Multivariate Quantitative Structure–Pharmacokinetic Relationships (QSPKR) Analysis of Adenosine A_1 Receptor Agonists in Rat

PIETER H. VAN DER GRAAF,[†] JONAS NILSSON,[‡,§] ERNO A. VAN SCHAICK,[†,$] AND MEINDERT DANHOF*[,†]

Contribution from *Division of Pharmacology, Leiden/Amsterdam Center for Drug Research, P.O. Box 9503, 2300RA Leiden, The Netherlands, and Department of Medicinal Chemistry, University Centre for Pharmacy, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands.*

Received July 21, 1998. Final revised manuscript received October 21, 1998.
Accepted for publication October 26, 1998.

Abstract □ The aim of this study was to investigate the feasibility of a quantitative structure–pharmacokinetic relationships (QSPKR) method based on contemporary three-dimensional (3D) molecular characterization and multivariate statistical analysis. For this purpose, the programs SYBYL/CoMFA, GRID, and Pallas, in combination with the multivariate statistical technique principal component analysis were employed to generate a total of 16 descriptor variables for a series of 12 structurally related adenosine A_1 receptor agonists. Subsequently, the multivariate regression method, partial least squares, was used to predict clearance (CL), volume of distribution (Vd_{SS}) and protein binding (fraction unbound, f_U). The QSPKR models obtained could account for most of the variation in *CL*, Vd_{SS}, and f_U ($R^2 = 0.82$, 0.61 and 0.78, respectively). Cross-validation confirmed the predictive ability of the models ($Q^2 = 0.59$, 0.41 and 0.62 for *CL*, Vd_{SS}, and f_U, respectively). In conclusion, we have developed a multivariate 3D QSPKR model that could adequately predict overall pharmacokinetic behavior of adenosine A_1 receptor agonists in rat. This methodology can also be used for other classes of compounds and may facilitate the further integration of QSPKR in drug discovery and preclinical development.

Introduction

Since the pioneering work of Hansch and co-workers,[1] quantitative structure−activity relationships (QSAR) analysis has become a widely used tool in the design of modern drugs. In recent years, sophisticated alternatives for traditional "Hansch-type" QSAR methods have been developed, in particular in the area of three-dimensional (3D) molecular characterization and multivariate data analysis.[2−4] In contrast to these developments in the QSAR field, the vast majority of quantitative structure−pharmacokinetic relationships (QSPKR[5]) studies have only focused on univariate correlations of individual pharmacokinetic parameters with lipophilicity and ionization, and it has been well established that log *P* and pK_A are important determinants of drug absorption, distribution, protein binding and elimination (for example[3,6−10]). A major drawback of such QSPKR models is that they provide no insight into the influence of other physicochemical properties and therefore their conceptual and predictive value is rather limited. However, despite these limitations, only a few attempts have been made to develop more comprehensive multivariate QSPKR models that predict pharmacokinetics on the basis of various molecular physiocochemical descriptors.[11,12] Furthermore, most QSPKR studies have primarily focused on the prediction of single parameters instead of multivariate modeling of overall pharmacokinetic behavior.[6] Therefore, in the present study we have started to explore the possibilities of generating QSPKR models based on contemporary QSAR methods of 3D molecular characterization and multivariate statistical analysis. The 3D molecular descriptor methods, SYBYL/CoMFA[13] and GRID,[14] in combination with the multivariate statistical techniques, principal component analysis (PCA) and partial least squares (PLS[15,16]), were used to build models for simultaneous prediction of the primary pharmacokinetic parameters, clearance (*CL*), volume of distribution at steady-state (Vd_{SS}), and protein binding (fraction unbound, f_U) for a series of structurally related adenosine A_1 receptor agonists in rat. The pharmacokinetic data were obtained during the course of a program focused on the design of partial agonists for the adenosine A_1 receptor[17,18] as reported previously.[19−23]

Materials and Methods

Chemicals—All compounds used in this study are analogues of the endogenous purine nucleoside, adenosine (Figure 1). Compound **1** is the reference adenosine A_1 receptor agonist, N^6-cyclopentyladenosine (CPA[19]). Compounds **2**−**4** and **5**−**9** are deoxyribose[20] and 8-alkylamino-substitued CPA analogues,[22] respectively. In addition to the CPA analogues, another widely used adenosine A_1 receptor agonist, R−N^6-phenylisopropyladenosine (R−PIA,[21] **10**) was included as well as the novel compounds **11** and **12** as hydrophilic and lipophilic non-CPA analogues, respectively.[23]

Pharmacokinetic Experiments—Details of the pharmacokinetic experiments have been published previously.[19−23] Briefly, 2 days before experimentation, the right femoral artery and the right jugular vein of male Wistar rat (200−250 g) were cannulated for the collection of blood samples and administration of drugs, respectively. Conscious, freely moving rats received an intravenous infusion of vehicle (20% DMSO/water) or compound over 15 min. Serial arterial blood samples were taken over a period of at least 100 min, hemolyzed immediately, and stored at −35 °C until HPLC analysis of blood concentrations.

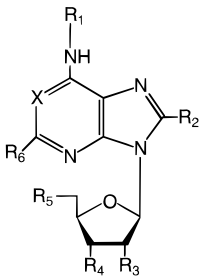
Compartmental analysis of blood concentration−time profiles was performed by fitting the data to a biexponential equation from which systemic clearance (*CL*) and volume of distribution at steady state (Vd_{SS}) were calculated.[19] The fraction unbound drug in blood (f_U) was determined using standard ultrafiltration methods. The pharmacokinetic parameters for all 12 compounds included in the study are summarized in Table 1.

Molecular Modeling—The 3D structures of the compounds were provided by Dr A. P. IJzerman (Division of Medicinal Chemistry, Leiden/Amsterdam Center for Drug Research), based on the adenosine A_1 receptor model published before.[24]

Generation of Molecular Descriptors—In both the SYBYL/CoMFA[13] (Molecular Modeling Software 6.3, Tripos Inc., St. Louis,

* Corresponding author. Telephone: ++-31-71-5276211. Fax: ++-31-71-5276292. E-mail: m.danhof@lacdr.leidenuniv.nl.
[†] Division of Pharmacology.
[‡] Department of Medicinal Chemistry.
[§] Current affiliation: Pharmacia & Upjohn, Department of Structural Chemistry, Uppsala, Sweden.
[$] Current affiliation: Janssen Research Foundation, Department of Clinical Pharmacokinetics, Beerse, Belgium.

10.1021/js980294a CCC: $18.00
Published on Web 01/12/1999



	R_1	R_2	R_3	R_4	R_5	R_6	X
1	cyclopentyl	H	OH	OH	OH	H	N
2			H				
3				H			
4					H		
5		$NH\text{-}CH_3$					
6		$NH\text{-}CH_2CH_3$					
7		$NH\text{-}(CH_2)_2CH_3$					
8		$NH\text{-}(CH_2)_3CH_3$					
9		NH-cyclopentyl					
10	phenylisopropyl						
11	(p-sulfophenyl)						
12						Cl	C

Figure 1—Chemical structure of the adenosine A_1 receptor agonists. For the sake of clarity, only the modifications compared with the reference compound **1** (CPA) are indicated.

Table 1—Pharmacokinetic Parameters for the Adenosine A_1 Receptor Agonists[a]

compound[a]	CL (mL/min/kg)[b]	Vd_{SS} (mL/kg)[c]	f_U (%)[d]
1	76	320	72
2	33	1050	63
3	58	660	68
4	55	740	61
5	65	1000	24
6	81	860	16
7	92	1000	23
8	72	1190	15
9	62	1130	11
10	24	940	41
11	5.6	365	67
12	62	1155	17

[a] Compound numbers correspond to the chemical strucutres shown in Figure 1. [b] Clearance. [c] Volume of distribution at steady state. [d] Fraction unbound.

MO) and GRID[14] (Molecular Discovery Ltd., University of Oxford, U.K.) programs, a grid large enough to enclose all the aligned ligands is utilized. In each grid point, interactions between a probe atom and the target molecules are calculated. SYBYL/CoMFA and GRID use different force fields and different types of probe atoms, and the interactions are calculated differently. Interactions accounted for in the GRID force field are steric, electrostatic and hydrogen bonding interactions represented by the Lennard−Jones energy (E_{ste}), the Coulombic energy (E_{ele}) and a hydrogen bonding (E_{hb}) term, respectively. In contrast to SYBYL/CoMFA, where the interaction energies (i.e., E_{ste} and E_{ele}) are considered separately, the sum of all the different interaction energies is calculated in each grid point with GRID. An attractive interaction between the probe atom and the ligand produces a negative field (E_{tot}), whereas a repulsive interaction is positive

$$E_{tot} = E_{ele} + E_{ste} + E_{hb} \quad (1)$$

Different probes reflect different types of interactions and may selectively be included to mimic specific interactions between the ligand and the target protein,[14,25,26] and often more than one probe is necessary for a complete description of the interactions involved in the ligand−protein interaction.

The second type of descriptors considered were molecular surface volumes, such as, POP1, POT0, and POM1 (see Table 2), corresponding to the volumes (Å^3) of the electrostatic potential surfaces at the charge levels plus one, zero, and minus one, respectively (SYBYL 6.1, Molecular Modeling Software 6.3, Tripos Inc., St. Louis, MO).

The third type of descriptors were the electronic descriptors [heat of formation (HEFO), electronic energy (ELEN), dipole moment (DIPO), filled levels (FILE), and ionization potentials (IOPO)], obtained from Mopac 6.0 (SYBYL 6.1) AM1 single point calculations[27] with the keywords SCF1, MULLIK, AM1 and T = 3600 activated (Table 2). Four steric descriptors were included in the present investigation, namely the core−core repulsion (COCO), molecular weight (MW) and the areas and the volumes enclosed by the Connolly surfaces, COAR and COVO, respectively. The Connolly surface area (solvent accessible area) is defined as the area a theoretical water molecule (1.4 Å in diameter) produces when it moves over the van der Waals surface of a ligand.

Finally, the log *P* descriptor (the logarithm of the partition coefficient, *P*, between 1-octanol and water) was estimated with the Prolog *P* module (CDR database) of the Pallas program (version 1.2, CompuDrug Chemistry Ltd., Budapest, Hungary), which utilizes a modification of the method presented by Rekker and De Kort,[28] where the contributions of the hydrophobic molecular fragment constants are added. It is well-known that different methods of calculation may yield different theoretical estimates of log *P*. Therefore, we also calculated log *P* with the ATOMIC5 database of the Pallas program, which is based on a modification of the work of Ghose and Crippen.[29] Although the estimates varied slightly between the two methods, the correlation was high (r = 0.88) and the slope and intercept of the best-line fit were not significantly different from unity and zero, respectively. Hence, for the sake of simplicity only estimates based on the CDR database were used in the present analysis, and the log *P* parameter should be interpreted as such. In other cases, however, it might be advantageous to include several log *P* estimates calculated by different methods in the model.

Principal Component Analysis (PCA)—Principal component analysis (PCA) was used in the first stage of the study to replace the original molecular descriptors generated in the GRID program by so-called "3D principal properties" (3DPPs, see *Results*) and to summarize in a graphical manner the information contained in the dataset. Details of PCA can be found in numerous references.[15,16] Briefly, let **X** be a matrix with *m* rows and *n* columns, representing a dataset of *m* compounds with *n* descriptor variables ($\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}]$). In PCA, the original *n* descriptors in **X** are replaced by a limited number (*a*) of new variables, called principal components (PCs), which are linear combinations of the columns in **X**. Algebraically, PCA decomposes **X** in *a* PCs as follows

$$\mathbf{X} = \mathbf{t_1}\,\mathbf{p_1}^T + \mathbf{t_2}\,\mathbf{p_2}^T + ... + \mathbf{t_a}\,\mathbf{p_a}^T + \mathbf{E} \quad (2)$$

where the $\mathbf{t_i}$ and $\mathbf{p_i}$ vectors are known as the principal components scores and variable loadings, respectively, and **E** is the residual matrix not described by the model. Each consecutive PC is calculated orthogonal to all previous PCs and accounts for a decreasing percentage of the variation in **X**. The purpose of PCA is to describe the complete dataset with less PCs than original descriptors without significant loss of information. Plots of scores and loadings obtained from the first few PCs can then be made to reveal the relationships between objects (compounds) and variables (descriptors), respectively.

Partial Least Squares (PLS) Regression—Partial least squares (PLS) is a relatively new multivariate statistical method that has become the most widely used regression tool in the area of QSAR.[16] PLS is a generalization of ordinary multiple linear regression (MLR) and can be seen as a least-squares regression extension of PCA. Like MLR, PLS aims to provide a statistical model that describes biological properties (**y**) in terms of the descriptor variables in matrix **X**. In contrast to MLR, PLS can deal with high correlations between the descriptor variables in **X** (collinearity) and with the situation where the number of descriptors exceeds the number of compounds ($n > m$), which is often the case in QSAR/QSPKR). A detailed description of the PLS algorithm can be found in the review by Geladi and Kowalski.[15]

For the interpretation of the model, it is particularly useful that PLS models can be expressed in terms of regression coefficients ($\mathbf{b_{PLS}}$):

Table 2—Physicochemical Descriptors[b] Used for the QSPKR Analysis of the Adenosine A_1 Receptor Agonists[a]

compound	1 3DPP1	2 3DPP2	3 3DPP3	4 POP1	5 POT0	6 POM1	7 HEFO	8 ELEN	9 DIPO	10 FILE	11 IOPO	12 COAR	13 COVO	14 MW	15 COCO	16 log *P*
1	109.0	–40.2	14.9	1120	12 116	18 426	–27.3	–32 308	4.21	65	8.63	322.4	322.2	335.4	27 785	0.12
2	105.1	–30.2	8.5	1094	13 202	18 272	10.2	–29 569	3.36	62	8.66	317.1	313.2	319.4	25 367	0.64
3	100.9	–38.1	16.8	1232	9 511	17 683	12.4	–29 811	5.13	62	8.56	316.0	313.4	319.4	25 609	0.64
4	96.2	–33.8	5.6	1108	11 548	17 689	15.5	–29 995	3.48	62	8.65	313.9	312.0	319.4	25 793	1.22
5	–117.1	5.4	–5.1	1461	10 162	17 713	148.3	–38 336	6.96	71	7.64	334.1	337.4	364.4	33 444	0.92
6	–130.9	4.0	–5.4	1512	11 248	18 820	142.9	–40 810	7.12	74	7.63	348.2	357.0	378.4	35 763	1.44
7	–154.7	–5.5	5.6	1545	11 366	18 828	140.8	–43 295	7.08	77	7.62	363.0	374.7	392.5	38 092	1.96
8	–156.4	–3.4	–1.0	1583	12 177	19 393	185.0	–46 016	7.03	80	7.62	374.1	389.1	406.5	40 659	2.48
9	–158.2	–1.7	–1.7	1619	12 210	19 412	160.0	–48 099	7.00	82	7.62	379.5	398.9	418.5	42 613	2.63
10	93.7	131.7	63.8	1335	14 004	19 957	25.5	–38 180	4.17	74	8.63	374.2	376.0	385.4	33 119	1.11
11	112.1	51.9	–116.7	1374	12 110	18 324	–99.3	–40 984	8.70	77	8.85	364.2	361.9	423.4	35 234	–2.23
12	100.3	–40.1	14.7	1178	11 994	18 766	–41.0	–34 615	3.71	78	8.59	340.6	342.2	368.8	29 798	1.67

[a] Compound numbers refer to the chemical structures shown in Figure 1. [b] Abbreviations: 3DPP, three-dimensional principal property; COAR, area enclosed by Connolly surface; COCO, core–core repulsion; COVO, volume enclosed by Connolly surface; DIPO, dipole moment; ELEN, electronic energy; FILE, filled levels; HEFO, heat of formation; IOPO, ionization potential; log *P*, logarithm of partition coefficient; MW, molecular weight; POM1, volume of electrostatic potential surface at charge level minus one; POP1 volume of electrostatic potential surface at charge level plus one; and POT0, volume of electrostatic potential surface at charge level zero.

$$\mathbf{y} = \mathbf{X}\mathbf{b}_{\mathbf{PLS}} + \mathbf{F} \tag{3}$$

where **F** is the residual matrix. As in the case of MLR, the regression coefficients can be used to determine the influence of each variable in **X** in the model.[30]

A crucial step in the development of PLS models is the determination of the number of significant components.[4,15] Although it is possible to calculate as many PLS components as the number of descriptors in the **X** matrix, the use of too many components will result in a model that fits the data well but has poor predictability. Cross-validation has become the method of choice to determine the optimal number of components in PLS[25,26,30] and was also used in the present study. Briefly, with cross-validation, a PLS model is developed with a group of compounds omitted from the dataset. Subsequently, the model is used to predict the dependent variable (y_{pred}) for the omitted compounds and the differences between actual (y_{obs}) and predicted values are calculated. This procedure is repeated several times until all compounds have been omitted once. The predictive ability of the model can then be quantified with the Q^2 statistic:[25,26,30]

$$Q^2 = 1 - \sum (y_{ipred} - y_{iobs})^2 / \sum (y_{iobs} - y_{mean})^2 \tag{4}$$

The optimum number of PLS components is given by the number of components that maximizes Q^2. A model with good predictive capability will have a Q^2 close to unity, whereas a negative value for Q^2 indicates that the model does not predict better than random. Formally, real predictability can only be validated with an external test set, which is not included in the calibration process of the model. However, even in QSAR studies in which it is usually easier to test a large number of compounds than in pharmacokinetic experiments, it is common to test predictability by cross-validation because the dataset is often too small to be split into a training and test set. A compromise between an external test set and leave-one-out cross-validation as used in the present study is leave-more-out cross-validation where multiple compounds are left out simultaneously several times, like in bootstrapping. However, this method still requires rather large datasets and we therefore decided to use the leave-one-out method; that is, compounds were omitted one by one.

The Q^2 value can be compared with the R^2 statistic which indicates the fraction of the variation accounted for by the model:[25,26,30]

$$R^2 = 1 - \sum (y_{icalc} - y_{iobs})^2 / \sum (y_{iobs} - y_{mean})^2 \tag{5}$$

The R^2 values range between zero (meaningless model) and unity (perfect correlation). The value of R^2 increases with the number of components included in the model, and the PLS regression model converges toward the MLR solution when the number of components is identical to the number of variables in the **X** matrix. In that case, the fraction explained variance is maximal but the model is usually overfitted, that is, the predictability is poor (low Q^2). On the other hand, an underfitted model with too few PLS components does not account for sufficient variation (low R^2). Eriksson and Johansson[30] have suggested that a difference between R^2 and Q^2 values of >0.3 is indicative of an inappropriate model.

The outcomes of PCA and PLS are dependent on the scaling of the variables. In the present study we always employed the autoscaling procedure, that is, all variables were scaled to unit variance and centered around the mean.[15]

All PLS and PCA calculations were performed with the PLS_Toolbox version 1.5.2 (Eigenvector Research, Inc., Manson WA) using the MATLAB software package version 4.2c.1 (The MathWorks Inc., Natick MA).

Computer Hardware—The generation of the 3D molecular descriptors was carried out on a R4600 Indy Silicon Graphics workstation. All other calculations were done on standard Pentium personal computers.

Results

Generation and PCA of Molecular Descriptors—The grid created in GRID enclosed all the aligned ligands with 4 Å with a resolution of 1 Å in all directions and consisted of 9025 points. For each of the 12 compounds, interactions with three probes (CA2+, C3, and OH2) were calculated at each grid point, thus yielding three grids with 9025 data points for each compound (i.e., a total of 27 075 data points per compound). Prior to the multivariate analysis, the grids were unfolded to row vectors, which were subsequently combined to form a 12 × 27 075 matrix. A PCA was performed on this matrix to reduce this large set of data into a smaller number of new, orthogonal (uncorrelated) variables. The first three PCs explained 91% of the total variance in the data, indicating that the original 27 075 GRID variables could be replaced by the score vectors associated with these PCs without significant loss of information. These scores vectors can be regarded as "principal properties" (PP) of the compounds and, because they were derived from a 3D structure analysis, will be referred to as "3DPPs". The 3DPPs for the first, second, and third PC (3DPP1, 3DPP2, and 3DPP3, respectively) are given in Table 2 together with the other descriptors, which were obtained as described in the *Methods* section.

Subsequently, a PCA was performed on the combined data of Table 2 to obtain insight into the relationships between descriptors and compounds. One-, two-, and three-component models described 60, 79, and 94% of the total variance, respectively. The plot of the scores of the compounds for the first two PCs reveals three clusters: the first one with the reference ligand CPA (**1**), the deoxyribose

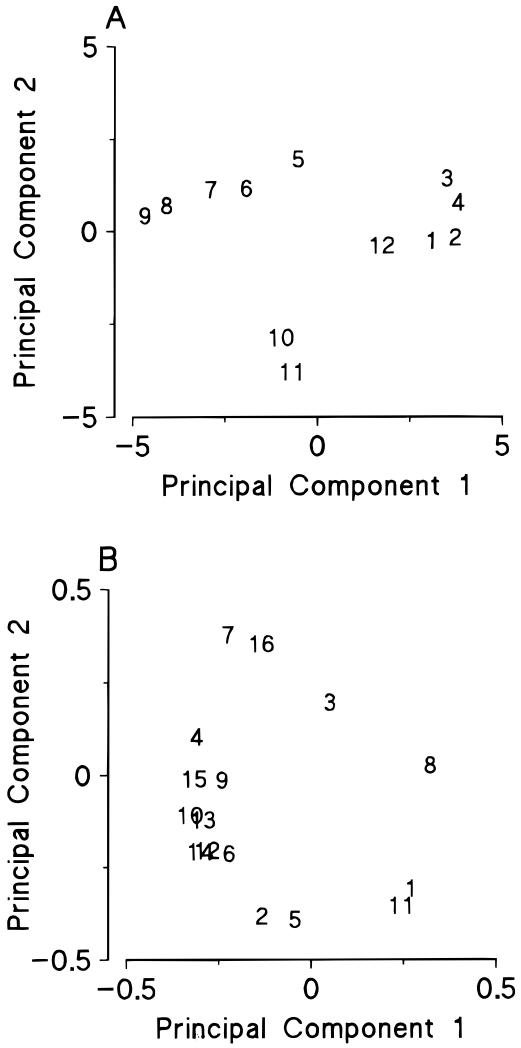


Figure 2—Graphical representation of the first two principal components, which explained 79% of the total variance of the data given in Table 2. (A) Plot of the scores of the 12 adenosine A_1 receptor agonists. Numbers correspond to the compounds shown in Figure 1. (B) Plot of the loadings of the 16 molecular descriptors. Numbers correspond to the descriptor variables in Table 2.

analogues **2**, **3**, and **4**, and the lipophilic non-CPA analogue **12**; a second cluster with the 8-alkylamino-substitued CPA analogues **5**, **6**, **7**, **8**, and **9**; and a third cluster with R−PIA (**10**) and the hydrophilic non-CPA analogue **11** (Figure 2A). From the loading plot for the first two PCs, one clear cluster can be recognized that contains the steric descriptors 12 (COAR), 13 (COVO), 14 (MW), and 15 (COCO), the electronic descriptors 9 (DIPO) and 10 (FILE), and the molecular surface volumes 4 (POP1) and 6 (POM1, Figure 2B). The clustering of the remaining descriptors is less clear. It is of interest to note, however, that 3DPP1 and 3DPP2 seem to be closely associated with IOPO and POT0, respectively (Figure 2B).

Correlation between Pharmacokinetic Parameters and Molecular Descriptors—Correlation coefficients were calculated between the pharmacokinetic parameters and each of the molecular descriptors using the MATLAB program. All three pharmacokinetic parameters were significantly correlated ($P < 0.05$) with HEFO ($r = 0.63$, 0.63, and −0.75 for CL, Vd_{SS}, and f_U, respectively) and log P ($r = 0.66$, 0.80 and −0.74 for CL, Vd_{SS} and f_U, respectively). In the case of Vd_{SS}, no other significant correlations were found, whereas CL displayed additional significant correlations with 3DPP1 ($r = -0.62$) and IOPO ($r = -0.68$), and f_U had significant correlations with 3DPP1 ($r = 0.81$), IOPO ($r = 0.81$), POP1 ($r = -0.75$), ELEN ($r = 0.73$), FILE ($r = -0.68$), COAR ($r = -0.61$), COVO ($r = -0.69$), MW ($r = -0.59$), and COCO ($r = -0.74$).

Table 3—Summary of the Results of the Prediction of *CL*, Vd_{SS}, and f_U of the Adenosine A_1 Receptor Agonists (Table 1) using PLS

pharmacokinetic parameter	no. of components[a]	Q^2	R^2
CL	4	0.59	0.82
Vd_{SS}	2	0.41	0.61
f_U	2	0.62	0.78

[a] Number of PLS components were determined using leave-one-out cross-validation (see *Methods* for details).

PLS Regression—PLS regression models to predict CL, Vd_{SS}, and f_U were built using the 16 descriptor variables listed in Table 2. The results of the analysis are summarized in Table 3 and plots of the observed versus predicted values are shown in Figure 3. With the exception of Vd_{SS} for **1** and f_U for **12**, most of the variation in the pharmacokinetic parameters could be accounted for by the PLS models (R^2 values 0.6−0.8). Furthermore, the moderate differences between R^2 and Q^2 (∼0.2 for each model) found using cross-validation indicate adequate model predictability and number of PLS components (Table 3). The relationship between R^2 and Q^2 with increasing number of PLS components is exemplified for CL in Figure 4.

Figure 5 shows the weighted PLS regression coefficients, which provide insight into the contribution of each descriptor to the modeling of the pharmacokinetic parameters. For example, 3DPP2 (descriptor 2) has a high impact on CL but is relatively unimportant in the models for Vd_{SS} and f_U. On the other hand, lipophilicity has considerable, but opposite, influence on both Vd_{SS} and f_U; that is, the log P regression coefficient is positive for the former and negative for the latter.

Discussion

To date, QSPKR studies have mainly focused on the relationship between pharmacokinetic properties and single, conventional physiocochemical parameters. To our knowledge, this study is the first example of the application of the molecular descriptor techniques, SYBYL/CoMFA and GRID, in combination with multivariate analysis methods, PCA and PLS, for the prediction of pharmacokinetic properties. Our multivariate approach has several advantages compared with the conventional methodologies employed in previously published QSPKR studies. First, the use of a whole series of advanced molecular descriptors instead of a single "classical" physicochemical property may yield more comprehensive 3D QSPKR models that can provide better insight into the relationship between chemical structure and pharmacokinetic behavior and a better and more robust prediction. Second, our approach can be used to predict multiple parameters and thus overall pharmacokinetic profile. Third, this approach could allow for integrated QSPKR−QSAR analysis and thus for simultaneous optimization of pharmacodynamic and pharmacokinetic properties at a very early stage of drug discovery. This integration could become a feasible strategy in the light of the emerging in vivo high-throughput screening technologies[31−33] that could provide considerable savings of laboratory animals, time, and money.

In the current study, QSPKR models were obtained for the prediction of three main pharmacokinetic parameters (CL, Vd_{SS}, and f_U) of a series of 12 adenosine A_1 receptor agonists. It should be noted that the present study was primarily focused on the development of methodology and

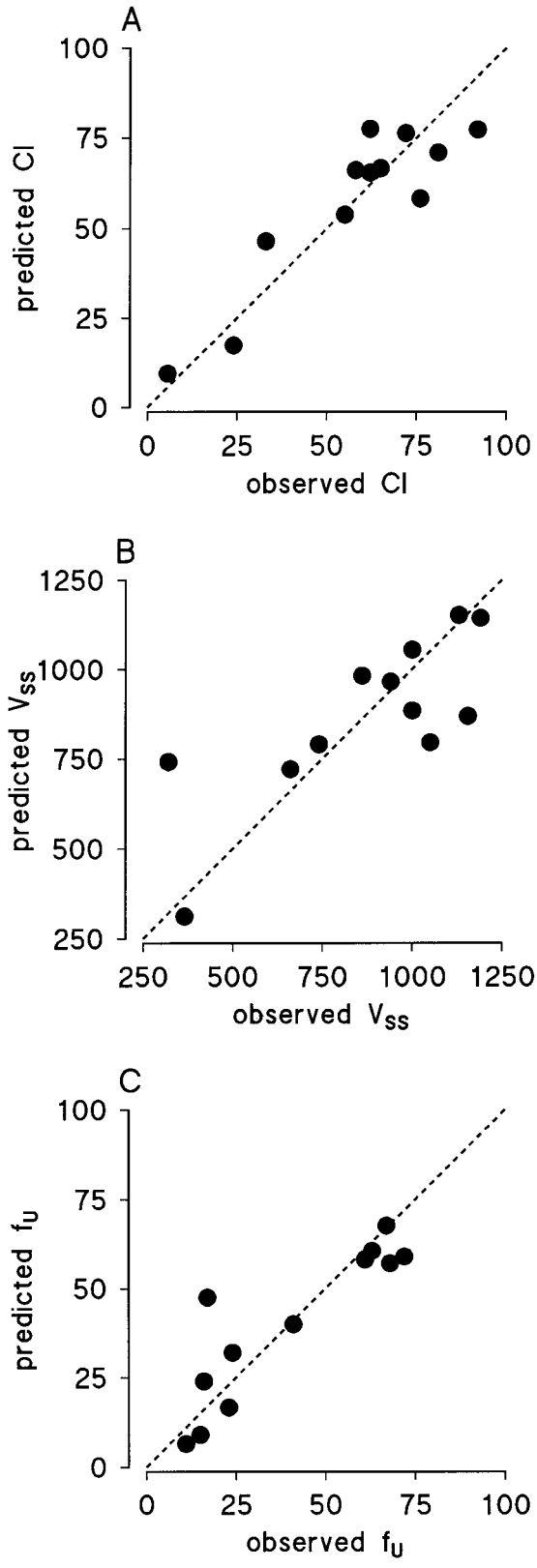


Figure 3—Relationship between observed and predicted (A) *CL*, (B) Vd_{SS}, and (C) f_U using the PLS models summarized in Table 3. The dashed lines represent lines of identity.

the number of compounds used for the modeling was limited. However, despite the relatively small dataset, the predictive ability of the QSPKR model as judged by leave-one-out cross validation was fair (i.e., difference between R^2 and Q^2 was ~0.2 in each case). The between-ligand variation in the pharmacokinetic parameters was accounted for in an adequate manner by the models in most cases (Figure 3). The exceptions were the predictions of Vd_{SS} for CPA (**1**, predicted 742 mL/kg, observed 320 mL/kg) and f_U for **12** (predicted 47%, observed 17%). At present we have no explanation for these outliers, but in the case of CPA, the overestimation of Vd_{SS} may at least in part be due to variability in the experimental estimate because we have recently[23] obtained a >30% higher value (421 mL/kg). Furthermore, the use of the standard PLS algorithm implies the assumption of a linear relationship between molecular descriptors and pharmacokinetic parameters. If this assumption would not be valid, nonlinear variations of PLS might provide better predictions. However, the nonlinear algorithms with polynomial, spline, or neural-network inner relations as implemented in the MATLAB PLS_Toolbox have not yet been widely applied and require further validation.

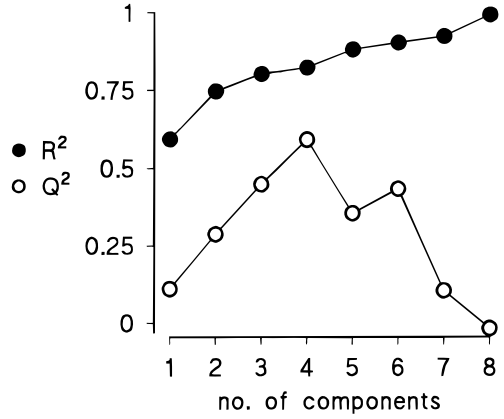


Figure 4—Effect of increasing the number of PLS components in the model for *CL* on the values of R^2 and Q^2. In this case, the optimum number of components was found to be 4 (Table 3).

The GRID program was used to generate 3D molecular descriptors in this study. Because GRID generates datasets with a very large number (27 075 in the present study) of highly correlated variables, it is desirable to employ some form of data pretreatment to reduce the number of datapoints in the descriptor matrix. Because it was found that three principal components could describe >90% of the total variance in the GRID matrix, we decided to replace the original variables with these principal components (3DPPs). An advantage of the use of principal properties, particularly when the number of compounds is relatively small, is that it is much more efficient in reducing the number of variables than other data pretreatment methods generally used in conjunction with GRID, such as D-optimal variable preselection.[25,26] Furthermore, being principal components, the 3DPPs display no collinearity. Another advantage is that the 3DPPs can be related to other molecular descriptors, which allows for a physicochemical interpretation of the GRID analysis. In the present study, it was found that the first and second 3DPP were closely associated with IOPO and POT0, respectively (Figure 2B). To our knowledge, this is the first example where such a relationship between GRID and physicochemical properties is shown.

A qualitative interpretation of the QSPKR models can be made on the basis of the PLS weighted regression coefficients (Figure 5). High values of 3DPP2 and the related descriptor POT0, and to a lesser degree 3DPP1 and the related descriptor IOPO, are associated with low *CL*. In contrast, high values of 3DPP3 and POM1 result in an increase of *CL*. The influences on *CL* of the remaining molecular properties, including log *P* (as calculated), appear to be less important (Figure 5A).

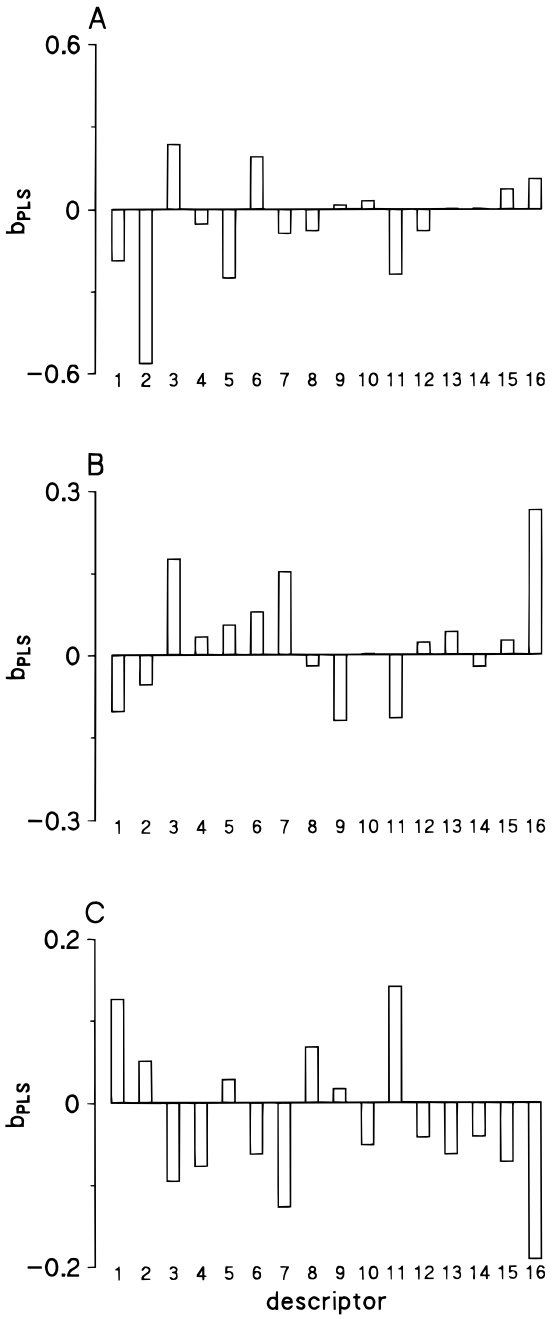


Figure 5—Weighted PLS regression coefficients (b_{PLS}) for (A) *CL*, (B) Vd_{SS}, and (C) f_U. Numbers correspond to the descriptor variables in Table 2.

In the case of Vd_{SS}, the high value of the regression coefficient for log *P* is consistent with many previously published reports that high lipophilicity is associated with large Vd_{SS}.[7] However, from Figure 5B it can be seen that 3DPP3 and HEFO are also important factors that are positively correlated with Vd_{SS}, whereas increased values of 3DPP1, IOPO, and DIPO will reduce Vd_{SS}.

Figure 5C shows that many factors contribute to the model for f_U, which would be expected on the basis of the high number of significant correlations found previously (see *Results*). The most influential factors are log *P* and to a lesser degree HEFO, which are negatively correlated, and the two related descriptors 3DPP1 and IOPO, which are positively correlated with f_U. This result confirms the conclusion made by others that lipophilicity is a key determinant of protein binding but that other physicochemical features, in particular electronic properties, also play an important role.[7]

To date, MLR has been used successfully as a regression technique in many QSPKR studies. The main disadvantage of MLR compared with PLS is that it performs very poorly in the case of collinearities among descriptor variables and that it can only be applied when the number of compounds exceeds the number of descriptors by at least a factor 3−5.[15,30] Therefore, PLS has replaced MLR as method of choice in modern QSAR studies and it may be expected that this will also happen in the area of QSPKR when the methods presented in this paper are developed further and applied to large datasets with even more molecular descriptors (for example original GRID points) and pharmacokinetic properties. The main appeal of MLR is that it may yield models that are apparently easy to interpret. When a dependent variable is strongly correlated with one of the independent variables, it should, in general, be possible to generate a simple MLR model based on the influential descriptor and one or two additional orthogonal variables that explain a fraction of the variation similar to that explained by PLS. For example, on the basis of the strong influence of log *P* on Vd_{SS} and f_U, one can calculate the simple MLR relationships $Vd_{SS} = 236.2 + 181.6 \log P + 1.2$ MW and $f_U = 92.4 - 15.4 \log P - 6.4$ DIPO which yield R^2 values (0.67 and 0.77, respectively) that are as good as the ones obtained with PLS (Table 3). However, the molecular diversity of the ligands in this study was limited, and it may be expected that the chances of finding such simple relationships will decrease with increasing complexity of the dataset. Furthermore, multivariate QSPKR modeling should aim to predict overall pharmacokinetic behavior, whereas MLR on the basis of a few descriptors may only provide an adequate alternative for PLS for some of the parameters of interest. In the present case, for example, the best MLR model for *CL* on the basis of log *P* and one other descriptor ($CL = 156.5 + 12.9 \log P - 0.01$ POT0) predicted poorly compared to PLS ($R^2 = 0.65$ and 0.82, respectively).

Herman and Veng-Pedersen[11] have suggested that QSPKR models need not be limited to congeneric series but can be used to predict distribution kinetics of a wide variety of drugs. However, in our view, QSPKR and QSAR models should generally be regarded as having "local validity" only, that is, their utility is restricted to series of chemically related ligands with similar biological properties. Hence, we believe that QSPKR modeling may be particularly useful to analyze pharmacokinetic databases of structural analogues obtained in the early stages of drug development programs. Obviously, the aim is always to develop a model that predicts the pharmacokinetic parameters of the whole set of compounds as well as possible. However, the limitations of a QSPKR model may be just as valuable and the outliers that are poorly predicted may provide new insight into the mechanisms underlying the pharmacokinetic processes. For this purpose, however, a more physiological approach than the one used in the present study should be adopted in which the dependent variables represent intrinsic pharmacokinetic properties of individual organ systems rather than overall pharmacokinetic behavior.[10,34]

In conclusion, in the present paper we have described a QSPKR method based on 3D molecular characterization and multivariate statistical analysis. By using this method, we obtained a model that could adequately predict overall pharmacokinetic behavior of 12 adenosine A_1 receptor ligands in rat. Notwithstanding the fact that analysis of a larger dataset would be required to draw any general conclusion about the pharmacokinetic behavior of adenosine analogues, these first results obtained with a relatively small number of ligands warrant further validation and application of multivariate 3D QSPKR on larger datasets of other classes of compounds. In combination with emerg-

ing methods for high-throughput in vivo pharmacokinetic screening of mixtures of structurally related compounds[31−33] and with novel approaches in physiologically based pharmacokinetic modeling,[10,34] multivariate 3D QSPKR may become a useful tool in drug discovery and preclinical development.

References and Notes

1. Hansch, C.; Fujita, T. rho-sigma-pi Analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616−1626.
2. van de Waterbeemd, H. *Chemometric Methods in Molecular Design*, VCH: Weinheim, 1995.
3. Kubinyi, H. The quantitative analysis of structure−activity relationships. In *Burger's Medicinal Chemistry and Drug Discovery*; Wolff, M. E., Ed.; John Wiley & Sons: New York, 1995; pp 497−571.
4. Kubinyi, H. QSAR and 3D QSAR in drug design. Part 2: Applications and problems. *Drug Discovery Today* **1997**, *2*, 538−546.
5. Others have used the abbreviation "QSPR" for this purpose, but to avoid confusion, we propose to reserve this term for "quantitative structure−property relationships".[2]
6. Toon, S.; Rowland, M. Strucure-pharmacokinetic relationships among the barbiturates in the rat. *J. Pharmacol. Exp. Ther.* **1983**, *225*, 752−763.
7. Mayer, J. M.; van de Waterbeemd, H. Development of quantitative structure-pharmacokinetic relationships. *Environ. Health Perspect.* **1985**, *61*, 295−306.
8. Yamada, Y.; Ito, K.; Nakamura, K.; Sawada, Y.; Iga., T. Prediction of therapeutic doses of beta-adrenergic receptor blocking agents based on quantitative structure-pharmacokinetic/pharmacodynamic relationship. *Biol. Pharm. Bull.* **1993**, *16*, 1251−1259.
9. Gobburu, J. V. S.; Shelver, W. H. Quantitative structure-pharmacokinetic relationships (QSPR) of beta blockers derived using neural networks. *J. Pharm. Sci.* **1995**, *84*, 862−865.
10. Blackey, G. E.; Nestorov, I. A.; Arundel, P. A.; Aarons, L. J.; Rowland, M. Quantitative structure-pharmacokinetics relationships: I. Development of a whole-body physiologically based model to characterize changes in pharmacokinetics across a homologous series of barbiturates in the rat. *J. Pharmacokinet. Biopharm.* **1997**, *25*, 277−312.
11. Herman, R. A.; Veng-Pedersen, P. Quantitative structure-pharmacokinetic relationships for systemic drug distribution kinetics not confined to a congeneric series. *J. Pharm. Sci.* **1994**, *83*, 423−428.
12. Cupid, B. C.; Beddell, C. R.; Lindon, J. C.; Wilson, L. D.; Nicholson, J. K. Quantitative structure-metabolism relationships for substituted benzoic acids in the rabbit: prediction of urinary excretion of glycine and glucuronide conjugates. *Xenobiotica* **1996**, *26*, 157−176.
13. Cramer III, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analyses (COMFA). 1. Effects of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
14. Goodford, P. Multivariate characterization of molecules for QSAR analysis. *J. Chemom.* **1996**, *10*, 107−117.
15. Geladi, P.; Kowalski, B. R. Partial least squares: a tutorial. *Anal. Chem. Acta* **1985**, *185*, 1−17.
16. Cecchetti, V.; Filipponi, E.; Fravolini, A.; Tabarrini, O.; Bonelli, D.; Clementi, M.; Cruciani, G.; Clementi, S. Chemometric methodologies in a quantitative structure−activity relationship study: the antibacterial activity of 6-aminoquinolones. *J. Med. Chem.* **1997**, *40*, 1698−1706.
17. IJzerman, A. P.; van der Wenden, E. M.; Roelen, H. C. P. F.; Mathôt, R. A. A.; Von Frijtag Drabbe Künzel, J. K. Partial agonists for adenosine receptors. In *Perspectives in Receptor Research*; Giardinà, D.; Piergentili, A.; Pigini M., Eds.; Elsevier Science B. V.: Amsterdam, 1996; pp 181−192.
18. Van der Graaf, P. H.; Van Schaick, E. A.; Mathôt, R. A. A.; IJzerman, A. P.; Danhof, M. Mechanism-based pharmacokinetic-pharmacodynamic modeling of the effects of N^6-cyclopentyladenosine analogues on heart rate in rat: estimation of in vivo operational affinity and efficacy at adenosine A_1 receptors. *J. Pharmacol. Exp. Ther.* **1997**, *283*, 809−816.
19. Mathôt, R. A. A.; van Schaick, E. A.; Langemeijer, M. W. E.; Soudijn, W.; IJzerman, A. P.; Danhof, M. Pharmacokinetic-pharmacodynamic relationship of the cardiovascular effects of adenosine A_1 receptor agonist N^6-cyclopentyladenosine in the rat. *J. Pharmacol. Exp. Ther.* **1994**, *268*, 616−624.
20. Mathôt, R. A. A.; van der Wenden, E. M.; Soudijn, W.; IJzerman, A. P.; Danhof, M. Deoxyribose analogues of N^6-cyclopentyladenosine (CPA): partial agonists at the adenosine A_1 receptor in vivo. *Br. J. Pharmacol.* **1995**, *116*, 1957−1964.
21. Mathôt, R. A. A.; van den Aarsen, B. C.; Soudijn, W.; Breimer, D. D.; IJzerman, A. P.; Danhof, M. Pharmacokinetic-pharmacodynamic modelling of the cardiovascular effects of R- and S−N^6-phenylisopropyladenosine (PIA) in conscious normotensive rats. *J. Pharmacol. Exp. Ther.* **1995**, *273*, 405−414.
22. van Schaick, E. A.; Mathôt, R. A. A.; Gubbens-Stibbe, J. M.; Langemeijer, M. W. E.; Roelen, H. C. P. F.; IJzerman, A. P.; Danhof, M. 8-Alkylamino substituted analogues of N^6-cyclopentyladenosine (CPA) are partial agonists for the cardiovascular adenosine A_1 receptors in vivo. *J. Pharmacol. Exp. Ther.* **1997**, *283*, 800−809.
23. van Schaick, E. A.; Kulkarni, C.; Von Frijtag Drabbe Künzel, J. K.; Mathôt, R. A. A.; Cristalli, G.; IJzerman, A. P.; Danhof, M. Time course of action of three adenosine A_1 receptor agonists with differing lipophilicity in rats: comparison of pharmacokinetic, hemodynamic and EEG effects. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **1997**, *356*, 827−837.
24. IJzerman, A. P.; Van Galen, P. J. M.; Jacobson, K. A. Molecular modeling of adenosine receptors. I. The ligand binding site on the A_1 receptor. *Drug Design Discov.* **1992**, *9*, 49−67.
25. Nilsson, J.; Wikström, H.; Smilde, A.; Glase, S.; Pugsley, T.; Cruciani, G.; Pastor, M.; Clementi S. GRID/GOLPE 3D quantitative structure−activity relationship study on a set of benzamides and naphthamides, with affinity for the dopamine D_3 receptor subtype. *J. Med. Chem.* **1997**, *40*, 833−840.
26. Nilsson, J.; Homan, E. J.; Smilde, A. K.; Grol, C. J.; Wikström, H. A multiway 3D QSAR analysis of a series of (S)-N-[(1-ethyl-2-pyrrolidinyl)methyl]-6-methoxybenzamides. *J. Computer-Aided Mol. Design* **1998**, *12*, 81−93.
27. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
28. Rekker, R. F.; De Kort, H. M. The hydrophobic fragmental constant: an extension to a 1000 data point set. *Eur. J. Med. Chem.* **1979**, *6*, 479−488.
29. Ghose A. K.; Crippen G. M. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure−activity relationships. I. Partition coefficients as a measure of hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565−577.
30. Eriksson, L.; Johansson, E. Multivariate design and modeling in QSAR. *Chemom. Intell. Lab. Syst.* **1996**, *34*, 1−19.
31. Berman, J.; Halm, K.; Adkison, K.; Shaffer, J. Simultaneous pharmacokinetic screening of a mixture of compounds in the dog using API LC/MS/MS analysis for increased throughput. *J. Med. Chem.* **1997**, *40*, 827−829.
32. Olah, T. V.; McLoughlin, D. A.; Gilbert J. D. The simultaneous determination of mixtures of drug candidates by liquid chromatography atmospheric pressure chemical ionization mass spectrometry as an in vivo drug screening procedure. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 17−23.
33. Korfmacher, W. A.; Cox, K. A.; Bryant, M. S.; Veals, J.; Ng, K.; Watkins, R.; Lin, C.-C. HPLC-API/MS/MS: a powerful tool for integrating drug metabolism into the drug discovery process. *Drug Discovery Today* **1997**, *2*, 532−537.
34. Nestorov, I. A.; Aarons, L. J.; Rowland, M. Physiologically based pharmacokinetic modeling of a homologous series of barbiturates in the rat: a sensitivity analysis. *J. Pharmacokinet. Biopharm.* **1997**, *25*, 413−448.

JS980294A