

Structure–Activity Relationships for a Class of Inhibitors of Purine Nucleoside Phosphorylase

Victor Farutin,[†] Lori Masterson, Adriano D. Andricopulo, Jianming Cheng, Brad Riley, Ryan Hakimi, Jack W. Frazer, and Eugene H. Cordes*

College of Pharmacy and Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109

Received January 22, 1999

Values of inhibition constants, K_i , for a family of structurally related, competitive inhibitors of calf spleen purine nucleoside phosphorylase (PNP) have been determined employing both inosine as substrate and a manual assay and 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG) as substrate and a robot-based enzyme kinetics facility. Several of the values determined robotically were confirmed employing the same substrate and a manual assay. Surprisingly, for many of the inhibitors examined, values of K_i determined with MESG as substrate are smaller than those obtained employing inosine as substrate by a factor that varies from less than 2 to 10. Values of concentrations required for 50% inhibition of PNP, IC_{50} , have also been determined for the same family of inhibitors employing inosine as substrate. Values of IC_{50}^{ino} and those for K_i^{ino} and K_i^{mesg} for subsets of the inhibitors have been employed as training sets to create quantitative structure–activity relationships (QSAR) which have substantial power to predict values of IC_{50} and K_i for inhibitors outside the training set. These QSAR models should be useful in guiding future medicinal chemistry efforts designed to discover inhibitors of PNP having increased potency.

Introduction

The ability to efficiently design or discover novel, patentable molecules which are potent, specific inhibitors of enzymes or potent, specific agonists or antagonists at biological receptors is of great importance. Many such molecules contribute to the prevention of or therapy for human diseases. The pharmaceutical industry worldwide continues to search for and employ novel technologies to improve their ability to rapidly identify and characterize molecules in these classes.

Quantitative structure–activity relationships (QSAR) [or quantitative structure–property relationships (QSPR)] have been employed, and continue to be developed and employed, both to correlate information in data sets and as a tool to facilitate, for example, the discovery of enzyme inhibitors. Extensive reviews of the enormous QSAR literature are available.^{1,2} Work employing classical QSAR technology has proved useful in any number of settings. At the same time, we believe that QSAR technology has a substantial unexploited potential to facilitate molecular design. Three specific goals include: (i) to improve the ability of QSAR statistical models to *predict* property values for molecules outside of the training set; (ii) to *extrapolate* to property values outside those included by members of the training set; and (iii) to *qualify* these models in a way that provides reliable measures of the accuracy of such predictions.

In an effort to build on classical QSAR technology in a way that would meet these objectives, a team of scientists was organized, initially in the pharmaceutical industry at Sterling Winthrop Pharmaceuticals Inc., with the objective of realizing much of the unexploited potential of QSAR in the service of drug discovery. This

Laboratory functions as a component of that team, which is led by J. W. Frazer. Our specific objective is to create high-quality standard databases which can be employed (i) to test and refine the developing QSAR technology and (ii) to provide the basis for discovery of novel molecules having promise of utility in clinical medicine.

Work reported herein is the result of an effort to build upon the foundation for discovery of structurally novel, potent inhibitors of human purine nucleoside phosphorylase (PNP). Such inhibitors have multiple, plausible utilities in clinical medicine.^{3–5} Scientists at BioCryst Pharmaceuticals have employed structure-based drug discovery technologies to create a strong scientific basis for drug discovery and have used this basis to discover several classes of promising inhibitors of PNP,^{6–9} one of which is currently in advanced clinical trials. We have determined values of K_i and IC_{50} for one of these classes of inhibitors, employing calf spleen PNP, and used the data to create QSAR statistical models which show substantial predictive promise. These models may be employed to predict potency of candidate inhibitors not yet synthesized, to search for synergies between QSAR technology and structure-based drug design technologies, and to use as a basis for creation of more robust predictive QSAR models through incorporation of additional data into the training set.

Results

Determination of Inhibition Constants. Structures for inhibitors employed in this study and corresponding values for K_i employing 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG) as substrate, K_i^{mesg} , K_i employing inosine as substrate, K_i^{ino} , and IC_{50} employing inosine as substrate, IC_{50}^{ino} , are collected in Table 1. Values of IC_{50} for a number of these inhibitors

[†] Present address: Millennium Pharmaceuticals, Inc., 238 Main St, Cambridge, MA 02142.

Table 1. Values of Inhibition Constants for a Family of Inhibitors of Calf Spleen PNP^a

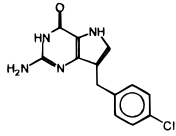
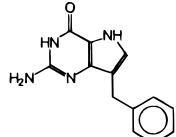
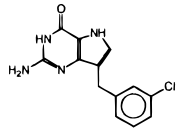
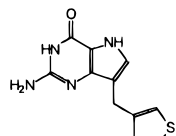
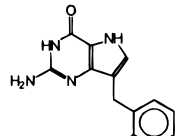
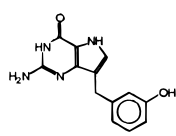
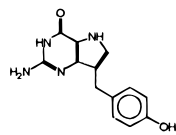
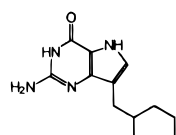
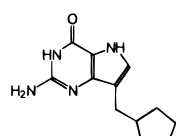
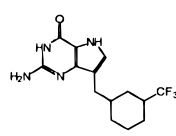
Compound	Structure	K_i^{mesg} (nM)	K_i^{ino} (nM)	K_i^{ino}/K_i^{mesg}	$IC_{50}^{ino\ b}$ (nM)	IC_{50}^{ino}/K_i^{ino}
001		3.9 ± 0.85 n=4	10.6 ± 0.40 n=2	2.72 ± 0.22	17.4 ± 0.80 n=2	1.64 ± 0.06
002		3.6 ± 0.50 n=4	11 ± 1.5^c n=2	3.06 ± 0.20	16.5 ± 1.95 n=2	1.50 ± 0.18
003		4.3 ± 0.69 n=4	18.9 ± 7.1 n=4	4.45 ± 0.41	25.8 ± 3.8 n=4	1.36 ± 0.40
004		3.5 ± 0.69 n=4	15.6 ± 3.4 n=2	4.47 ± 0.29	22.6 ± 1.30 n=3	1.45 ± 0.22
005		14.4 ± 0.90 n=7	65 ± 11.9 n=3	4.50 ± 0.19	102.6 ± 7.0 n=3	1.58 ± 0.20
006		1.7 ± 0.39 n=3	16.5 ± 1.7 n=3	9.92 ± 0.25	36.1 ± 2.9 n=4	2.18 ± 0.13
007		35 ± 2.8 n=3	109.1 ± 10.1 n=2	3.12 ± 0.12	238.9 ± 23.6 n=3	2.19 ± 0.14
008		6.4 ± 0.25 n=2	13.2 ± 4.8 n=3	2.07 ± 0.36	21 ± 0.12 n=2	1.60 ± 0.36
009		6.2 ± 1.30 n=2	9.1 ± 2.1 n=3	1.97 ± 0.31	20.2 ± 1.9 n=3	2.20 ± 0.25
010		3.9 ± 0.10 n=2	9.4 ± 4.1 n=3	2.41 ± 0.43	14.8 ± 3.0 n=4	1.38 ± 0.48

Table 1 (Continued)

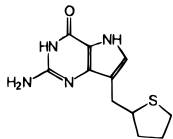
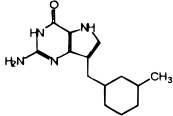
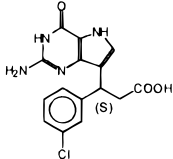
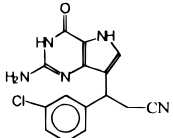
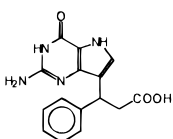
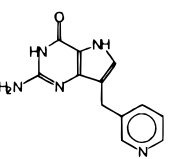
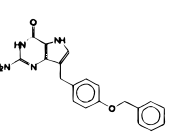
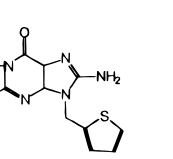
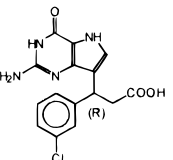
Compound	Structure	K_i^{mesg} (nM)	K_i^{ino} (nM)	$K_i^{\text{ino}}/K_i^{\text{mesg}}$	$IC_{50}^{\text{ino b}}$ (nM)	$IC_{50}^{\text{ino}}/K_i^{\text{ino}}$
011		6 ± 0.55 n=2	8.5 ± 1.6 n=3	1.43 ± 0.21	17.4 ± 0.10 n=2	2.05 ± 0.19
012		3.3 ± 0.41 n=3	11.4 ± 2.4 n=3	3.44 ± 0.24	19.2 ± 3.5 n=3	1.69 ± 0.28
013		22 ± 6.2 n=7	4.0 n=1	0.18 ± 0.27	35.6 n=1	8.92
014		4.6 ± 0.69 n=2	10.9 ± 2.2 n=3	2.39 ± 0.25	16.3 ± 0.8 n=2	1.50 ± 0.21
015		61 ± 1 n=2	9.0 n=1	0.15 ± 0.02	46.8 ± 4.7 n=2	5.18 ± 0.10
016		3.1 ± 0.11 n=2	13.5 ± 0.1 n=2	4.35 ± 0.04	27.0 ± 6.0 n=5	1.63 ± 0.27
017		20.3 ± 7.9 n=4	37.8 ± 6.6 n=3	6.11 ± 0.41	123.7 ± 4.5 n=4	2.14 ± 0.12
018		20.7 ± 2.4 n=3	35.7 ± 1.4 n=2	1.73 ± 0.12	69.4 ± 5.7 n=3	1.95 ± 0.09
019		67.7 ± 7.0 n=7	148.8 ± 2.3 n=3	2.20 ± 0.10	245.1 ± 11.9 n=6	1.65 ± 0.05

Table 1 (Continued)

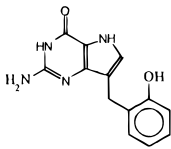
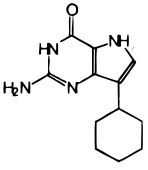
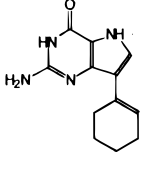
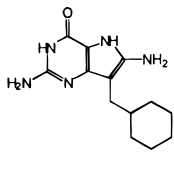
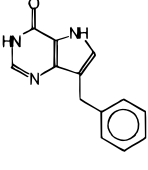
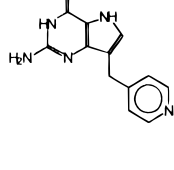
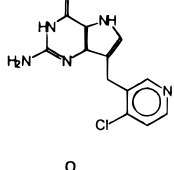
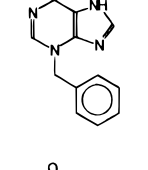
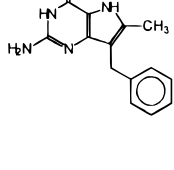
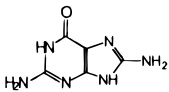
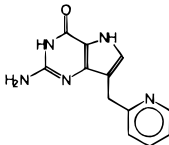
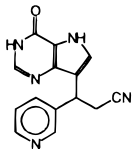
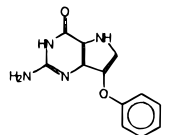
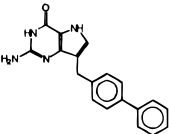
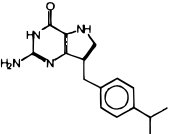
Compound	Structure	K_i^{mesg} (nM)	K_i^{ino} (nM)	$K_i^{\text{ino}}/K_i^{\text{mesg}}$	$IC_{50}^{\text{ino } b}$ (nM)	$IC_{50}^{\text{ino}}/K_i^{\text{ino}}$
020		31.8 ± 0.4 n=4	84 ± 19.3 n=4	2.65 ± 0.23	182.5 ± 15.9 n=5	2.17 ± 0.25
021		332.7 ± 66.8 n=6	335.3 ± 80.1 n=3	1.01 ± 0.31	640 ± 26 n=6	1.91 ± 0.24
022		500.3 ± 75.6 n=4	862 ± 151.0 n=4	1.72 ± 0.23	1969 ± 83 n=3	2.28 ± 0.8
023		766 ± 218 n=9	885 ± 110 n=3	1.16 ± 0.31	1471 ± 68 n=3	1.66 ± 0.13
024		5.4 ± 0.9 n=5	18.7 ± 3.5 n=2	3.43 ± 0.25	24.5 ± 1.6 n=3	1.31 ± 0.20
025		16.9 ± 1.85 n=4	25.4 ± 1.2 n=3	1.51 ± 0.12	63.5 ± 4.2 n=4	2.50 ± 0.08
026		35.6 ± 5.1 n=5	66.7 ± 30.2 n=5	1.87 ± 0.47	134.9 ± 51 n=10	2.02 ± 0.99
027		—	24645 ± 1365 n=2	—	49646 ± 9962 n=5	2.01 ± 0.21
028		67501 ± 5100 n=2	10321 ± 2249 n=3	0.15 ± 0.23	20201 ± 2584 n=5	1.96 ± 0.25

Table 1 (Continued)

Compound	Structure	K_i^{mesg} (nM)	K_i^{ino} (nM)	$K_i^{\text{ino}}/K_i^{\text{mesg}}$	$IC_{50}^{\text{ino } b}$ (nM)	$IC_{50}^{\text{ino}}/K_i^{\text{ino}}$
029		680.6 ± 141 n=5	982 ± 32 n=2	1.44 ± 0.21	2299 ± 221 n=2	2.34 ± 0.10
030		—	—	—	15.3 ± 0.2 n=2	—
031		—	—	—	38.9 ± 3.2 n=2	—
032		—	—	—	41.5 ± 2.3 n=2	—
033		—	—	—	546.4 ± 320 n=2	—
034		—	—	—	96.6 ± 39.9 n=2	—

^a All values of K_i and IC_{50} were determined at pH 7.4, 1 mM phosphate, and 25 °C. Values of IC_{50}^{ino} were determined at 10 μM inosine. ^b Values of IC_{50}^{ino} for inhibitors **001**–**024** measured at 10 μM inosine and 1 mM phosphate are generally in good agreement with those reported by the BioCryst group^{6–8} measured under comparable conditions with the exception of that for inhibitor **021** for which a value of 1300 nM has been reported.⁷ See text for a discussion of the values for inhibitors **013** and **015**. Comparative values are not available for inhibitors **025**–**034**. ^c Lit.⁶ 11 μM .

employing inosine as substrate and calf spleen PNP have previously been reported.^{6–8} Where comparisons are possible, our results and those previously reported are in good agreement. Ratios of values of K_i^{ino} to K_i^{mesg} and IC_{50}^{ino} to K_i^{ino} are also collected in Table 1.

Table 1 includes data for 34 inhibitors of calf spleen PNP. For 28 of these, we have determined values of K_i^{mesg} ; these vary from 1.7 to 67 500 nM, a factor of about 40 000-fold. The corresponding values for K_i^{ino} , for 29 inhibitors, vary from 4.0 to about 24 000 nM, a factor of 6 000-fold. Values of IC_{50}^{ino} for the complete set of 34 inhibitors vary from 14.8 to 50 000 nM, a factor of 3 400.

The family of inhibitors examined has modest structural diversity. Twenty-seven of the 34 inhibitors are 9-substituted-9-deazaguanines; one is 9-substituted-8-aminoguanine (compound **018**); one is 8-aminoguanine (compound **029**); one is a 9-substituted-8-methyl-9-

deazaguanine (compound **028**); one is a 9-substituted-8-amino-9-deazaguanine (compound **023**); two are 9-substituted-9-deazahypoxanthines (compounds **024** and **031**); the last is an *N*-3-substituted-hypoxanthine (compound **027**). Thus, the bulk of the structural diversity lies in the nature of the substituent linked to the 9-position of the purine ring.

Of the 34 inhibitors studied, three (**013**, **015**, and **019**) have 9-substituents which are predominantly anionic at the value of pH, 7.4, for which we have determined inhibition constants. Two of these, compounds **013** and **019**, are enantiomers. The *S*-enantiomer is the more potent of the two, although the difference is not profound. The *S*-enantiomer is unusual in that, like the racemic, anionic compound **015**, its potency is greater when employing inosine, rather than MESG, as substrate (Table 1).

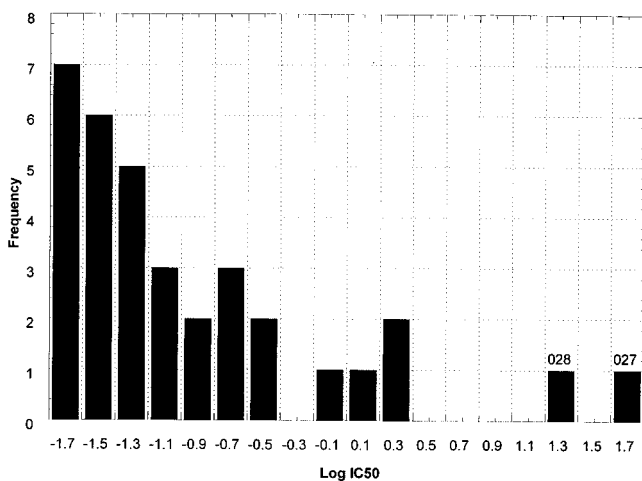


Figure 1. Distribution of values of $\log IC_{50}^{ino}$ for a set of inhibitors of calf spleen PNP.

A number of other inhibitors bear 9-substituents that have the potential to bind to PNP in an ionic form: these include the phenolic and pyridine substituents. However, all such inhibitors exist predominantly as the neutral form at pH 7.4 and are likely to bind to PNP in that form.

Surprisingly, values of K_i determined employing MESG or inosine as substrate are not identical within experimental error. For 25 of the 28 compounds for which we can make comparisons, inhibitors are equally potent or more potent, by a factor which is usually in the range 1–4 but is nearly 10 in one case, when MESG is employed as substrate. For the other three inhibitors, which include two of the three anionic substrates (compounds **013** and **015**) and the single 9-substituted-8-methyl-9-deazaguanine inhibitor (compound **028**), potency is greater by a factor of 5–7 when inosine is employed as substrate. As noted in Table 1, most values of K_i for both substrates have been determined repeatedly and the observed differences are substantially greater than experimental error.

For 27 of the 29 inhibitors for which comparisons are possible, the ratio of IC_{50}^{ino} to K_i^{ino} varies over a narrow range, from 1.3 to 2.5, reassuring consistency. The two outliers are two of the three anionic compounds, **013** and **015**, for which this ratio is 5 or 9 (Table 1).

Statistical Modeling. For two reasons, this PNP inhibitor data set is not particularly well-suited for QSAR modeling. First, as noted above, the structural diversity is modest with the bulk of the compounds being 9-substituted-9-deazaguanines. Thus little structure space is covered by 27 of the 34 compounds in the data set. At the same time, there is a small number of compounds, most notably **027** and **029** and the three anionic compounds, **013**, **015**, and **019**, which are structurally distinct. Second, 28 out of the 34 values of IC_{50}^{ino} lie within a rather narrow range: $-1.83 \geq \log IC_{50}^{ino} (\mu M) \geq -0.61$; four of the remaining compounds have $\log IC_{50}^{ino}$ values in the range -0.19 to 0.36 , and the final two values are 1.3 and 1.75. The distribution of values of $\log IC_{50}^{ino}$ is shown graphically in Figure 1. A more nearly suitable data set for statistical modeling would include more structural diversity and a larger number of values of $\log IC_{50}^{ino}$ in the range 0.0–2.0.

For statistical modeling of values of $\log IC_{50}^{ino}$, the data set was divided into four arbitrary subsets, two

containing 8 and two containing 9 inhibitors. The first subset contained those inhibitors numbered **001**, **005**, **009**,... in Table 1; the second contained those inhibitors numbered **002**, **006**, **010**,... in Table 1, and so forth. Sets of molecular descriptors based on earlier experience were selected for the modeling work. These include about 400 2D and 3D descriptors traditionally employed in development of QSAR models as well as a larger set of descriptors based on electron densities at the molecular surface derived employing TAE technology. Four modeling runs were completed, using each of the subsets once as test set and the remaining three subsets as training set. Thus, either 25 or 26 compounds were employed as training sets to predict values of IC_{50}^{ino} for either 8 or 9 test compounds. Each compound appeared once in a test set (a prediction) and three times in a training set (an estimation). Similarly, training sets of 21 or 22 compounds were employed to predict values of $\log K_i^{ino}$ values for 8 or 7 test compounds, and training sets of 21 compounds were employed to predict values of $\log K_i^{mesg}$ for 7 test compounds. Values of $\log K_i^{mesg}$, $\log K_i^{ino}$, and $\log IC_{50}^{ino}$ at 10 μM inosine were modeled. Results of the modeling effort for values of $\log K_i^{mesg}$ are collected in Table 2 and presented graphically in Figure 2. Those for values of $\log K_i^{ino}$ and $\log IC_{50}^{ino}$ are presented graphically in Figures 3 and 4.

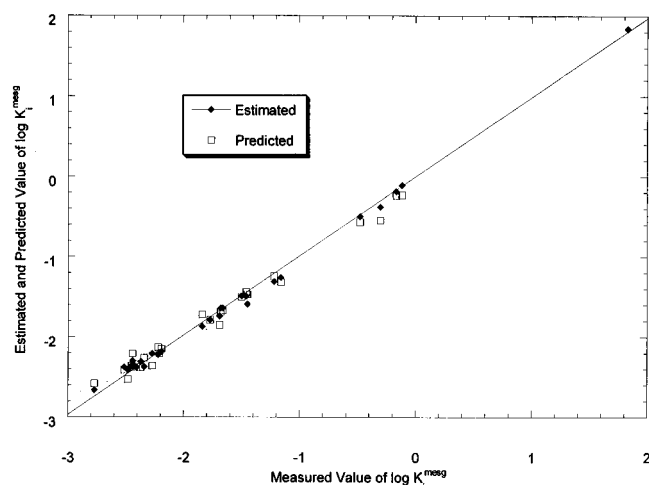
Training Set Characterization. In each of the 12 modeling efforts (four each for values of $\log K_i^{mesg}$ and $\log K_i^{ino}$ and four for values of $\log IC_{50}^{ino}$), we were able to adequately characterize the training set. Plots (not shown) of $\log(\text{estimated})$ vs $\log(\text{measured})$ have values of R^2 which vary from 0.93 to 0.99 for the 12 runs. In Table 2 and Figures 2–4, estimated values were taken as the average of the three values obtained for each compound. For 27 of the 28 compounds for which values of $\log K_i^{mesg}$ were estimated and for 28 of the 29 inhibitors for which values of $\log K_i^{ino}$ were estimated, the three values were highly consistent, with standard deviations generally less than 0.10 log unit (Table 2 and Figures 2 and 3). The exceptions are compound **028** for $\log K_i^{mesg}$, for which only one qualified estimate was obtained (Table 2), and compound **019**, for which the estimated value for $\log K_i^{ino}$ is 1.29 ± 0.39 . Results are presented graphically in Figures 2 and 3. Results of comparable quality were obtained for values of $\log IC_{50}^{ino}$ (Figure 4). Again, the single exception to consistency in the estimated values of $\log IC_{50}^{ino}$ was compound **019** for which a value of 0.76 ± 0.23 was obtained.

Predictions. Comparisons of measured and predicted values for $\log K_i^{mesg}$, $\log K_i^{ino}$, and $\log IC_{50}^{ino}$ are provided in Table 2 and Figures 2–4. For 25 of the 28 inhibitors, predicted values of $\log K_i^{mesg}$ are within the estimated experimental error (± 0.2 log unit) of the measured values (Table 2). No qualified prediction for compound **028** was obtained, and predicted values for compounds **002** and **022** are just slightly beyond the limits of experimental error. Similarly, for 24 of the 29 inhibitors, predicted values for $\log K_i^{ino}$ are within the estimated experimental error of the measurements. For an additional three inhibitors, predictions are within 2 standard deviations of the estimated measurement error. Predictions for the two weakest inhibitors in the data set, **027** and **028**, are more substantially in error.

Table 2. Measured, Estimated, and Predicted Values of $\log K_i^{\text{mesg}}$ for a Series of Inhibitors of Calf Spleen PNP^a

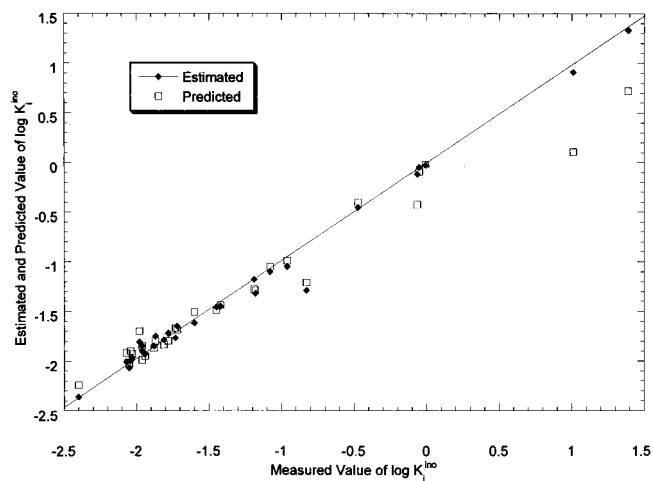
inhibitor	$\log K_i^{\text{mesg}}$ measured	$\log K_i^{\text{mesg}}$ estimated or predicted ^b				Δ	
		1	2	3	4	estimated ^c	predicted ^d
001	-2.41	-2.38*	-2.37	-2.45	-2.31	-0.03	-0.03
002	-2.44	-2.30	-2.21*	-2.27	-2.33	-0.11	-0.23
003	-2.37	-2.33	-2.31	-2.38*	-2.29	-0.06	0.01
004	-2.45	-2.35	-2.38	-2.35	-2.36*	-0.09	-0.09
005	-1.84	-1.72*	-1.86	-1.84	-1.91	0.03	-0.12
006	-2.77	-2.78	-2.58*	-2.75	-2.86	-0.11	-0.19
007	-1.46	-1.51	-1.50	-1.44*	-1.46	0.03	-0.02
008	-2.19	-2.23	-2.16	-2.16	-2.15*	-0.01	-0.04
009	-2.21	-2.21*	-2.20	-2.20	-2.18	-0.02	0
010	-2.41	-2.40	-2.34*	-2.40	-2.35	-0.03	-0.07
011	-2.22	-2.21	-2.28	-2.131*	-2.16	0	-0.09
012	-2.48	-2.39	-2.38	-2.47	-2.53*	-0.07	0.05
013	-1.66	-1.67*	-1.59	-1.63	-1.69	-0.02	0.01
014	-2.34	-2.32	-2.26*	-2.42	-2.36	0.03	-0.08
015	-1.22	-1.39	-1.31	-1.24*	-1.24	0.09	0.02
016	-2.51	-2.45	-2.39	-2.31	-2.41*	-0.13	-0.10
017	-1.69	-1.85*	-1.67	-1.85	-1.70	0.05	0.16
018	-1.68	-1.65	-1.69*	-1.59	-1.68	-0.04	0.01
019	-1.16	-1.28	-1.29	-1.32*	-1.21	0.10	0.16
020	-1.50	-1.46	-1.53	-1.48	-1.50*	-0.01	0
021	-0.478	-0.567*	-0.516	-0.521	-0.473	-0.01	0.089
022	-0.301	-0.311	-0.549*	-0.453	-0.372	0.08	0.248
023	-0.116	-0.110	-0.095	-0.231*	-0.116	-0.01	0.115
024	-2.27	-2.25	-2.23	-2.16	-2.36*	-0.06	0.09
025	-1.77	-1.79*	-1.77	-1.84	-1.77	0.02	0.02
026	-1.45	-1.62	-1.47*	-1.61	-1.55	0.14	0.02
028	1.83	1.84	-	-*	-	-	-
029	-0.167	-0.180	-0.153	-0.227	-0.245*	0.02	0.078

^a All values of K_i were determined at pH 7.4, 1 mM phosphate, and 25 °C. ^b Predicted values are the single values obtained for each inhibitor as a member of a test set and are designated by asterisks; estimated values are averages of the three values obtained for each inhibitor as a member of training sets. ^c The difference between logarithms of measured and estimated values. ^d The difference between the logarithms of measured and predicted values.

**Figure 2.** Plot of logarithms of measured values for K_i^{mesg} against the corresponding estimated (squares) and predicted (circles) values for a set of inhibitors of calf spleen PNP.

Values of $\log K_i^{\text{ino}}$ for the last two inhibitors are substantially greater those for any compound employed in the training set, requiring an extrapolation of more than 1 order of magnitude. Results are presented graphically in Figure 3. Similarly, 28 of the 34 predictions of $\log IC_{50}^{\text{ino}}$ are within the estimated experimental error, and those for four additional compounds are within 2 standard deviations of the measured value (Figure 4). Here, too, the estimates for **027** and **028** are substantially in error.

We carried out an independent (and chronologically earlier) effort to model the values of $\log IC_{50}^{\text{ino}}$ (Table 1) employing a different set of molecular descriptors.

**Figure 3.** Plot of logarithms of measured values for K_i^{ino} against the corresponding estimated (squares) and predicted (circles) values for a set of inhibitors of calf spleen PNP.

The inhibitors of Table 1 were divided into three sets: (i) a training set of 23 compounds, (ii) a test set of 5 compounds, and (iii) an unknown set of 5 compounds.

At the outset of this modeling work, the values of IC_{50}^{ino} for compounds in the training and test sets were known to us. Values of IC_{50}^{ino} for the five compounds in the unknown set were not known to us at the time the modeling was done. Samples of these inhibitors were subsequently supplied to us and values of IC_{50}^{ino} determined.

Measured, estimated, and predicted values of $\log IC_{50}^{\text{ino}}$ for the data set are collected in Table 3. The statistical model does an adequate job of characterizing

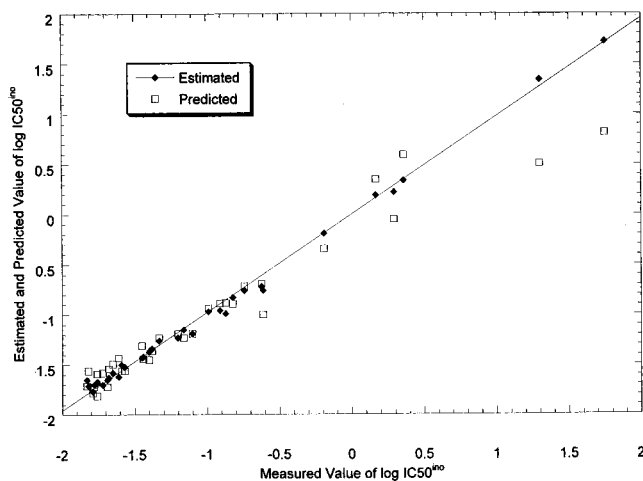


Figure 4. Plot of logarithms of measured values of IC_{50}^{ino} against the corresponding estimated (squares) and predicted (circles) values for a set of inhibitors of calf spleen PNP.

the training set: the difference between measured and estimated values is less than 2 standard deviations of the measurement error for all 23 compounds and within 1 standard deviation of the measurement error for 16 of the 23 compounds. There is a tendency for the model to estimate values of $\log IC_{50}^{ino}$ slightly greater than the measured values. The more detailed modeling effort described earlier does a somewhat better job of characterizing the data set than does this effort. However, comparison of the average of the three estimates in Figure 4 with the single estimates in Table 3 reveals that they are generally consistent, with no major discrepancies.

In addition to adequately characterizing the training set, this model has predictive power for structures and values of $\log IC_{50}^{ino}$ for compounds outside the training set. Of particular note are the results for the test set (Table 3). For all five compounds, predicted values are within 1 standard deviation of the experimental error. This is notable since two of the compounds, **027** and **028**, have property values at least 1 order of magnitude greater than that for any compound in the training set (Figure 1). Hence this statistical model has at least some power to extrapolate to property values outside those of the training set. In addition, three of the compounds in the test set are structurally distinct from the training set inhibitors. With three exceptions, the training set consists of 9-substituted-9-deazaguanines. The test set includes the only 3-substituted-hypoxanthine, the only 9-substituted-8-methyl-9-deazaguanine, and the only inhibitor not bearing a substituent on the purine ring. Thus, it appears that the model has at least modest ability to extrapolate to structures beyond those in the training set.

The predicted values for the unknown set reveal that further progress is required to develop a really robust statistical model. For three of the five unknowns, the model predicts the property values within 1 standard deviation of the experimental error. However, estimated values of IC_{50}^{ino} for two of the inhibitors are 1 order of magnitude greater than the measured values.

Discussion

Estimates and predictions made employing a set of molecular descriptors which have generally worked well

Table 3. Logarithms of Measured, Estimated, and Predicted Values of IC_{50}^{ino} for a Set of Inhibitors of Calf Spleen PNP^a

A. Training Set			
compd	log		Δ estimated ^b
	measured	estimated	
001	-1.76	-1.65	-0.11
002	-1.78	-1.86	0.08
003	-1.59	-1.24	-0.35
004	-1.65	-1.54	-0.11
005	-0.99	-0.87	-0.12
006	-1.44	-1.15	-0.29
007	-0.62	-0.36	-0.26
009	-1.69	-1.61	-0.08
010	-1.83	-1.96	0.13
011	-1.76	-1.79	0.03
012	-1.72	-1.51	-0.21
013	-1.45	-1.28	-0.17
014	-1.79	-1.87	0.08
015	-0.60	-0.61	0.01
016	-1.66	-1.34	-0.32
017	-0.91	-0.87	-0.04
018	-1.16	-1.06	-0.10
019	-0.61	-0.66	0.05
020	-0.76	-0.52	-0.24
021	-0.19	-0.18	-0.01
022	0.29	0.24	0.05
023	0.17	0.15	0.02
024	-1.61	-1.37	-0.24

B. Test Set			
compd	log		Δ predicted ^c
	measured	predicted	
025	-1.20	-1.08	-0.12
026	-0.87	-0.90	0.03
027	1.75	1.70	0.05
028	1.31	1.15	0.16
029	0.36	0.56	-0.20

C. Unknown Set			
compd	log		Δ predicted ^c
	measured	predicted	
030	-1.82	-0.72	-1.10
031	-1.40	-0.26	-1.14
032	-1.38	-1.34	-0.04
033	-0.82	-1.00	0.18
034	-1.10	-1.10	0.00

^a Logarithms have been taken of values expressed in micromolar. ^b The difference between logarithms of measured and estimated values. ^c The difference between logarithms of measured and predicted values.

for us in QSAR modeling efforts adequately characterize the training sets and generally provide predicted values within 1 standard deviation of the measurement error (Table 2 and Figures 2–4). Given that the data set contains a few structural outliers among a larger series of 9-substituted-9-deazaguanines and a few rather weak inhibitors among a larger collection of substantially more potent ones (Figure 1), these results are certainly encouraging. They suggest that the current QSAR model has substantial power to predict values of K_i and IC_{50} for novel, potent 9-substituted-9-deazaguanines and useful power to predict such values for reasonably potent inhibitors outside this structural class. Creation of a more robust QSAR model will require greater structural diversity and a larger and more uniform distribution of K_i or IC_{50} values in the data set. Work to achieve these goals is underway.

It is not surprising that the largest differences between measured and predicted values are those for

the weakest inhibitors, for which a substantial extrapolation is required. Our models underestimate the values for K_i and IC_{50} for both **027** and **028**; that is, we predict these inhibitors to be more potent than they are. Thus, our QSAR model pulls the estimated values back toward the mean of the values of the training set. However, using a different set of molecular descriptors, we were successful in developing a QSAR model which predicted values for these inhibitors within experimental error (Table 3). Thus, it is clear that QSAR models for this data set can be created which have the power to extrapolate, but it is not clear how to choose a set of molecular descriptors to achieve this goal.

Among compounds that have values of K_i and IC_{50} that fall within the range of those in the training sets, **019** proved most problematic, in terms of both training set characterization and prediction; our predicted value is in error by about 2 standard deviations of the measurement error. It has previously been observed (J. A. Montgomery, personal communication) that the measured value of IC_{50}^{ino} for **019** varies by more than 1 order of magnitude depending on whether the reaction is initiated by addition of inosine or enzyme. This behavior has been confirmed in our Laboratory; the mechanistic basis of this behavior is unclear. It is not observed for a number of other inhibitors examined in this study. It is not clear whether our difficulties in accurately predicting values for this inhibitor reflect an inadequacy of our QSAR model at its present state of development or a subtle mechanistic difference between this inhibitor and others in the data set leading to an uncertainty in the pertinent measured value.

Measured values of K_i for most of the inhibitors employed in this study are dependent on the nature of the substrate used (Table 1). In most, but not all, cases values of K_i^{mesg} are smaller than those of K_i^{ino} . It has been established through structural studies⁶⁻⁹ as well as kinetic characterization⁶⁻⁹ that these are competitive inhibitors, a result confirmed in the present work. Substrate-dependent values of K_i for a set of competitive inhibitors is a highly surprising result, for which we know of no precedent. Since values of K_i are equilibrium constants for dissociation of E·I complexes, this result demands that E is in some sense different when MESG and inosine are employed as substrates.

Onuffer and Kirsch have provided a possible model for our results.¹⁰ Anomalous behavior of a site-specific mutant (D222A) of *Escherichia coli* aspartate aminotransferase demands the existence of two slowly interconverting enzyme forms. By analogy, it is possible that (a) PNP emerges from a catalytic event in one of two forms, depending on whether MESG or inosine is employed as substrate; (b) interconversion of these two forms is slow compared to the rate of a catalytic cycle; and (c) the two forms have differential affinities for the set of inhibitors employed in this work. Thus, in effect, the inhibitors are binding to two different enzyme structures. Definition of the precise basis for substrate-dependent values of K_i for competitive inhibitors will require detailed further study.

Experimental Section

Materials. 2-Amino-6-mercapto-7-methylpurine ribonucleoside (MESG) was synthesized as previously described.¹¹ All PNP inhibitors employed were synthesized by scientists at

BioCryst Pharmaceuticals Inc. and are the gift of that organization. With the exception of compounds **013** and **019** (see Table 1 for structures), which are enantiomers, other inhibitors bearing one or two chiral centers were studied as racemates or diastereomeric mixtures. Calf spleen PNP (phosphate-free, purity greater than 98%) and xanthine oxidase were the best grades available from Sigma Chemicals and were used without further purification. Other reagents were obtained commercially from either Sigma or Aldrich and were the best grade available.

Kinetic measurements were carried out spectrophotometrically with aid of an HP 8453 diode array spectrophotometer. Rates were demonstrated to increase linearly with increasing enzyme concentration. Values of K_i employing MESG and 1 mM phosphate as substrate were generally carried out with an automated enzyme kinetic workbench built around an HP ORCA laboratory robot equipped with an HP 8453 spectrophotometer.¹¹ In several cases, values of K_i employing MESG as substrate were also measured manually. Results obtained with the robotic workbench and those obtained manually gave concordant results. Values of K_i employing inosine and 1 mM phosphate as substrate were determined manually employing a standard coupled assay.¹² It was demonstrated that these values are independent of the concentration of the coupling enzyme employed in the assay. All values of K_i were determined by making rate measurements at five inhibitor concentrations for each of five substrate concentrations. In all cases, the kinetic data revealed, as expected,⁶⁻⁹ competitive inhibition. Each determination of K_i also yielded five values of IC_{50} , one for each concentration of inosine or MESG employed. As expected, values of IC_{50} increased linearly with substrate concentration, and extrapolation of these values to zero substrate concentration provided values of K_i consistent with those based on statistical analysis of all data points. Values of IC_{50} were independently determined by making rate measurements for at least five inhibitor concentrations employing 10 μ M inosine and 1 mM phosphate as substrates. Measurements were usually made at an enzyme concentration of approximately 4.0 nM (as the monomer). In several cases, measurements of K_i and IC_{50} were made over the enzyme concentration range 2.3–4.6 nM. Over this concentration range, these values were independent of enzyme concentration. All measurements were made at 25 °C, pH 7.4 (maintained with 0.1 M *N*-(2-hydroxyethyl)piperazine-*N*-(2-ethanesulfonic acid) sodium salt (HEPES) buffer). Reactions employing MESG as substrate were followed at 356 or 357 nm. Those employing inosine as substrate were followed at 293 nm. Kinetic parameters were estimated from the collected data employing the Leonora statistics package. Values of pH were determined with a Radiometer PHM240 pH meter.

Statistical modeling was carried out employing software based on a novel QSAR paradigm. Several essential features of this paradigm have been published.¹³ Briefly, a set of molecular descriptors, including those 2D and 3D descriptors generally employed in QSAR calculations, a set of quantum-mechanical descriptors,¹⁴⁻¹⁶ and many based on transferable atom equivalent (TAE) technology,^{13,17} was calculated for a single optimized conformation of each molecule in the data set. Optimized conformations were obtained employing CONCORD and Mopac; precisely the same sequence of operations was performed for each member of the data set. The data set modeled was then decomposed into subsets employing a mixture of regression models algorithm. QSAR models for each subset were then qualified for their ability to predict property values for each molecule in the training set but outside the subset, employing one of several sets of molecular descriptors. Only those subsets passing this criterion were retained. Partial least-squares (PLS) statistical models were then developed and optimized for each qualified subset. Those PLS models passing qualifying hurdles were then incorporated into a matrix-of-models. The matrix-of-models was then employed to predict property values for molecules outside those of the training set. Details will be reported elsewhere.

Acknowledgment. We are grateful to Dr. J. A. Montgomery and his colleagues at BioCryst Pharmaceuticals, Inc. for the gift of the inhibitors employed in this work and for useful scientific guidance.

Supporting Information Available: Detailed tables in the format of Table 2 for measured, estimated, and predicted values of $\log K_i^{\text{ino}}$ and $\log \text{IC}_{50}^{\text{ino}}$. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Hansch, C.; Hoekman, D.; Gao, H. Comparative QSAR: Toward a Deeper Understanding of Chemobiological Interactions. *Chem. Rev.* **1996**, *96*, 1045–1075.
- Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
- Sircar, J. C.; Gilbertsen, R. B. Purine Nucleoside Phosphorylase (PNP) Inhibitors: Potential Selective Immunosuppressive Agents. *Drugs Future* **1988**, *13*, 653–688.
- Montgomery, J. A. Inhibitors of Purine Nucleoside Phosphorylase. *Expert. Opin. Invest. Drugs* **1994**, *3*, 1303–1313.
- Stoeckler, J. D. Purine Nucleoside Phosphorylase: A Target for Chemotherapy. In *Developments in Cancer Chemotherapy*; Glazer, R. I., Ed.; CRC Press: Boca Raton, FL, 1984; pp 35–60.
- Montgomery, J. A.; Niwas, S.; Rose, J. D.; Secrist, J. A., III; Babu, Y. S.; Bugg, C. E.; Erion, M. D.; Guida, W. C.; Ealick, S. E. Structure-Based Design of Inhibitors of Purine Nucleoside Phosphorylase. 1. 9-(Arylmethyl) Derivatives of 9-Deazaguanine. *J. Med. Chem.* **1993**, *36*, 55–69.
- Secrist, J. A., III; Niwas, S.; Rose, J. D.; Babu, Y. S.; Bugg, C. E.; Erion, M. D.; Guida, W. C.; Ealick, S. E.; Montgomery, J. A. Structure-Based Design of Inhibitors of Purine Nucleoside Phosphorylase. 2. 9-Alicyclic and 9-Heteroalicyclic Derivatives of 9-Deazaguanine. *J. Med. Chem.* **1993**, *36*, 1847–1854.
- Erion, M. D.; Niwas, S.; Rose, J. D.; Anathan, S.; Allen, M.; Secrist, J. A., III; Babu, Y. S.; Bugg, C. E.; Guida, W. C.; Ealick, S. E.; Montgomery, J. A. Structure-Based Design of Inhibitors of Purine Nucleoside Phosphorylase. 3. 9-Arylmethyl Derivatives of 9-Deazaguanine Substituted on the Methylene Group. *J. Med. Chem.* **1993**, *36*, 3771–3783.
- Guida, W. C.; Elliott, R. D.; Thomas, H. J.; Secrist, J. A., III; Babu, Y. S.; Buff, C. E.; Erion, M. D.; Ealick, S. E.; Montgomery, J. A. Structure-Based Design of Inhibitors of Purine Nucleoside Phosphorylase. 4. A Study of Phosphate Mimics. *J. Med. Chem.* **1994**, *37*, 1109–1114.
- Onuffer, J. J.; Kirsch, J. F. Characterization of the Apparent Negative Cooperativity Induced in *Escherichia coli* Aspartate Aminotransferase by the Replacement of Asp222 with Alanine. Evidence for an Extremely Slow Conformational Change. *Protein Eng.* **1994**, *7*, 413–424.
- Cheng, J.; Farutin, V.; Wu, Z.; Jacob-Mosier, G.; Riley, B.; Hakimi, R.; Cordes, E. H. Purine Nucleoside Phosphorylase Catalyzed Phosphate-Independent Hydrolysis of 2-Amino-6-Mercapto-7-Methylpurine Ribonucleoside. *Bioorg. Chem.* **1999**, *42*, in press.
- Tuttle, J. V.; Krenitsky, T. A. Effects of Acyclovir and its Metabolites on Purine Nucleoside Phosphorylase. *J. Biol. Chem.* **1984**, *259*, 4065–4069.
- Breneman, C. M.; Rhem, M. QSPR Analysis of HPLC Column Capacity Factors for a Set of High-Energy Materials Using Electronic van der Waals Surface Property Descriptors Computed by Transferable Atom Equivalent Method. *J. Comput. Chem.* **1997**, *18*, 182–197.
- Karleson, M.; Labanov, V. S.; Katritsky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR. *Chem. Rev.* **1996**, *96*, 1027–1043.
- Wilson, L. Y.; Famini, G. R. Using Theoretical Descriptors in Quantitative Structure–Activity Relationships: Some Toxicological Indices. *J. Med. Chem.* **1991**, *34*, 1668–1674.
- Breneman, C. M.; Martinov, M. The Use of Electrostatic Potential Fields in QSAR and QSPR. In *Molecular Electrostatic Potentials: Concepts and Applications*; Murray, J. S., Sen, K., Eds.; Elsevier Science B.V.: Amsterdam, 1996.
- Breneman, C. M.; Thompson, T. R.; Rhem, M.; Dung, M. Electron Density Modeling of Large Systems Using the Transferable Atom Equivalent Method. *Comput. Chem.* **1995**, *19*, 161–179.

JM990037Y