select another subset among 181 descriptors. The element of the subset was measured mainly by its contribution to r^2. Finally we obtained a 32–descriptor subset, which keeps most interpretive information for pIC_{50} and have the fewest number of descriptors at the same time. The statistical parameters of the model based on the 32 descriptors are: $r^2 = 1.00$, $s = 0.00$. The names of the 32 descriptors are listed in Table 5 of supplementary materials.

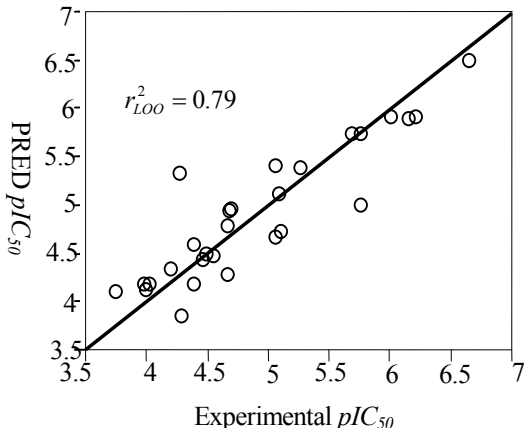


Figure 7. Leave–one–out cross–validated prediction versus experimental pIC_{50} values for dataset2.

Then we transformed the 32 descriptors into a set of uncorrelated and normalized variables using PCA. To capture 100% of the variance in the previous 32–descriptor subset, 26 principal components (PCs) are needed. The accumulative percentage of variance explained by the first five PCs is 81.38%; with the 1st PC explaining 34.87%, the 2nd 16.23%, 3rd 13.22%, 4th 6.01%, and 5th 5.05%.

After stepwise selection, four PCs (PC_2, PC_3, PC_4, PC_5) were determined to best describe the tendency of pIC_{50}. We obtained the following linear model:

$$\begin{aligned} pIC_{50} &= 4.9108 + 0.2218 \times PC_2 - 0.4067 \times PC_3 - 0.5010 \times PC_4 + 0.1535 \times PC_5 \\ r^2 &= 0.85, \quad r = 0.30, \quad r^2_{LOO} = 0.79, \quad s^2_{LOO} = 0.35, \quad n = 28, \quad F = 32.03, \quad k = 4 \end{aligned} \tag{7}$$

The leave–one–out cross–validated predictive pIC_{50} values were listed in Table 2 of the supplementary materials and plotted in Figure 7.

We then carried out the leave–three–out procedure just as we did on model (3) to test whether the PCA model have predictive value. The selection of test compounds followed Martin *et al*. [5] on the CoMFA model (Eq. (2)). The results for three training–test sets of calculations are given in Table 2 of supplementary materials. The graphical representation of the results is shown in Figure 8. The rms error for the test set compounds was 0.66, and the correlation coefficient between experimental and predicted pIC_{50} values was $r^2 = 0.7$ (Test r^2). The results indicate that Eq. (7) predicts the test set quite well and is not over fitting for the training set.

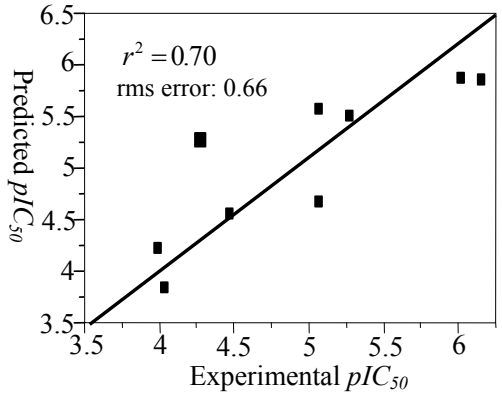


Figure 8. Predicted pIC_{50} values versus experimental pIC_{50} values for 9–compound test set of dataset2.

The comparison of PCA model (Eq. (7)) and the CoMFA model (Eq. (2)) [5] of the dataset2 is listed in Table 4. The rms error value of the test set (Test RMSE) is 0.32 for CoMFA (Eq. (2)) and 0.66 for Model (7). Then, $F_{9,9} = 4.25$ is larger than the boundary value $F_{0.05;\,9,9}$=3.18. It seems that model (7) is inferior to CoMFA model in predictive ability. However, compared to r^2 and r^2_{LOO}, Test r^2 value for the CoMFA model seems artificially high. General trend should be Test r^2 < $r^2_{LOO} < r^2$ according to statistical principle. This may be resulted from chance correlation of the test

compounds to the CoMFA model (Eq. (2)). Therefore we cannot claim that predictive ability of the two models has significant difference at α = 0.05.

Table 4. Statistical comparison of Eq. (7) and CoMFA model (Eq. (2))

Model	r^2	r^2_{LOO}	n	k	F	rms error	Test r^2
Model (2)	0.87	0.79	26	4	34.80	0.32	0.87
Model (7)	0.85	0.79	28	4	32.03	0.66	0.70

Note: Dataset2 includes pamidronate (compound **2**) and T.B.2–13 (compound **118**) from Ref. [4], which did not include in CoMFA dataset [5].

3.3 QSAR model of dataset3

The biological complexity of dataset3 is much greater than those of dataset1 and dataset2. The normal descriptor selection procedure suggested 16 descriptors for the dataset, and the statistical parameters of the model based on the 16 descriptors are: $r^2 = 0.84,\ \ s = 0.40,\ \ r^2_{LOO} = 0.50,\ \ s_{LOO} = 0.85,\ \ n = 44,\ \ F = 8.36,\ \ k = 16$. The names of the 16 descriptors are listed in Table 5 of supporting materials. Principal component analysis was carried out on the 16 descriptors. 16 PCs are required to capture the 100% variance in the previous descriptor subset. The accumulative percentage of variance explained by the first five PCs is 91.06%; with the 1st PC explaining 62.48%, the 2nd 10.29%, 3rd 8.27%, 4th 5.56%, and 5th 4.47%. After stepwise selection, six PCs (PC_2, PC_6, PC_9, PC_{12}, PC_{14}, PC_{15}) were selected to build the final model:

$$pED_{50} = 8.10 - 0.35 \times PC_2 - 0.25 \times PC_6 + 0.23 \times PC_9 - 0.65 \times PC_{12} + 0.22 \times PC_{14} - 0.26 \times PC_{15}$$
$$r^2 = 0.80, \quad s = 0.44, \quad r^2_{LOO} = 0.72, \quad s^2_{LOO} = 0.53, \quad n = 44, \quad F = 24.83, \quad k = 6 \qquad (8)$$

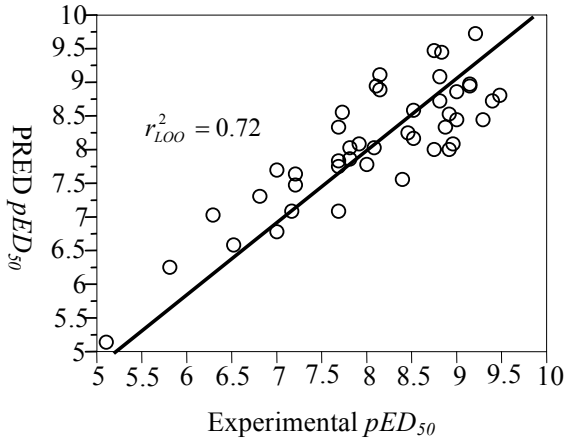


Figure 9. Leave–one–out cross–validated prediction versus experimental pED_{50} values for dataset3.

The percentage of variance explained by the 6 descriptors was listed in Table 5 respectively along with the result of parameter effect test of Eq. (8). The most correlative PC of Eq. (8) is PC_{12}

(F = 80.98), which only explains 0.12% variance of original descriptor subset. The PCA procedure succeeded in extracting useful information and getting rid of noisy information from original dataset.

Table 5. Effect tests of the descriptors for Eq. (8)

Source	Correlation to pED_{50} (r^2)	Sum of Squares	F Ratio	Prob > F	Percentage of variance (%)
PC_{12}	0.44	18.76	80.98	<.0001	0.12
PC_2	0.13	5.42	23.39	<.0001	10.29
PC_{15}	0.07	3.09	13.33	0.0008	0.01
PC_6	0.07	2.81	12.13	0.0013	3.56
PC_9	0.05	2.33	10.07	0.0030	1.31
PC_{14}	0.05	2.12	9.13	0.0045	0.02

The leave–one–out cross–validated predictive pIC_{50} values were listed in the Table 3 of supplementary materials and plotted in Figure 9.

To further investigate the predictive ability of this model, we removed four compounds from the training set randomly before recomputing the QSAR equation on the reduced dataset. The pED_{50} values of the removed compounds were predicted using the QSAR model derived from the reduced training set. The procedure was repeated four times and the predicted 16 pED_{50} values are given in the Table 3 of supplementary materials in bold and plotted in Figure 10. The graphical representation of the results is shown in Figure 10. The rms error between predicted pED_{50} and the experimental pED_{50} of the test set compounds was 0.34, and the correlation coefficient between experimental and predicted values is $r^2 = 0.91$. The quite good predictive result indicates that the PCA model (8) is robust and not seriously over fitting for the training set. Of course, general trend should be Test $r^2 < r^2_{LOO} < r^2$; the particularly high Test r^2 should be attributed to chance correlation.

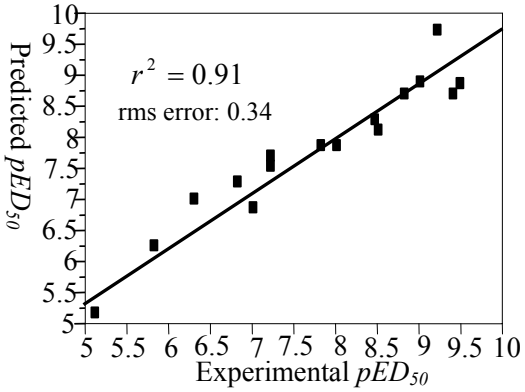


Figure 10. Predicted pED_{50} values versus experimental pED_{50} values for 16–compound test set of dataset3

3.4 QSAR for dataset4

15 descriptors were selected through normal descriptor selection procedure. The statistical parameters of the model based on the 15 descriptors are: $r^2 = 0.86$, $s = 0.39$, $r^2_{LOO} = 0.68$,

$s_{LOO} = 0.60,\ \ n = 42,\ \ F = 10.54,\ \ k = 15,$ The names of the 15 descriptors are listed in Table 5 of supplementary materials. Principal component analysis was carried out on the 15 descriptors. 15 PCs are required to capture the 100% variance of the previous descriptor subset. The accumulative percentage of variance explained by the first five PCs is 83.58%; with the 1st PC explaining 39.92%, the 2nd 18.85%, 3rd 9.86%, 4th 8.00%, and 5th 6.95%. After stepwise selection, seven PCs (PC_3, PC_7, PC_9, PC_{10}, PC_{12}, PC_{14}, PC_{15}) were selected to build the final model:

$$pED_{50} = 7.69 + 0.53 \times PC_3 + 0.15 \times PC_7 + 0.17 \times PC_9 - 0.35 \times PC_{12} - 0.24 \times PC_{14} + 0.24 \times PC_{10} + 0.56 \times PC_{15} \quad (9)$$

$$r^2 = 0.80, \quad s = 0.46, \quad r^2_{LOO} = 0.71, \quad s_{LOO} = 0.57, \quad n = 42, \quad F = 19.99.$$

The percentage of variance explained by the seven descriptors was listed in Table 6 respectively, along with the result of parameter effect test of Eq. (9). The most interpretive PC of the Eq. (9) is PC_{15} (F = 49.94), which only explains 0.01% variance of original descriptor subset. The PCA procedure also succeeded in extracting useful information and getting rid of noisy information from original dataset.

Table 6. Effect Tests of the descriptors for Eq. (9)

Source	Correlation to pED_{50}	Sum of Squares	F Ratio	Prob > F	Percentage of variance (%)
PC_{15}	0.29	13.14	49.94	<.0001	0.01
PC_3	0.25	11.59	44.08	<.0001	9.86
PC_{12}	0.11	5.21	19.82	<.0001	0.47
PC_{14}	0.05	2.35	8.95	0.0051	0.04
PC_{10}	0.05	2.32	8.84	0.0054	0.92
PC_9	0.03	1.22	4.66	0.0381	1.68
PC_7	0.02	0.95	3.63	0.0654	4.61

The leave–one–out cross–validated predictive pED_{50} values were listed in the Table 4 of supplementary materials and plotted in Figure 11.

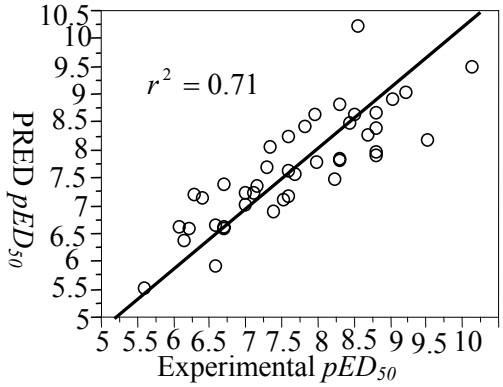


Figure 11. Leave–one–out cross–validated prediction versus experimental pED_{50} values for dataset4.

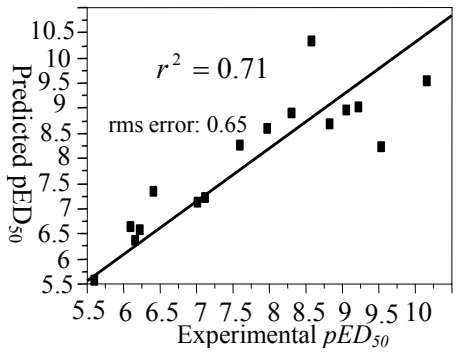


Figure 12. Predicted pED_{50} values versus experimental pED_{50} values for 16–compounds test set of dataset4.

A QSAR model with seven descriptive variables for a dataset of 42 compounds may have a tendency of over–fitting. The leave–four–out procedure was carried out to test the predictive ability and robustness of the model. The predicted pED_{50} values for the 16 test compounds are listed in bold in Table 4 of supplementary materials and plotted in Figure 12. The rms error between predicted pED_{50} and the experimental pED_{50} of the test set compounds was 0.65, and the correlation coefficient between experimental and predicted values is $r^2 = 0.71$.

4 CONCLUSIONS

We have collected over 118 bisphosphonates with different bioactivities from various literature sources and performed QSAR studies on datasets according different bioactivities. For the GGPPSase inhibitor dataset (dataset1), we built a simple and explicit QSAR model based on the enzymatic activity of 28 compounds. This model (Eq. (3), $r^2 = 0.86$, $r^2_{LOO} = 0.82$, $s = 0.45$, $s_{LOO} = 0.51$, $F = 77.56$) has comparable predictive ability with that of the CoMFA model (Eq. (1)) reported by Szabo *et al*. [13] for the same dataset. The QSAR of Dataset2 of 28 compounds with bioactivities against the growth of *T. Brucei rhodesiense* was studied using principal component analysis followed by stepwise variable selection. The PCA model (Eq. (7), $r^2 = 0.85$, $r^2_{LOO} = 0.79$, $s = 0.30$, $s_{LOO} = 0.35$, $F = 32.03$) based on the dataset also has nearly equal predictive ability with that of the CoMFA model (Eq. (2)) built by Martin *et al*. [5] We divided the 86 bisphosphonates reported by Novartis with *in vivo* activity data in TPTX rats into two sub datasets according their structural features and rough speculations of their mode of action. A six PCs model (Eq. (8), $r^2 = 0.80$, $r^2_{LOO} = 0.72$, $s = 0.44$, $s_{LOO} = 0.53$, $F = 24.83$) elucidated the dataset of 44 compounds in which containing aliphatic linked nitrogen atoms. The other dataset includes 42 BPs containing a heterocyclic moiety with at least one nitrogen atom. Its PCA model (Eq. (9), $r^2 = 0.80$, $r^2_{LOO} = 0.71$, $s = 0.46$, $s_{LOO} = 0.57$, $F = 19.99$) consists of seven PCs. A leave–four–out test procedure shows that though the QSAR models based on *in vivo* bone resorption pED_{50} values cannot provide explicit indications for drug design, their predictive ability for related compounds is quite good.

Supplementary Material

Table S1. Experimental IC_{50}, pIC_{50} and predicted pIC_{50} values for GGPPSase inhibitors (dataset1) and statistical parameters for QSAR models

Cmpd		Experimental activity		QSAR model predicted pIC_{50}			
Serial	Cmpd code	IC_{50} (μM)	pIC_{50}	Training set	3 compd test set		
111	Homorisedronate	410	3.39	3.67	**3.80**	3.76	3.77
2	Pamidronate	180	3.74	2.91	2.93	2.94	2.89
3	Alendronate	440	3.36	3.36	3.26	3.22	3.23
10	Ibandronate	83	4.08	4.83	4.72	4.70	4.70
112	Risedronate	350	3.46	3.46	3.63	3.61	**3.60**
108	NE97220	220	3.66	3.74	3.80	3.76	3.77
109	N–(2–(4–picolyl))AMDP	260	3.59	3.40	3.56	3.48	3.53
101	T.B. 006	2.2	5.66	5.50	5.41	5.44	5.38
100	T.B. 007	0.92	6.04	6.20	6.24	6.30	**6.21**
8	T.B. 009	330	3.48	3.16	3.10	3.09	3.07
99	T.B. 010	1.4	5.85	6.02	**6.00**	6.05	5.97
107	T.B. 012	220	3.66	3.33	3.26	3.22	3.22
110	T.B. 013	550	3.26	3.41	3.68	3.59	3.65
98	T.B. 014	0.72	6.14	5.62	5.55	5.61	5.52
105	T.B. 016	220	3.66	3.94	4.51	**4.66**	4.47
5	Neridronate	690	3.16	3.90	3.79	3.73	3.76
106	T.B. 020	180	3.74	3.61	4.01	4.02	3.98
95	T.B. 023	53	4.28	4.25	4.14	4.19	4.10
93	T.B. 024	620	3.21	3.58	3.41	3.46	3.37
94	T.B. 025	200	3.70	3.90	3.78	3.83	3.75
87	3–azaGGPP	0.14	6.85	6.57	6.69	6.44	6.69
91	3–azahomoFPP	0.31	6.51	5.93	5.91	**5.68**	5.90
90	3–azahomoGGPP	0.37	6.43	6.92	6.96	6.67	6.97
88	3–azaFPP	0.74	6.13	5.60	5.59	5.40	5.57
97	GGPP017	4.3	5.37	4.94	4.84	**4.90**	4.81
96	GGPP018	11	4.96	4.61	4.50	4.56	**4.47**
102	GGPP031	19	4.72	4.79	**4.74**	4.76	4.71
89	3–azaGPP	240	3.61	4.52	4.44	4.33	4.42
r^2				0.86	0.85	0.88	0.85
s				0.45	0.47	0.40	0.46
r^2_{LOO}				0.82	0.81	0.85	0.81
s_{LOO}				0.51	0.54	0.46	0.54
F				77.56	64.53	83.83	64.81
k				2	2	2	2
n				28	25	25	25

Table S2. Experimental IC50, pIC50 and predicted pIC50 values for bisphosphonates against T. Brucei Trypomastigotes (dataset2) and statistical parameters for QSAR models

Cmpd		Experimental activity		QSAR model predicted pIC_{50}			
Serial	Cmpd code	IC_{50} (μM)	pIC_{50}	Training set	3 compd test set		
111	Homorisedronate	1.7	5.77	5.89	5.89	5.93	5.89
2	Pamidronate	177	3.75	4.22	4.24	4.23	4.21
10	Ibandronate	0.96	6.01	5.88	5.87	**5.86**	5.96
112	Risedronate	8.6	5.06	5.52	5.51	**5.57**	5.51
108	NE97220	0.70	6.15	5.87	**5.85**	5.88	5.96
109	N–(2–(4–picolyl))AMDP	0.61	6.21	5.83	5.81	5.83	5.94
113	NE58018	0.22	6.66	6.52	6.51	6.53	6.59
114	N–(2–(5–chloro)–pyridyl)AMDP	53.30	4.27	5.16	5.12	5.16	**5.27**
101	T.B. 006	1.70	5.77	5.20	5.22	5.23	5.18
100	T.B. 007	2.0	5.70	5.75	5.73	5.78	5.75
7	Olpadronate	5.4	5.27	5.47	**5.50**	5.45	5.53
8	T.B. 009	7.8	5.11	4.89	4.92	4.91	4.90
99	T.B. 010	8.0	5.10	5.16	5.15	5.20	5.13
107	T.B. 012	8.6	5.07	4.64	4.62	4.67	**4.66**
110	T.B. 013	19.8	4.70	5.00	4.99	5.03	5.02
98	T.B. 014	20.5	4.69	4.74	4.69	4.76	4.76
115	T.B.015	20.9	4.68	4.59	4.62	4.64	4.50
105	T.B. 016	21.3	4.67	4.77	4.73	4.82	4.75
5	Neridronate	31.7	4.50	4.68	4.71	4.71	4.66
116	T.B. 018	34.4	4.46	4.50	4.53	4.49	**4.54**
117	T.B. 019	39.5	4.40	4.38	4.34	4.39	4.47
106	T.B. 020	40.0	4.39	4.39	4.37	4.49	4.31
103	T.B. 021	50.6	4.30	3.98	3.99	4.00	3.93
95	T.B. 023	62.4	4.20	4.02	3.97	4.02	4.07
93	T.B. 024	92.0	4.04	3.87	**3.83**	3.86	3.92
95	T.B. 025	99.8	4.00	3.98	3.95	3.99	3.99
104	T.B. 026	102.0	3.99	4.19	4.19	**4.21**	4.20
118	T.B. 2–13	27.9	4.55	4.39	4.35	4.41	4.53
r^2				0.85	0.83	0.85	0.91
s				0.30	0.30	0.29	0.24
r^2_{LOO}				0.79	0.76	0.79	0.88
s_{LOO}				0.35	0.37	0.34	0.28
F				32.03	24.88	27.43	50.13
k				4	4	4	4
n				28	25	25	25

Table S3. Experimental ED_{50}, pED_{50} and predicted pED_{50} values for dataset3 and statistical parameters for QSAR models

Cmpd		Experimental activity		QSAR model predicted pIC_{50}				
Serial	Cmpd code	ED_{50} (μg /kg)	pED_{50}	Training set	4 compd test set			
2	Pamidronate	61	7.21	7.48	7.50	**7.50**	7.45	7.45
3	Alendronate	8	8.10	8.04	8.05	8.05	8.08	8.10
10	Ibandronate	1.1	8.96	8.10	8.16	8.25	8.24	8.30
7	Olpadronate	12	7.92	8.10	8.03	8.12	8.05	8.08
8	T.B. 009	3.4	8.47	8.25	8.23	**8.26**	8.27	8.30
5	Neridronate	60	7.22	7.65	**7.69**	7.60	7.61	7.61
1	Novartis 1a	150	6.82	7.31	7.20	7.29	**7.25**	7.25
4	Novartis 1d	20	7.70	7.86	7.92	7.84	7.87	7.88
6	Novartis 1g	15	7.82	8.03	7.95	8.04	8.04	8.07
8	T.B. 009	3	8.52	8.61	8.49	8.60	8.59	8.66
9	Novartis 1j	3	8.52	8.17	**8.09**	8.23	8.23	8.29
11	Novartis 1l	100	7	6.80	6.90	6.87	6.87	**6.84**
13	Novartis 1o	18	7.74	8.57	8.30	8.36	8.29	8.31
14	Novartis 1p	65	7.19	7.08	7.21	7.11	7.08	7.06
29	Novarris 4a	300	6.52	6.58	6.63	6.62	6.62	6.59
30	Novartis 4b	1.4	8.85	9.47	9.33	9.37	9.32	9.39
31	Novartis 4c	20	7.70	8.33	8.29	8.30	8.27	8.31
32	Novartis 4d	1	9	8.87	8.83	**8.87**	8.83	8.89
33	Novartis 4e	15	7.82	7.87	7.87	7.89	**7.86**	7.87
34	Novartis 4f	1.5	8.82	8.73	8.67	8.73	**8.67**	8.72
35	Novartis 4g	0.7	9.15	8.99	8.98	8.98	8.93	8.99
36	Novartis 4i	1	9	8.45	8.46	8.56	8.62	8.68
37	Novartis 4j	0.4	9.40	8.74	**8.67**	8.81	8.79	8.87
38	Novartis 4k	20	7.70	7.09	7.10	7.24	7.17	7.18
39	Novartis 4l	1500	5.82	6.26	6.09	**6.25**	6.21	6.18
40	Novartis 5a	1.5	8.82	9.08	9.05	9.04	8.98	9.04
41	Novartis 5b	1.7	8.77	9.45	9.38	9.35	9.36	9.43
42	Novartis 5c	1.2	8.92	8.01	8.22	8.17	8.14	8.17
43	Novartis 5d	0.5	9.30	8.45	8.56	8.50	8.52	8.55
44	Novartis5e	1.7	8.77	8.01	8.10	8.07	8.04	8.05
46	Novartis 5g	1.3	8.89	8.34	8.42	8.35	8.36	8.38
45	Novartis 5f	0.6	9.22	9.72	9.52	9.63	9.61	**9.70**
47	Novartis 5h	1.2	8.92	8.55	8.56	8.63	8.64	8.67
48	Novartis 5i	20	7.70	7.75	7.84	7.77	7.76	7.76
49	Novartis 5j	10	8	7.78	7.80	7.82	7.83	**7.86**
50	Novartis 5k	500	6.30	7.03	6.93	6.94	6.98	**6.97**
51	Novartis 5l	4	8.40	7.57	7.67	7.64	7.64	7.64
52	Novartis 5m	7500	5.12	5.15	**5.17**	5.28	5.21	5.14
53	Novartis 5n	100	7	7.71	7.66	7.72	7.70	7.71
54	Novartis 5p	0.7	9.15	8.97	9.05	8.95	8.98	9.01
55	Novartis 5q	7	8.15	8.91	8.72	8.78	8.82	8.87
56	Novartis 5r	0.33	9.48	8.82	8.85	8.90	**8.84**	8.89
57	Novartis 5s	7.8	8.11	8.95	8.76	8.65	8.66	8.66
72	Novartis 7e	7	8.15	9.13	8.61	8.59	8.60	8.66
r^2				0.80	0.75	0.77	0.79	0.79
s				0.44	0.44	0.46	0.45	0.44
r^2_{LOO}				0.72	0.65	0.67	0.70	0.68
s_{LOO}				0.53	0.55	0.55	0.54	0.55
F				24.83	16.61	18.42	21.24	20.47
k				6	6	6	6	6
n				44	40	40	40	40

Table S4. Experimental ED_{50}, pED_{50} and predicted pED_{50} values for dataset4 and statistical parameters for QSAR models

Cmpd		Experimental activity		QSAR model predicted pIC_{50}				
Serial num	Cmpd code	ED_{50} (μg /kg)	pED_{50}	Training set	4 compd test set			
65	Zoledronate	0.07	10.15	9.50	**9.53**	9.55	9.75	**9.66**
15	Novartis 2a	10	8	7.79	7.82	7.88	7.88	7.65
16	Novartis 2b	25	7.60	7.63	7.66	7.65	7.68	7.37
17	Novartis 2c	250	6.60	6.65	6.73	6.63	6.66	6.63
18	Novartis 2d	70	7.15	7.37	7.35	7.30	7.36	7.28
19	Novartis 2e	3.5	8.46	8.49	8.46	8.48	8.55	8.50
20	Novartis 2f	5.6	8.25	7.50	7.58	7.62	7.64	7.47
21	Novartis 2g	11	7.96	8.64	**8.56**	8.54	8.68	8.58
22	Novartis 2h	100	7	7.02	7.05	6.98	7.09	**7.09**
23	Novartis 2j	30	7.52	7.13	7.20	7.18	7.21	7.38
24	Novartis 2k	25	7.60	8.24	8.14	8.16	**8.25**	8.09
25	Novartis 2m	400	6.40	7.13	6.82	6.84	6.67	**7.31**
26	Novartis 3a	50	7.30	7.70	7.59	7.66	7.70	7.41
27	Novartis 3b	250	6.60	5.94	6.11	6.12	6.12	5.92
28	Novartis 3c	2500	5.60	5.52	5.62	5.67	5.65	**5.54**
58	Novartis 6a	5	8.30	8.83	8.75	8.70	**8.87**	8.90
59	Novartis 6b	0.6	9.22	9.04	**8.99**	9.05	9.15	9.32
60	Novartis 6c	25	7.60	7.17	7.31	7.31	7.40	7.44
61	Novartis 6d	0.3	9.52	8.21	8.35	**8.21**	8.38	8.52
62	Novartis 6e	20	7.70	7.59	7.65	7.48	7.65	7.82
63	Novartis 6f	15	7.82	8.45	8.36	8.25	8.32	8.50
64	Novartis 6h	1.5	8.82	8.41	8.47	8.39	8.47	8.65
66	Novartis 6j	45	7.35	8.08	7.92	7.86	8.03	7.77
67	Novartis 6k	3	8.52	8.64	8.52	8.63	8.70	8.79
68	Novartis 6l	1.5	8.82	7.92	8.02	7.98	8.03	8.17
69	Novartis 6n	600	6.22	6.60	6.60	**6.56**	6.59	6.52
70	Novartis 7c	800	6.10	6.65	6.57	6.53	**6.60**	6.42
71	Novartis 7d	40	7.40	6.91	7.04	7.01	7.11	7.17
73	Novartis 8a	5	8.30	7.85	7.83	7.91	7.97	7.87
74	Novartis8b	100	7	7.24	7.20	7.24	7.23	7.21
75	Novartis 8c	1.5	8.82	7.98	8.20	8.07	8.18	8.54
76	Novartis 8d	1.5	8.82	8.69	8.66	**8.65**	8.73	8.90
77	Novartis 8e	2	8.70	8.29	8.31	8.23	8.32	8.56
78	Novartis 8f	0.9	9.05	8.91	8.95	8.80	**8.94**	9.07
79	Novartis 8g	200	6.70	6.63	6.64	6.66	6.65	6.75
80	Novartis 8h	2.7	8.57	10.22	9.66	9.53	9.74	10.32
81	Novartis 8j	500	6.30	7.21	7.15	7.13	7.15	7.12
82	Novartis 8k	5	8.30	7.83	7.88	8.00	7.94	8.06
83	Novartis 8l	75	7.12	7.25	7.21	**7.19**	7.19	7.20
84	Novartis 8m	200	6.70	7.40	7.24	7.27	7.24	7.34
85	Novartis 9a	200	6.70	6.62	6.69	6.60	6.71	6.41
86	Novartis 9b	700	6.15	6.38	**6.34**	6.23	6.30	6.29
r^2				0.80	0.76	0.81	0.80	0.83
s				0.46	0.46	0.44	0.46	0.45
r^2_{LOO}				0.71	0.65	0.69	0.69	0.75
s_{LOO}				0.55	0.60	0.56	0.58	0.51
F				19.99	13.84	18.47	17.44	20.59
k				7	7	7	7	7
n				42	38	38	38	38

Table S5. Original descriptors adopted for PCA procedure in each dataset

Datasets	Original descriptors
Dataset2	a_nH, zagreb, PEOE_VSA+0, PEOE_VSA+1, PEOE_VSA+2, PEOE_VSA–1, Q_VSA_POS, Q_VSA_HYD, Q_VSA_PPOS, E_sol, E_stb, E_strain, E_tor, E_vdw, KierFlex, apol, vsa_don, vsa_other, SlogP_VSA1, SlogP_VSA4, SlogP_VSA7, SlogP_VSA8, SlogP_VSA9, SMR_VSA1, SMR_VSA2, SMR_VSA3, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7, vol, VSA
Dataset3	VSA, DASA, vol, SlogP_VSA8, E_tor, E_ang, Q_VSA_POS, Zagreb, ASA+, SMR_VSA6, Q_VSA_PNEG, Q_VSA_HYD, PEOE_VSA_HYD, PEOE_VSA–1, weinerPath
Dataset4	weinerPol, PEOE_VSA+3, E_ang, SlogP_VSA5, DASA, DCASA, E_vdw, apol, SlogP_VSA6, SMR_VSA2, ASA_H, PEOE_VSA+4, PEOE_VSA–3, Q_VSA_HYD, SMR_VSA5

Appendix 1

Denotations of original descriptors adopted for PCA procedure in each dataset

I. Physical Properties that can be calculated from the connection table (with no dependence on conformation) of a molecule:

Code	Description
apol	Sum of the atomic polarizabilities (including implicit hydrogens) with polarizabilities taken from ref. [24].

II.Subdivided Surface Areas

The Subdivided Surface Areas are descriptors based on an approximate accessible van der Waals surface area calculation for each atom, v_i along with some other atomic property, p_i. The v_i are calculated using a connection table approximation. Each descriptor in a series is defined to be the sum of the v_i over all atoms i such that p_i is in a specified range (a,b).

In the descriptions to follow, L_i denotes the contribution to logP(o/w) for atom i as calculated in the SlogP descriptor [17]. R_i denotes the contribution to Molar Refractivity for atom i as calculated in the SMR descriptor [17]. The ranges were determined by percentile subdivision over a large collection of compounds.

Code	Description
SlogP_VSA0	Sum of v_i such that $L_i <= -0.4$.
SlogP_VSA1	Sum of v_i such that L_i is in (–0.4,–0.2].
SlogP_VSA2	Sum of v_i such that L_i is in (–0.2,0].
SlogP_VSA3	Sum of v_i such that L_i is in (0,0.1].
SlogP_VSA4	Sum of v_i such that L_i is in (0.1,0.15].
SlogP_VSA5	Sum of v_i such that L_i is in (0.15,0.20].
SlogP_VSA6	Sum of v_i such that L_i is in (0.20,0.25].
SlogP_VSA7	Sum of v_i such that L_i is in (0.25,0.30].
SlogP_VSA8	Sum of v_i such that L_i is in (0.30,0.40].
SlogP_VSA9	Sum of v_i such that $L_i > 0.40$.
SMR_VSA0	Sum of v_i such that R_i is in [0,0.11].
SMR_VSA1	Sum of v_i such that R_i is in (0.11,0.26].
SMR_VSA2	Sum of v_i such that R_i is in (0.26,0.35].
SMR_VSA3	Sum of v_i such that R_i is in (0.35,0.39].
SMR_VSA4	Sum of v_i such that R_i is in (0.39,0.44].
SMR_VSA5	Sum of v_i such that R_i is in (0.44,0.485].
SMR_VSA6	Sum of v_i such that R_i is in (0.485,0.56].
SMR_VSA7	Sum of v_i such that $R_i > 0.56$.

II. Atom Counts and Bond Counts and Kier&Hall Connectivity and Kappa Shape Indices

Code	Description
a_nH	Number of hydrogen atoms (including implicit hydrogens). This is calculated as the sum of h_i over all non–trivial atoms i plus the number of non–trivial hydrogen atoms.
zagreb	Zagreb index: the sum of d_i^2 over all heavy atoms i.
KierFlex	Kier molecular flexibility index: (KierA1) (KierA2) /n [18].

III. Adjacency and Distance Matrix Descriptors

Code	Description
weinerPath	Wiener path number: half the sum of all the distance matrix entries as defined in ref. [19] and [25].
weinerPol	Wiener polarity number: half the sum of all the distance matrix entries with a value of 3 as defined in ref. [19].

IV. Pharmacophore Feature Descriptors

Code	Description
vsa_don	Approximation to the sum of VDW surface areas of pure hydrogen bond donors (not counting basic atoms and atoms that are both hydrogen bond donors and acceptors such as –OH).
vsa_other	Approximation to the sum of VDW surface areas of atoms typed as "other".

V. Partial Charge Descriptors (Let q_i denote the partial charge of atom i as defined above. Let v_i be the van der Waals surface area of atom i.)

Code	Description
Q_PC+ PEOE_PC+	Total positive partial charge: the sum of the positive q_i. Q_PC+ is identical to PC+ which has been retained for compatibility.
Q_PC– PEOE_PC–	Total negative partial charge: the sum of the negative q_i. Q_PC– is identical to PC– which has been retained for compatibility.
Q_RPC+ PEOE_RPC+	Relative positive partial charge: the largest positive q_i divided by the sum of the positive q_i. Q_RPC+ is identical to RPC+ which has been retained for compatibility.
Q_PRC– PEOE_RPC–	Relative negative partial charge: the smallest negative q_i divided by the sum of the negative q_i. Q_RPC– is identical to RPC– which has been retained for compatibility.
Q_VSA_POS PEOE_VSA_POS	Total positive van der Waals surface area. This is the sum of the v_i such that q_i is non–negative. The v_i are calculated using a connection table approximation.
Q_VSA_NEG PEOE_VSA_NEG	Total negative van der Waals surface area. This is the sum of the v_i such that q_i is negative. The v_i are calculated using a connection table approximation.
Q_VSA_PPOS PEOE_VSA_PPOS	Total positive polar van der Waals surface area. This is the sum of the v_i such that q_i is greater than 0.2. The v_i are calculated using a connection table approximation.
Q_VSA_PNEG PEOE_VSA_PNEG	Total negative polar van der Waals surface area. This is the sum of the v_i such that q_i is less than –0.2. The v_i are calculated using a connection table approximation.
Q_VSA_HYD PEOE_VSA_HYD	Total hydrophobic van der Waals surface area. This is the sum of the v_i such that $\|q_i\|$ is less than or equal to 0.2. The v_i are calculated using a connection table approximation.
Q_VSA_POL PEOE_VSA_POL	Total polar van der Waals surface area. This is the sum of the v_i such that $\|q_i\|$ is greater than 0.2. The v_i are calculated using a connection table approximation.
Q_VSA_FPOS PEOE_VSA_FPOS	Fractional positive van der Waals surface area. This is the sum of the v_i such that q_i is non–negative divided by the total surface area. The v_i are calculated using a connection table approximation.
Q_VSA_FNEG PEOE_VSA_FNEG	Fractional negative van der Waals surface area. This is the sum of the v_i such that q_i is negative divided by the total surface area. The v_i are calculated using a connection table approximation.
Q_VSA_FPPOS PEOE_VSA_FPPOS	Fractional positive polar van der Waals surface area. This is the sum of the v_i such that q_i is greater than 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.
Q_VSA_FPNEG PEOE_VSA_FPNEG	Fractional negative polar van der Waals surface area. This is the sum of the v_i such that q_i is less than –0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.
Q_VSA_FHYD PEOE_VSA_FHYD	Fractional hydrophobic van der Waals surface area. This is the sum of the v_i such that $\|q_i\|$ is less than or equal to 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.
Q_VSA_FPOL PEOE_VSA_FPOL	Fractional polar van der Waals surface area. This is the sum of the v_i such that $\|q_i\|$ is greater than 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.
PEOE_VSA+6	Sum of v_i where q_i is greater than 0.3.
PEOE_VSA+5	Sum of v_i where q_i is in the range [0.25,0.30).
PEOE_VSA+4	Sum of v_i where q_i is in the range [0.20,0.25).
PEOE_VSA+3	Sum of v_i where q_i is in the range [0.15,0.20).

V. (Continued)

Code	Description
PEOE_VSA+2	Sum of v_i where q_i is in the range [0.10,0.15).
PEOE_VSA+1	Sum of v_i where q_i is in the range [0.05,0.10).
PEOE_VSA+0	Sum of v_i where q_i is in the range [0.00,0.05).
PEOE_VSA–0	Sum of v_i where q_i is in the range [–0.05,0.00).
PEOE_VSA–1	Sum of v_i where q_i is in the range [–0.10,–0.05).
PEOE_VSA–2	Sum of v_i where q_i is in the range [–0.15,–0.10).
PEOE_VSA–3	Sum of v_i where q_i is in the range [–0.20,–0.15).
PEOE_VSA–4	Sum of v_i where q_i is in the range [–0.25,–0.20).
PEOE_VSA–5	Sum of v_i where q_i is in the range [–0.30,–0.25).
PEOE_VSA–6	Sum of v_i where q_i is less than –0.30.

VI. Potential Energy Descriptors

Code	Description
E_ang	Angle bend potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.
E_sol	Solvation energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.
E_stb	Bond stretch–bend cross–term potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.
E_strain	Local strain energy: the current energy minus the value of the energy at a near local minimum. The current energy is calculated as for the E descriptor. The local minimum energy is the value of the E descriptor after first performing an energy minimization. Current chirality is preserved and charges are left undisturbed during minimization. The structure in the database is not modified (results of the minimization are discarded).
E_tor	Torsion (proper and improper) potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.
E_vdw	van der Waals component of the potential energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.

VII. Surface Area, Volume and Shape Descriptors

Code	Description
ASA	Water accessible surface area calculated using a radius of 1.4 A for the water molecule. A polyhedral representation is used for each atom in calculating the surface area.
vol	van der Waals volume calculated using a grid approximation (spacing 0.75 A).
VSA	van der Waals surface area. A polyhedral representation is used for each atom in calculating the surface area.

VII. Conformation Dependent Charge Descriptors

Code	Description
ASA+	Water accessible surface area of all atoms with positive partial charge (strictly greater than 0).
ASA_H	Water accessible surface area of all hydrophobic ($\|q_i\|<0.2$) atoms.
DASA	Absolute value of the difference between ASA+ and ASA^-.
DCASA	Absolute value of the difference between CASA+ and $CASA^-$ [26].

5 REFERENCES

[1] R. G. G. Russell, M. J. Rogers, Bisphosphonates: From the Laboratory to the Clinic and Back Again, *Bone* **1999**, *25*, 97–106.

[2] T. H. Cromartie, K. J. Fisher, and J. N. Grossman, The Discovery of a Novel Site of Action for Herbicidal Bisphosphonates, *Pesticide Biochemistry and Physiology* **1999**, *63*, 114–126.

[3] S. Oura, T. Sakurai, G. Yoshimura, T. Tamaki, and T. Umemura, Study on the Safty of Rapid Infusion and the Efficacy of Incadronate Against Bone Metastase of breast Cancer, *Gan to Kagaku Ryoho* **1999**, *26*, 1623–1628.

[4] M. B. Martin, J. S. Grimley, J. C. Lewis, H. T. Heath, and B. N. Bailey, 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.

[5] M. B. Martin, J. M. Sanders, H. Kendrick, K. d. Luca–Fradley, J. C. Lewis, J. S. Grimley, E. M. Van Brussel, J. R. Olsen, G. A. Meints, A. Burzynska, P. Kafarski, S. L. Croft, and E. Oldfield, Activity of Bisphosphonates against Trypanosoma brucei rhodesiense, *J. Med. Chem.* **2002**, *45*, 2904–2914.

[6] D. Amin, S. A. Cornell, S. K. Gustafson, S. J. Needle, J. W. Ullrich, G. E. Bilder, and M. H. Perrone, Bisphosphonates used for the Treatment of Bone Disorders Inhibit Squalene Synthase and Cholesterol Biosynthesis, *J. Lipid. Res.* **1992**, *33*, 1657–1663.

[7] D. Amin, S. A. Cornell, M. H. Perrone, and G. E. Bilder, 1–hydroxy–3–(methylpentylamino)–propylidene–1,1–Bisphosphonic Acid as a Potent Inhibitor of Squalene Synthase, *Drug Res.* **1996**, *46*, 759–762.

[8] T. H. Cromartie, and K. J. Fisher, Method of Controlling Plants by Inhibition of Farnesyl Pyrophosphate Synthase, Zeneca Limited, London, England **1998**, 5, 756, 423 (US Patent).

[9] J. E. Grove, R. J. Brown, and D. J. Watts, 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.

[10] M. Sato, W. Grasser, N. Endo, R. Akins, H. Simmons, D. D. Thompson, E. Golub, and G. A. Rodan, Bisphosphonate action: Alendronate Localization in Rat Bone and Effects on Osteoclast Ultrastructure, *J. Clin. Invest.* **1991**, *88*, 2095–2105.

[11] J. D. Bergstrom, R. G. Bostedor, P. J. Masarachia, A. A. Reszka, and G. Rodan, Alendronate Is a Specific, Nanomolar Inhibitor of Farnesyl Diphosphate Synthase, *Biochemistry and Biophysics* **2000**, *373*, 231–241.

[12] C. M. Szabo, M. B. Martin, and E. Oldfield, An Investigation of Bone Resorption and Dictyostelium discoideum Growth Inhibition by Bisphosphonate Drugs, *J. Med. Chem.* **2002**, *45*, 2894–2903.

[13] C. M. Szabo, Y. Matsumura, S. Fukura, M. B. Martin, J. M. Sanders, S. Sengupta, J. A. Cieslak, T. C. Loftus, C. R. Lea, H. J. Lee, A. Koohang, R. M. Coates, H. Sagami, and E. Oldfield, Inhibition of Geranylgeranyl Diphosphate Synthase by Bisphosphonates and Diphosphates: A Potential Route to New Bone Antiresorption and Antiparasitic Agents, *J. Med. Chem.* **2002**, *45*, 2185–2196.

[14] L. Widler, K. A. Jaeggi, M. Glatt, K. Muller, R. Bachmann, M. Bisping, A. R. Born, R. Cortesi, G. Guiglia, H. Jeker, R. Klein, U. Ramseier, J. Schmid, G. Schreiber, Y. Seltenmeyer, and J. R. Green, Highly Potent Geminal Bisphos phonates. From Pamidronate Disodium (Aredia) to Zoledronic Acid (Zometa), *J. Med. Chem.* **2002**, *45*, 3721–3738.

[15] MOE 2001.01, Chemical Computing Group Inc., Montreal, Canada, E–mail info@chemcomp.com, www http://www.chemcomp.com.

[16] JMP 4.5, SAS Institute Inc., Cary, NC, USA, www http://www.jmp.com.

[17] S. A. Wildman, G. M. Crippen, Prediction of Physicochemical Parameters by Atomic Contributions, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868–873.

[18] L H. Hall and L. B. Kier, The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure–Property Modeling; in: *Reviews in Computational Chemistry*, Eds, K. B. Lipkowitz and D. B. Boyd, VCH Publishers, New York, **1991**, Vol. *2*, pp. 367–422.

[19] A. T. Balaban, Five New Topological Indices for the Branching of Tree–like Graphs, *Theor. Chim. Acta.* **1979**, *53*, 355–375.

[20] A. T. Balaban, Highly Discriminating Distance–based Topological Index, *Chem. Phys. Lett.* **1982**, *89*, 399–404.

[21] M. Petitjean, Applications of the Radius–Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds, *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 331–337.

[22] J. Gasteiger, M. Marsili, Iterative Partial Equalization of Obital Electronegativity – A Rapid Access to Atomic Charges, *Terahedron* **1980**, *36*, 3219–3228.

[23] F. E. Grubbs, Procedures for Detecting Outlying Observations in Samples, *Technometrics* **1969**, *11*, 1–21.

[24] D. R. Lide, CRC Handbook of Chemistry and Physics, CRC Press, 2000 NW Corporate Blvd Boca Raton, FL 33431, USA, 1994

[25] H. Wiener, Structural Determination of Paraffin Boiling Points, *J. Am. Chem. Soc.* **1947**, *69*, 17–20.

[26] D. T. Stanton and P. C. Jurs, Development and Use of Charged Partial Surface Area Structural Descriptors in Computer Assissted Quantitative Structure Property Relationship Studies, *Anal. Chem.* **1990**, *62*, 2323–2329.