Two- and three-dimensional quantitative structure–activity relationships for a series of purine nucleoside phosphorylase inhibitors

Marcelo S. Castilho,[a] Matheus P. Postigo,[a] Caroline B. V. de Paula,[a] Carlos A. Montanari,[b,†] Glaucius Oliva[a] and Adriano D. Andricopulo[a,*]

[a]*Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, 13560-970, São Carlos-SP, Brazil*
[b]*Núcleo de Estudos em Química Medicinal—NEQUIM, Universidade Federal de Minas Gerais, Campus da Pampulha, 31270-901, Belo Horizonte-MG, Brazil*

Abstract—Comparative molecular field analysis (CoMFA), comparative molecular similarity indices analysis, and hologram quantitative structure–activity relationship (HQSAR) studies were conducted on a series of 52 training set inhibitors of calf spleen purine nucleoside phosphorylase (PNP). Significant cross-validated correlation coefficients (CoMFA, $q^2 = 0.68$; CoMSIA, $q^2 = 0.66$; and HQSAR, $q^2 = 0.70$) were obtained, indicating the potential of the models for untested compounds. The models were then used to predict the inhibitory potency of 16 test set compounds that were not included in the training set, and the predicted values were in good agreement with the experimental results. The final QSAR models along with the information gathered from 3D contour and 2D contribution maps should be useful for the design of novel inhibitors of PNP having improved potency.
© 2005 Elsevier Ltd. All rights reserved.

1. Introduction

Purine nucleoside phosphorylase (PNP, EC 2.4.2.1) is a key enzyme in the purine salvage pathway. PNP catalyzes the reversible phosphorolysis of 2′-deoxypurine ribonucleosides to the corresponding bases and 2-deoxyribose 1-phosphate.[1–3] Inhibitors of PNP have received special attention because they possess a wide spectrum of plausible utilities in clinical medicine, including the treatment of T-cell proliferative diseases such as T-cell leukemias or lymphomas, for the prevention of transplant rejection, and in the treatment of T-cell autoimmune diseases such as rheumatoid arthritis and systemic lupus erythematosus (SLE).[4–6] Despite the significant therapeutic potential of PNP inhibitors, which has been recognized for more than twenty years, only two small-molecule inhibitors from BioCryst Pharmaceuticals, forodesine hydrochloride (BCX-1777) and BCX-4208, are currently in clinical trials for the control of T-cell malignancies.[7,8]

Quantitative structure–activity relationships (QSAR) have been successfully employed to create predictive models as a valuable tool to facilitate the discovery of enzyme inhibitors.[9–11] In the literature, there are only few studies on the quantitative relationship between the chemical structure information and the inhibitory potency of PNP inhibitors,[12,13] and to the best of our knowledge, no three-dimensional QSAR investigation has been reported to date.

In the present study, we have collected values of IC_{50} for a series of ground-state inhibitors of calf spleen PNP and used the data to create 3D and 2D QSAR models which show substantial predictive promise. The QSAR studies were carried out using comparative molecular field analysis (CoMFA),[14] comparative molecular similarity indices analysis (CoMSIA),[15] and hologram QSAR (HQSAR).[16] The results of modeling this data set are reported herein.

2. Materials and methods

2.1. Computational approach

The QSAR modeling analyses, calculations, and visualizations for CoMFA, CoMSIA, and HQSAR were performed using the SYBYL 6.9.2 package (Tripos Inc., St. Louis, USA) running on Red Hat Linux 7.3 workstations. Docking protocol as implemented in GOLD 2.1 (Cambridge Crystallographic Data Centre, Cambridge, UK) was employed to search the possible binding conformations of ligands into the PNP active site. The X-ray crystallographic data for PNP in complex with 1,4-dideoxy-4-aza-1-(S)-(9-deazahypoxanthin-9-yl)-D-ribitol (PDB code: 1B8O) used in docking simulations were retrieved from the Protein Data Bank (PDB).

2.2. Kinetic measurements

Calf spleen PNP (phosphate-free, purity greater than 98%) and xanthine oxidase were of 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 of the best grade available. Kinetic measurements were carried out spectrophotometrically with the aid of a Shimadzu UV–vis spectrophotometer, using a standard coupled assay as previously described.[12,13] All measurements were made at 25 °C and pH 7.4, employing 10 μM inosine and 1 mM phosphate as substrates. Kinetic parameters were estimated from the collected data employing the SigmaPlot enzyme kinetics module. Values of IC_{50} were independently determined by making rate measurements for at least five inhibitor concentrations. The values represent means of at least three individual experiments.

2.3. Data sets

The data set used for the 3D and 2D QSAR analyses contains 68 inhibitors of calf spleen PNP, which were selected from our own data and literature data.[12,13,17–21] Values of IC_{50} for a number of literature inhibitors, employing inosine as substrate, were determined in our laboratory and the results are in good agreement with those previously reported. The structures and corresponding IC_{50} values for the complete set of compounds are included in Table 1. The 3D structures of the PNP inhibitors were constructed using CONCORD and standard geometric parameters of molecular modeling software package SYBYL 6.9.2. Each single optimized conformation of each molecule in the data set was energetically minimized employing the atom-centered partial charge MNDO-ESP calculations implemented in MOPAC 6.0.

2.4. Molecular modeling

The advanced molecular docking program GOLD 2.1 uses a genetic algorithm as the conformational search tool and different scoring functions (e.g., GOLDscore, CHEMscore) to rank the pose thus obtained. All docking simulations were carried out with the X-ray coordinates of the calf spleen PNP structure to which hydrogen atoms were added in standard geometry, after

removal of the ligand and water molecules. The active site was centered on C1 atom of the 1,4-dideoxy-4-aza-1-(S)-(9-deazahypoxanthin-9-yl)-D-ribitol and a radius of 12 Å was considered during the docking procedure. His, Gln, and Asn were manually checked for possible flipped orientation, protonation, and tautomeric states with Pymol (DeLano Scientific, San Carlos, USA) side-chain wizard script. Minimized inhibitor structures were used in the docking protocol, which was repeated 10 times for each inhibitor due to the stochastic nature of the search algorithm. Default parameters and GOLDscore function were employed in all runs, and only the best ranked conformation of each inhibitor was considered for 3D QSAR studies.

2.5. Molecular alignment

Alignment is a crucial component in 3D QSAR studies, and a variety of useful approaches have been described in the literature.[14,15,22–25] Three alignment approaches were used in the present study:

1. Alignment 1. The lowest energy conformer of compound **28** (the most potent inhibitor of the data set) was used as template structure for the molecular alignment. All molecules were aligned by using simple align database (rigid fit) option in SYBYL 6.9.2. Alignment and common purine base atoms used for superimposition are shown in Figure 1.
2. Alignment 2. This alignment was generated by the docking protocol implemented in GOLD 2.1.[15,26,27] The aligned molecules were then exported to SYBYL and MOPAC charges were assigned before CoMFA and CoMSIA field calculations. The aligned data set is depicted in Figure 2A.
3. Alignment 3. Conformations generated by the docking program GOLD 2.1 were aligned on template molecule **28** according to the same protocol used in alignment 1. The aligned molecules of the data set are depicted in Figure 2B.
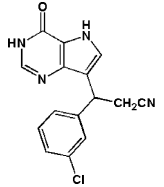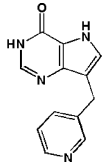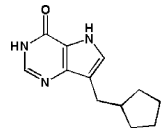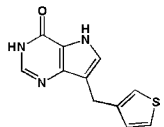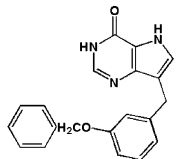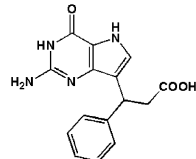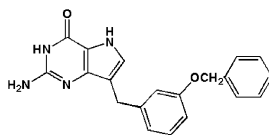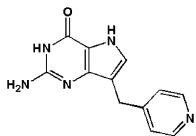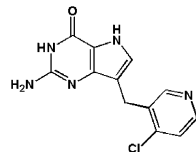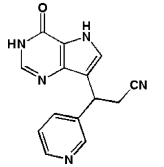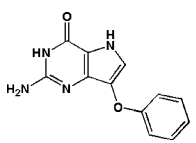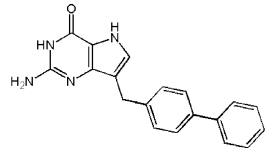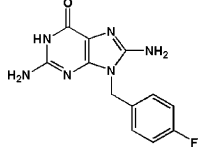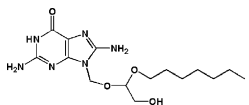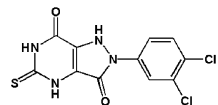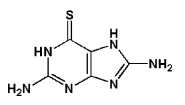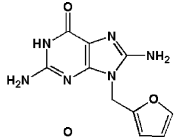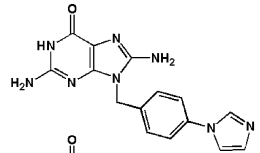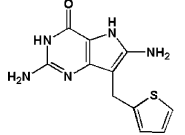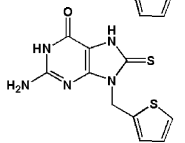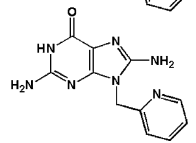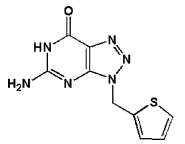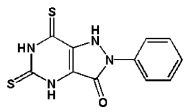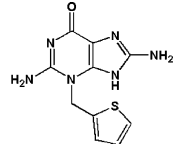
2.6. 3D QSAR studies

To better understand and explore the contributions of electrostatic, steric, and hydrophobic fields in the binding of the data set inhibitors to PNP, and to build predictive 3D QSAR models, CoMFA and CoMSIA studies were performed based on the three molecular alignments described. CoMFA calculates steric and electrostatic properties according to Lennard-Jones and Coulomb potentials, respectively, whereas, CoMSIA calculates similarity indices in the space surrounding each of the molecules in the data set.

2.6.1. CoMFA. The aligned training set molecules were placed in a 3D grid box such that the entire set was included in it. CoMFA steric and electrostatic fields were generated at each grid point with Tripos force field using an sp^3 carbon atom probe carrying a +1 net charge. The CoMFA grid spacing of 2.0 Å in the x, y, and z directions, and the grid region were automatically generated by the CoMFA routine to encompass all molecules with an extension of 4.0 Å in each direction. CoMFA
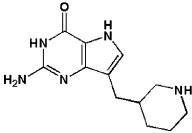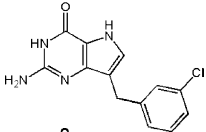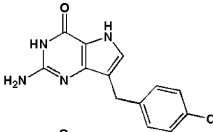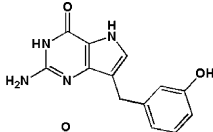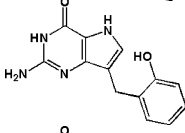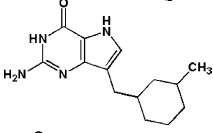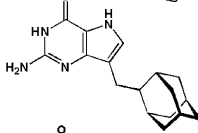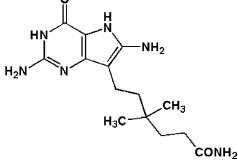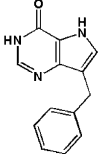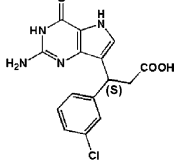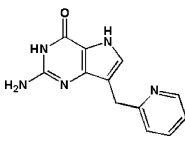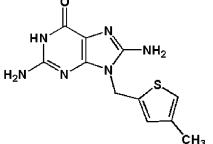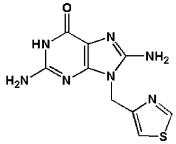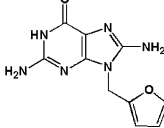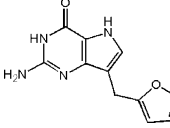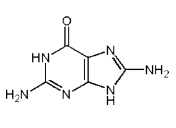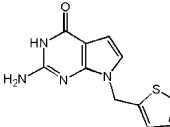
Table 1. Chemical structures and corresponding IC_{50} values (μM) for a series of inhibitors of calf spleen purine nucleoside phosphorylase

Compound	Structure	IC_{50}	Compound	Structure	IC_{50}	Compound	Structure	IC_{50}
				Training set molecules				
1		0.017 ± 0.007	2		0.021 ± 0.004	3		0.023 ± 0.006
4		0.024 ± 0.007	5		0.025 ± 0.013	6		0.025 ± 0.003
7		0.036 ± 0.012	8		0.051 ± 0.012	9		0.057 ± 0.011
10		0.082 ± 0.025	11		0.120 ± 0.040	12		0.147 ± 0.070
13		0.240 ± 0.020	14		0.160 ± 0.013	15		0.011 ± 0.009
16		0.018 ± 0.010	17		0.025 ± 0.006	18		0.029 ± 0.0019
19		0.030 ± 0.017	20		0.047 ± 0.014	21		0.070 ± 0.020
22		1.900 ± 0.033	23		6.600 ± 0.700	24		100.0 ± 8.0
25		260.0 ± 21.0	26		50.0 ± 5.0	27		0.011 ± 0.001

28	0.010 ± 0.01	**29**	0.040 ± 0.004	**30**	0.029 ± 0.003
31	0.028 ± 0.003	**32**	0.190 ± 0.015	**33**	0.047 ± 0.004
34	0.124 ± 0.018	**35**	0.064 ± 0.004	**36**	0.135 ± 0.051
37	0.039 ± 0.003	**38**	0.042 ± 0.002	**39**	0.546 ± 0.320
40	4.170 ± 0.410	**41**	1.520 ± 0.210	**42**	7.030 ± 0.200
43	8.490 ± 0.210	**44**	0.092 ± 0.004	**45**	0.852 ± 0.038
46	0.497 ± 0.026	**47**	2.866 ± 0.290	**48**	1.971 ± 0.150
49	1.839 ± 0.060	**50**	22.840 ± 1.820	**51**	1.758 ± 0.090

M. S. Castilho et al. / Bioorg. Med. Chem. 14 (2006) 516–527

519

Table 1 (*continued*)

Compound	Structure	IC$_{50}$	Compound	Structure	IC$_{50}$	Compound	Structure	IC$_{50}$
			52		1.000 ± 0.099			

<div align="center">Test set molecules</div>

Compound	Structure	IC$_{50}$	Compound	Structure	IC$_{50}$	Compound	Structure	IC$_{50}$
53		0.020 ± 0.007	54		0.025 ± 0.001	55		0.070 ± 0.004
56		0.270 ± 0.016	57		0.025 ± 0.003	58		0.090 ± 0.005
59		200.0 ± 25.0	60		0.035 ± 0.002	61		0.036 ± 0.003
62		0.015 ± 0.001	63		0.362 ± 0.014	64		1.043 ± 0.003
65		2.710 ± 0.002	66		0.083 ± 0.005	67		2.299 ± 0.221
			68		0.674 ± 0.032			
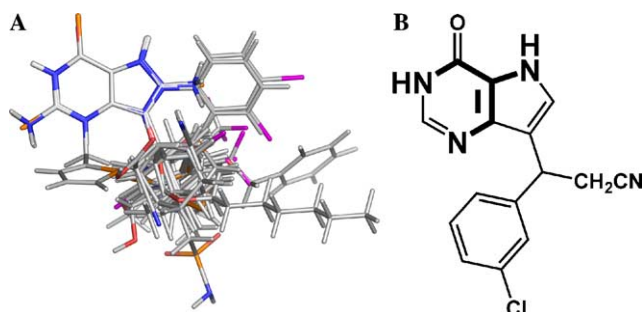
Figure 1. Three-dimensional data set alignment. (A) Alignment 1. (B) Template molecule (inhibitor **28**) used for the molecular alignment (reference atoms in the purine ring are shown in bold).
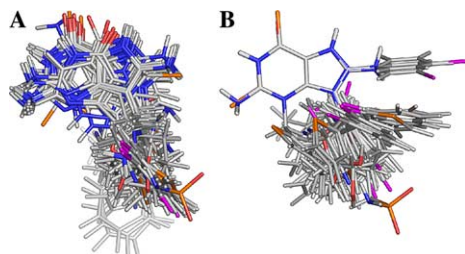
Figure 2. Alignment of the data set molecules for conformations generated by GOLD 2.1. (A) Alignment 2. (B) Alignment 3.

region focusing method was applied to increase the resolution of CoMFA models. The default value of 30 kcal mol^{-1} was set as the maximum steric and electrostatic energy cutoff. Minimum-sigma (column filtering) was set to be 2.0 kcal mol^{-1} to improve the signal-to-noise ratio by omitting those lattice points where energy variation is below the threshold. All models were investigated using full cross-validated (q^2), partial least squares (PLS), leave-one-out (LOO), and leave-many-out (LMO) methods with CoMFA standard options for scaling of variables. Progressive scrambling method was applied to determine the sensitivity of the QSAR models to chance correlations.

2.6.2. CoMSIA. The same training set was adopted for CoMSIA investigations. Electrostatic, steric, and hydrophobic fields were evaluated using the standard parameters implemented in SYBYL 6.9.2. The lattice dimensions were selected with a sufficiently large margin (4.0 Å) to enclose all molecules as in the CoMFA analysis. Any singularities were avoided at atomic positions in CoMSIA fields because a Gaussian-type distance dependence of the physicochemical properties was adopted; thus, no arbitrary cutoffs were required. Similarity indices were computed using a probe with a charge of +1, a radius of +1.0 Å, a hydrophobicity of +1, and 0.3 as attenuation factor for the Gaussian-type distance. The statistical evaluation for the CoMSIA analyses was performed in the same way as described for CoMFA.

2.7. 2D QSAR studies

The necessity of structural alignment and descriptor calculation makes use of 3D QSAR non-trivial. A powerful strategy is to search for synergies in combining 2D and 3D methods in the development of predictive QSAR models.[28] In this context, predictive 2D QSAR models were constructed for this series of PNP inhibitors using the HQSAR technique, which requires only 2D structures and biological activity as input.

2.7.1. HQSAR. In this method, each molecule in the data set is broken down into several unique structural fragments, which are arranged to form a molecular hologram. HQSAR encodes all possible molecular fragments (i.e., linear, branched, and overlapping). With the transformation of the chemical representation of a molecule into its corresponding molecular hologram, this method requires no explicit 3D information for the ligands (e.g., determination of 3D structure, putative binding conformations, and molecular alignment). HQSAR models can be affected by a number of parameters concerning hologram generation: hologram length, fragment size, and fragment distinction. Several combinations of fragment distinction were considered during the QSAR modeling runs. Holograms were generated using 6 distinct fragment sizes. HQSAR analysis was performed by screening the 12 default series of hologram lengths. The patterns of fragment counts from the training set inhibitors were then related to the experimental biological data using PLS analysis.

3. Results and discussion

3.1. Chemical and biological data

3D QSAR CoMFA and CoMSIA, and 2D HQSAR models were derived for a series of 68 ground-state inhibitors of calf spleen PNP. The corresponding values of IC$_{50}$ vary from 0.01 to 260 µM, a factor of 26,000 (Table 1). The in vitro IC$_{50}$ values were converted into the corresponding pIC$_{50}$ ($-$log IC$_{50}$) values. It is worth noting that the IC$_{50}$ values used in this work were measured under the same experimental conditions, a fundamental requirement for QSAR studies. Thus, the data set is appropriate for the purposes of QSAR model development. The generation of reliable models is dependent on the creation of appropriate training and test sets. The data set of 68 inhibitors was divided into training (52 compounds, **1–52**, Table 1) and test (16 compounds, **53–68**, Table 1) sets in the ratio of 3.25:1. A statistical cluster analysis was carried out with Tsar 3D version 3.3 (Accelrys, San Diego, USA) using the complete linkage clustering method (Euclidean distances) with no data standardization. The cluster analysis grouped the compounds into 2 major clusters (amalgamation distance of 0.6). The test set contains 9 compounds in cluster 1 and 7 compounds in cluster 2. The training set contains 33 compounds in cluster 1 and 19 in cluster 2. Another important point addressed in the analysis was the similarity between molecules. Considering that the structural diversity of the data set is modest with the bulk of the compounds being 9-substituted-9-deazaguanines (or 9-guanines), the composition of the QSAR training and test sets is necessarily representative of the whole data set in terms of similarity. Thus, these results show the appropriate chemical diversity and

distribution of biological property across the range of IC_{50} values for both training and test set compounds.

The pIC_{50} values employed in this study span approximately five orders of magnitude and property values, although weighted toward the high-potency end of the spectrum, are acceptably distributed across the range of values. The PLS method was used for all 3D and 2D QSAR analyses. CoMFA, CoMSIA, and HQSAR descriptors were used as independent variables, whereas pIC_{50} values were used as dependent variables in the PLS regression analyses to derive QSAR models. The predictive ability of the models was assessed by their q^2 values.

3.2. 3D QSAR analyses

3.2.1. CoMFA and CoMSIA models. CoMFA and CoMSIA methods are based on the assumption that changes in binding affinities of ligands are related to changes in molecular properties represented by molecular fields. The alignment rule and the bioactive conformation are crucial variables as both will affect the outcome of the 3D statistical analysis. In the present study, three different molecular alignments (alignments 1, 2, and 3) were submitted to CoMFA and CoMSIA analyses, and the statistical results are presented in Table 2. As it can be seen, the best statistical results among all models were obtained for CoMFA 1 (alignment 1), which produced a cross-validated correlation coefficient q^2 of 0.68 with 2 components, and CoMFA 2 (alignment 2), which produced a q^2 of 0.68 with 3 components. Hence, these two models were chosen for further investigations. A conventional non-cross-validated correlation coefficient r^2 of 0.89 was obtained for CoMFA 1 with a standard error of estimate (SEE) of 0.40, while an r^2 of 0.90 with a SEE of 0.40 for CoMFA 2. In all cases, no significant differences were observed using the LMO method. Progressive scrambling of the data set was also carried out to check for possible chance correlations and test the stability of the models. The results of progressive scrambling further confirmed consistency of the models as defined by the critical slope, and optimum statistics for cSDEP and Q^2 obtained at the end of different runs. The r^2 expresses the quality of the data fit rather than the quality of prediction, which is assessed by q^2. Additional components produced no significant increase in the q^2 values and were then discarded from the final models. On the other hand, the best statistical result among all the CoMSIA models was generated employ-

ing alignment 1 (CoMSIA 1). The region focusing was weighted by $SD \times$ coefficient values ranging from 0.3 to 1.5 and grid spacing ranging from 0.5 to 2.0. This strategy not only increased q^2 values during the process of model generation, but also resulted in the refinement of 3D contour maps. CoMFA 1 steric field descriptor explains 53% of the total variance, while the electrostatic descriptor explains the rest 47% (Table 2). CoMFA 2, steric field descriptor, explains 43% of the variance, while the electrostatic descriptor accounts for the rest 57%. CoMSIA 1 steric and hydrophobic fields explain 30% of the variance each, with the electrostatic field being responsible for the other 40%. The CoMFA 2 contour maps for the steric field are shown in Figure 3A and those for the electrostatic field in Figure 3B, both with compound **28** represented as PLS $SD \times$ coefficient plots.

CoMFA 2 contour maps show regions in 3D space where changes of steric and electrostatic fields for 9-substituted purines correlate strongly with concomitant changes in inhibitory potencies.

The CoMFA electrostatic contour map shows several red contours surrounding the group attached at the 9-position of the purine ring in the template molecule, encompassing the aromatic ring and the cyanil moiety, representing regions where electronegative environment would enhance activity. Another red contour, located near the oxygen atom at the 6-position of the purine ring, indicates that high electron density may play a favorable role in inhibitory potency. Two small blue-colored contours represent regions where electropositive groups would contribute to increase inhibitory potency. The steric contour map of the final CoMFA 2 model shows a green region forming a cavity near the cyanil moiety present in the group attached at the 9-position of the purine ring indicating the preference for bulkier substituents in that region. There is also a small green contour near O6 and N7 in the purine ring of the template molecule indicating a sterically favorable region. The disfavored steric yellow region seen above the plane of the aromatic ring in the group attached at the 9-position of the purine ring suggests that bulkier substituents may decrease the PNP inhibitory potency.

The q^2 LOO procedure used may give a suitable representation of the real predictive power of the models for novel inhibitors of PNP. However, the predictive ability of CoMFA 1 and CoMFA 2 models, derived with the 52 training set molecules, was assessed by

Table 2. CoMFA and CoMSIA results (alignments 1, 2, and 3)

Model	q^2	N	r^2	SEE	F	Fraction (%)		
						S	E	H
CoMFA 1	0.68	2	0.89	0.40	89.51	53	47	—
CoMFA 2	0.68	3	0.90	0.40	88.22	43	57	—
CoMFA 3	0.66	3	0.82	0.48	72.89	50	50	—
CoMSIA 1	0.66	3	0.85	0.45	86.64	30	40	30
CoMSIA 2	0.53	3	0.78	0.53	57.73	29	43	28
CoMSIA 3	0.60	2	0.78	0.70	89.46	30	34	36

q^2, Leave-one-out (LOO) cross-validated correlation coefficient; N, optimum number of components; r^2, non-cross-validated correlation coefficient; SEE, standard error of estimate; F, F test value; S, steric field; E, electrostatic field; H, hydrophobic field.
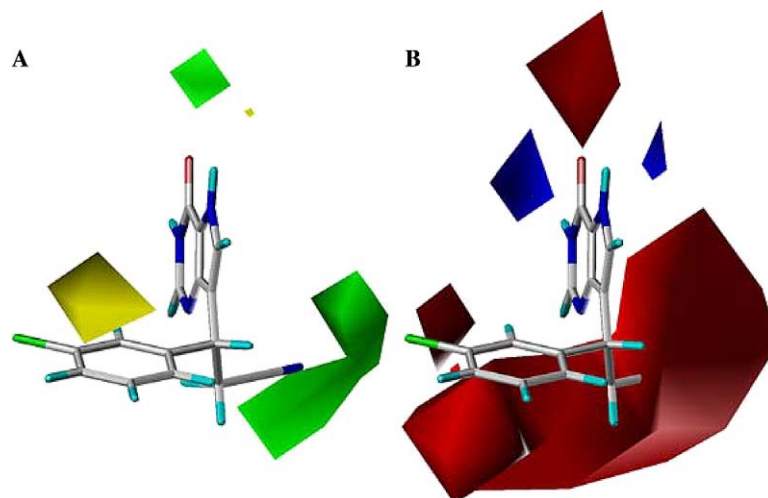
Figure 3. Representation of the 3D contour maps (SD × coefficient) for CoMFA 2. (A) steric field, (B) electrostatic field (see text for further explanation). The most potent PNP inhibitor (**28**) is displayed in the background for reference.

Table 3. Experimental and predictive activities (pIC$_{50}$) with residual values for the test set compounds for CoMFA 1, CoMFA 2, and HQSAR models

Test set compounds	Experimental[a]	Predicted CoMFA 1[a]	CoMFA 1 residuals[b]	Predicted CoMFA 2[a]	CoMFA 2 residuals[b]	Predicted HQSAR[a,c]	HQSAR residuals[b]
53	7.70	7.50	0.20	7.24	0.40	7.34	0.36
54	7.60	7.39	0.21	7.14	0.45	7.34	0.26
55	7.15	7.56	−0.38	7.72	−0.57	7.27	−0.12
56	6.57	7.05	−0.58	7.07	−0.39	7.23	−0.60
57	7.60	7.00	0.48	7.56	0.04	7.05	0.35
58	7.05	7.18	−0.13	7.09	−0.04	7.75	−0.55
59	3.70	3.89	−0.19	3.90	0.20	3.97	−0.27
60	7.46	7.57	−0.11	7.51	−0.05	7.56	−0.10
61	7.45	7.09	0.36	7.39	0.06	7.37	0.08
62	7.82	7.69	0.13	7.33	0.49	7.13	0.58
63	6.44	5.81	0.63	6.20	0.24	5.87	0.57
64	5.98	6.26	−0.28	6.28	−0.30	5.99	−0.01
65	5.57	5.54	0.03	5.64	0.10	5.56	0.01
66	7.08	7.68	−0.60	7.10	−0.11	7.26	−0.18
67	5.64	5.38	0.26	5.56	0.08	6.07	−0.40
68	6.17	6.84	−0.62	6.33	−0.16	6.07	0.10

[a] pIC$_{50}$ values.
[b] The difference between experimental and predicted values.
[c] Fragment distinction: atoms, bonds, connections, and chirality.

predicting biological activities of an external test set of 16 molecules (compounds **53**–**68**, Table 1). Prior to prediction, the test set compounds were processed identically to the training set compounds, as previously described. The external validation process can be considered the most valuable validation method, as these compounds were completely excluded during the training of the model. The results are listed in Table 3, and the graphic results simultaneously displayed in Figure 4.

The good agreement between experimental and predicted pIC$_{50}$ values for both models indicates the reliability of the constructed 3D QSAR models. The predicted values fall close to the experimental pIC$_{50}$ values, not deviating by more than 0.4 log units. The only exceptions are compounds **56**, **57**, **63**, **66**, and **68** for CoMFA 1, and compounds **54**, **55**, and **62** for CoMFA 2, for which the predicted values are more substantially in error (be-

tween 0.45 and 0.63 log units). None of the inhibitors were detected as outliers, the predictions for all compounds are within 2 standard deviations of the estimated error. The good agreement between experimental and predicted pIC$_{50}$ values for the test set compounds suggests that the models are of sufficient quality and can be used in the design of new inhibitors with improved potency.

3.3. Structure-based docking and 3D QSAR studies

The understanding of protein–ligand interactions is essential for the design of novel synthetic inhibitors when appropriate information is available. In the present study, we used an approach that combines 3D QSAR and structure-based molecular docking to investigate the detailed interactions between PNP active site and small-molecule inhibitors. The crystal structure of
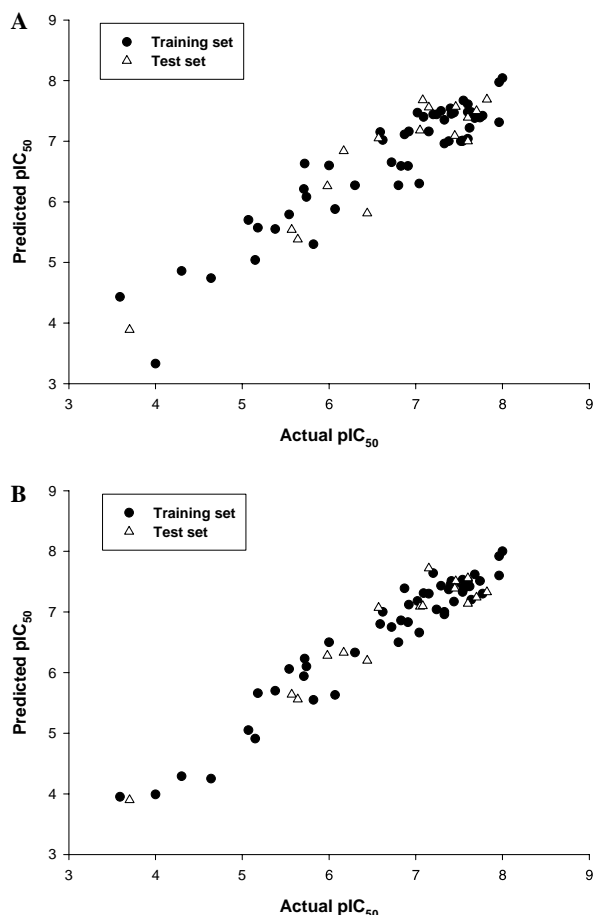
Figure 4. Plot of predicted values of pIC_{50} versus the corresponding experimental values for the training (solid circle) and test (open triangles) set compounds for CoMFA 1 (A) and CoMFA 2 (B) models.

calf spleen PNP (PDB code: 1B8O) was used for the docking studies. The bovine structure of PNP corresponds closely to that of human erythrocytes, with high identity and conservation of all amino acid residues in the catalytic region. The 68 data set inhibitors were docked into the PNP putative binding site using the

docking program GOLD 2.1, as shown in Figure 5A. The final 3D aligned molecules produced a statistically significant 3D QSAR model (CoMFA 2) with good correlation and predictive power.

Crystallographic studies of PNP-inhibitor complexes have identified Asn243 and Glu201 as the key residues in PNP catalytic mechanism. The proposed binding modes of the molecules generated with GOLD 2.1 are in good agreement with the crystallographic data. To illustrate the intermolecular interactions implicated in purine base binding, the most potent inhibitor (**28**) of the series is shown in Figure 5B for further analysis. As it can be seen, the carboxamide group of the Asn243 side chain forms a hydrogen bond with O6 of the purine ring, while the side-chain carboxylate of Glu201 is close to accepting a hydrogen bond from N1 (3.42 Å). Similar interactions are observed in other PNP crystallographic structures. The proposed binding mode also suggests additional interactions with residues Thr242 and Tyr192. In particular, N7 may hydrogen bond to Thr242 (2.49 Å), thus fulfilling the inhibitor hydrogen bonding network.

According to the proposed binding model, there is still room for bulkier groups at the 9-position of the purine ring. This observation is in good agreement with the CoMFA 2 steric contour map (Fig. 3A), which shows a large green region surrounding the cyanil moiety. The same trend could be observed near O6, which interacts with Asn243 (Fig. 5). However, the small green area in Figure 3A does not suggest any favorable structural variations in this region. The disfavored yellow contour located near the PO_4^{2-} binding site in the steric contour map suggests that bulkier groups would be in conflict with PO_4^{2-}. It is possible that electrostatic interactions with PO_4^{2-} would have some influence on activity in this situation, but the CoMFA 2 electrostatic contour map does not support this hypothesis (Fig. 3B). According to docking studies and the electrostatic contour map, the most important region, where additional positive groups would enhance activity, is located near Glu201.
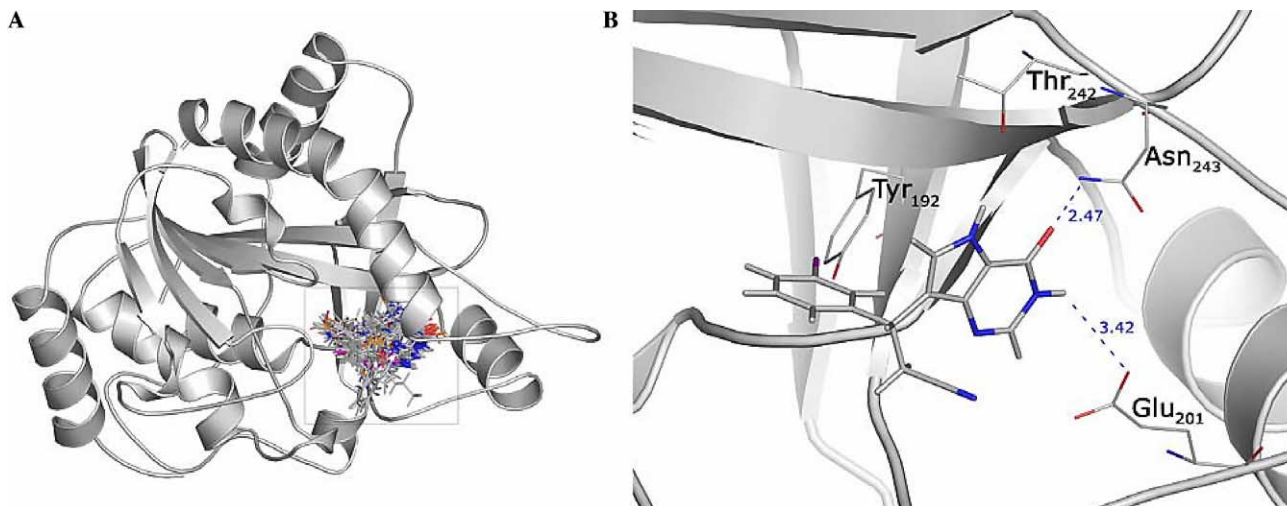

Figure 5. (A) Three-dimensional alignment of the 68 inhibitors in the active site of calf spleen PNP (alignment 2) generated by the docking program GOLD 2.1. (B) Representation of the intermolecular interactions implicated in purine base binding of inhibitor **28**.

3.4. 2D QSAR analyses

3.4.1. HQSAR models. HQSAR investigation requires selecting values for parameters that specify the length of the hologram, as well as the size and type of fragment that are to be encoded. In this study, holograms were generated using the standard parameters implemented in SYBYL 6.9.2. The generation of molecular fragments was carried out using the following fragment distinctions: atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA). Several combinations of these parameters were considered using the fragment size default (4–7), as follows: A/B, A/B/C, A/B/C/H, A/B/C/Ch, A/B/C/H/Ch, and A/C/DA. HQSAR analysis was performed over the 12 default series of hologram lengths of 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins. The statistical results from the PLS analyses for the 52 training set compounds using several fragment distinction combinations are summarized in Table 4. The best statistical results were obtained using either A/B ($q^2 = 0.67$, and $r^2 = 0.86$ with a SEE of 0.42) or A/B/C/ Ch as distinction information ($q^2 = 0.68$, and $r^2 = 0.85$ with a SEE of 0.42), with 3 components being optimum in both cases. The use of other fragment distinction into molecular hologram does not improve the quality of the model as measured by the statistical parameters shown in Table 4.

The influence of different fragment sizes in the statistical parameters was further investigated for the HQSAR model generated using A/B/C/Ch. Fragment size parameters control the minimum and maximum lengths of fragments to be included in the hologram fingerprint. These parameters can be changed to include larger or smaller fragments in the holograms. As mentioned previously, default fragments between 4 and 7 were used to derive the best HQSAR model. These default values had proved useful in a number of situations. The statistical results for the different fragment sizes evaluated (2–5, 3–6, 4–7, 5–8, 6–9, and 7–10) are summarized in Table 5. As it can be seen, the variation of the fragment size led to the generation of a better HQSAR model compared to that derived using the fragment size default (4–7). The best statistical result ($q^2 = 0.70$, and $r^2 = 0.90$ with a SEE of 0.38) among all models was ob-

Table 4. HQSAR analysis for various fragment distinction on the key statistical parameters using fragment size (4–7)

Fragment distinction	Statistical parameters				
	q^2	r^2	SEE	HL	N
A/B	0.67	0.86	0.42	71	3
A/B/C	0.65	0.84	0.46	307	3
A/B/C/H	0.59	0.72	0.60	257	3
A/B/C/Ch	0.68	0.85	0.42	97	3
A/B/C/H/Ch	0.57	0.70	0.62	257	3
A/C/DA	0.59	0.76	0.56	61	3

q^2, cross-validated correlation coefficient; r^2, non-cross-validated correlation coefficient; SEE, non-cross-validated standard error; HL, hologram length; N, optimal number of components. Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor.

Table 5. HQSAR analysis for the influence of various fragment sizes on the key statistical parameters using the best fragment distinction (A/B/C/Ch)

Fragment size	Statistical parameters				
	q^2	r^2	SEE	HL	N
2–5	0.66	0.80	0.50	151	3
3–6	0.66	0.85	0.44	151	3
4–7	0.68	0.84	0.46	97	3
5–8	0.66	0.85	0.45	401	3
6–9	0.70	0.90	0.38	401	3
7–10	0.64	0.79	0.52	401	3

tained with the fragment size 6–9, with 3 being the optimum number of PLS components. These results confirm the importance of testing the influence of different fragment sizes in the resultant molecular hologram. Moreover, the data in Tables 4 and 5 show that there is no correlation between predictive power of the HQSAR models and hologram length, which may be expected as the pattern of bin occupancies in the molecular hologram will change in a non-trivial manner with respect to changes in hologram length.

In terms of validation of a QSAR model, a measure of internal consistency is available in the form of q^2. However, the most important test of a QSAR model is its ability to predict the property value for new compounds. As the structure encoded in a 2D fingerprint is directly related to the biological activity of series of molecules, HQSAR models should be able to predict the activity of new structurally related PNP inhibitors from its fingerprint. Thus, the predictive power of the best HQSAR model derived using the 52 training set molecules (fragment distinction: A/B/C/Ch; fragment size: 6–9, Table 5) was assessed by predicting pIC_{50} values for 16 test set molecules (compounds **53**–**68**, Table 1), which were completely excluded during the training of the model. For simplification, the results are also listed in Table 3. The graphic results for the experimental versus predicted pIC_{50} values of both training and test sets are displayed in Figure 6. The results show that the test set
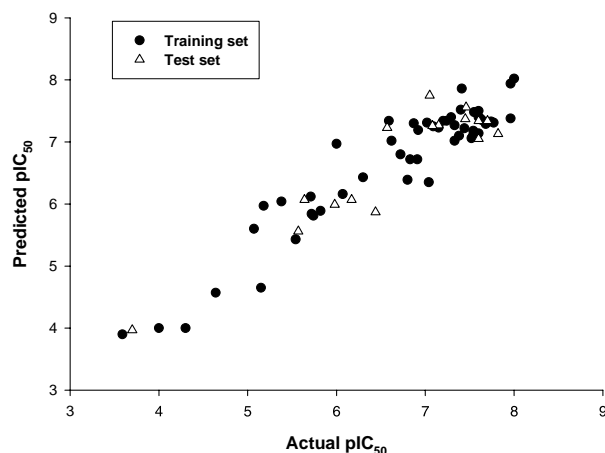

Figure 6. Plot of predicted versus experimental values of pIC_{50} for the 68 PNP inhibitors (training and test set compounds are discriminated).
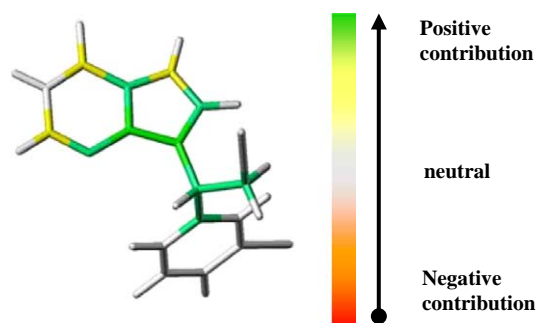
Figure 7. HQSAR contribution map for compound **28**.

compounds, which represent the different structural properties incorporated within the training set, are well predicted without any outliers (residual values are less than 2 standard deviations of the estimated error). The predicted values fall close to the experimental pIC50 values, deviating by no more than 0.4 log units. The only exceptions are compounds **56**, **58**, **62**, and **63**, for which the predicted values are more in error (between 0.55 and 0.60 log units), respectively. The good agreement between experimental and predicted pIC_{50} values for the test set compounds indicates the robustness of the HQSAR model (Fig. 6). Besides predicting the potency of untested inhibitors, an HQSAR analysis provides important hints about what molecular fragments are directly related to biological activity.

This information, combined with knowledge of synthetic chemistry, could lead to the synthesis of new molecules with improved properties. To better visualize this principle, the most important fragments of the compound **28** (the most potent inhibitor of the data set) are shown in Figure 7. The HQSAR module implemented in SYBYL 6.9.2 uses a color code to discriminate the main atomic contributions to activity. The colors at the red end of the spectrum (red, red orange, and orange) reflect poor contributions, whereas colors at the green end (yellow, green blue, and green) reflect favorable contributions. Atoms with intermediate contributions are colored white. As it can be seen in Figure 7, the molecular fragment represented by the purine ring of **28** was found to be strongly related to the biological activity of this compound. The same can be suggested for the cyanil moiety present in the group attached at the 9-position of the purine ring. It is important to note that the CoMFA steric counter map also reveals this region as potential target for molecular modification and further SAR studies aimed at increasing selectivity and potency. Moreover, in any molecule of the data set, regions with intermediate or poor contributions can be identified as potential targets for synthesis and SAR studies.

4. Conclusions

The 3D and 2D QSAR models described herein show both good internal and external consistency. The good correlation between experimental and predicted pIC_{50} values for the test set compounds further proved the reliability of the constructed QSAR models (CoMFA 1, CoMFA 2, and HQSAR). It is worth noting that we have employed the same training and test sets for all QSAR analyses, and the results showed that investigations can be carried out concomitantly to search for synergies between 3D and 2D QSAR technologies. Moreover, the 3D QSAR models generated are compatible with the 3D protein environment in the PNP binding site. These results suggest that the 3D model of PNP can be used in structure-based drug design and the QSAR models should be useful for the design of novel structurally related PNP inhibitors.

References and notes

1. Mao, C.; Cook, W. J.; Zhou, M.; Federov, A. A.; Almo, S. C.; Ealick, S. E. *Biochemistry* **1998**, *37*, 7135.
2. Beauchamp, L. M.; Tuttle, J. V.; Rodriguez, M. E.; Sznaidman, M. L. *J. Med. Chem.* **1996**, *39*, 949.
3. Kicska, G. A.; Tyler, P. C.; Evans, G. B.; Furneaux, R. H.; Kim, K.; Schramm, V. L. *J. Biol. Chem.* **2002**, *277*, 3219.
4. Bzowska, A.; Kulikowska, E.; Shugar, D. *Pharmacol. Ther.* **2000**, *88*, 349.
5. Kelley, J. L.; McLean, E. W.; Crouch, R. C.; Averett, D. R.; Tuttle, J. V. *J. Med. Chem.* **1995**, *38*, 1005.
6. Montgomery, J. A. *Med. Chem. Rev.* **1993**, *13*, 209.
7. Bantia, S.; Kilpatrick, J. M. *Curr. Opin. Drug Discov. Dev.* **2004**, *7*, 243.
8. BCX-4208, a second generation PNP inhibitor, has entered Phase I trial as a drug candidate for the treatment of T-cell mediated autoimmune diseases. Further information is available in the ByoCryst Pharmaceuticals website at www.biocryst.com.
9. Ragno, R.; Artico, M.; De Martino, G.; La Regina, G.; Coluccia, A.; Di Pasquali, A.; Silvestri, R. *J. Med. Chem.* **2005**, *48*, 213.
10. Murias, M.; Handler, N.; Erker, T.; Pleban, K.; Ecker, G.; Saiko, P.; Szekeres, T.; Jager, W. *Bioorg. Med. Chem.* **2004**, *12*, 5571.
11. Purushottamachar, P.; Kulkarni, V. M. *Bioorg. Med. Chem.* **2003**, *11*, 3487.
12. Farutin, V.; Masterson, L.; Andricopulo, A. D.; Cheng, J.; Riley, B.; Hakimi, R.; Frazer, J. W.; Cordes, E. H. *J. Med. Chem.* **2000**, *42*, 2422.
13. Andricopulo, A. D.; Yunes, R. A. *Chem. Pharm. Bull.* **2001**, *49*, 10–17.
14. Jayatilleke, P. R. N.; Nair, A. C.; Zauhar, R.; Welsh, W. J. *J. Med. Chem.* **2000**, *43*, 4446.
15. Söderholm, A. A.; Lehtovuori, P. T.; Nyrönen, T. H. *J. Med. Chem.* **2005**, *48*, 917.
16. Avery, M. A.; Alvim-Gaston, M.; Rodrigues, C. R.; Barreiro, E. J.; Cohen, F. E.; Sabnis, Y. A.; Woolfrey, J. R. *J. Med. Chem.* **2002**, *45*, 292.
17. Montgomery, J. A.; Niwas, S.; Rose, J. D.; Secrist, J. A., III; Babu, S.; Bugg, C. E.; Erion, M. D.; Guida, W. C.; Ealick, S. E. *J. Med. Chem* **1993**, *36*, 55.
18. Secrist, J. D., III; Niwas, S.; Rose, J. D.; Babu, S.; Bugg, C. E.; Erion, M. D.; Guida, W. C.; Ealick, S. E.; Montgomery, J. A. *J. Med. Chem.* **1993**, *36*, 1847.

M. S. Castilho et al. / Bioorg. Med. Chem. 14 (2006) 516–527

527

19. Guida, W. C.; Elliott, R. D.; Thomas, H. J.; Secrist, J. A., III; Babu, S.; Bugg, C. E.; Erion, M. D.; Ealick, S. E.; Montgomery, J. A. *J. Med. Chem.* **1994**, *37*, 1114.
20. Niwas, S.; Cahnd, P.; Pathak, V. P.; Montgomery, J. A. *J. Med. Chem.* **1994**, *37*, 2477.
21. Halazy, S.; Eharhard, A.; Eggenspiller, A.; Berges-Gross, V.; Danzin, C. *Tetrahedron* **1996**, *52*, 177.
22. Anu, J. T.; Tommi, H. N.; Toni, R.; Antti, P. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 807.
23. Masamoto, A.; Kiyoshi, H.; Kimito, F. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1390.
24. Nicholas, E. J.; David, B. T.; Peter, W.; Graham, J. S. *J. Mol. Graph. Model.* **2001**, *20*, 111.
25. Lemmen, C.; Lengauer, T. *J. Comput. Aided Mol. Des.* **2000**, *14*, 215.
26. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. D. *J. Mol. Biol.* **1997**, *267*, 727.
27. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins* **2003**, *52*, 609.
28. Zhu, W. L.; Gang, C.; Hu, L. H.; Luo, X. M.; Gui, C. S.; Cheng, L.; Puah, C. M.; Chen, K. X.; Jiang, H. L. *Bioorg. Med. Chem.* **2005**, *13*, 313.