J Mol Model (2003) 9:77–83
DOI 10.1007/s00894-002-0111-z

ORIGINAL PAPER

Tore Brinck · Ping Jin · Yuguang Ma · Jane S. Murray · Peter Politzer

Segmental analysis of molecular surface electrostatic potentials: application to enzyme inhibition

Received: 19 August 2002 / Accepted: 12 November 2002 / Published online: 21 February 2003

Abstract We have recently shown that the anti-HIV activities of reverse transcriptase inhibitors can be related quantitatively to properties of the electrostatic potentials on their molecular surfaces. We now introduce the technique of using only segments of the drug molecules in developing such expressions. If an improved correlation is obtained for a given family of compounds, it would suggest that the segment being used plays a key role in the interaction. We demonstrate the procedure for three groups of drugs, two acting on reverse transcriptase and one on HIV protease. Segmental analysis is found to be definitely beneficial in one case, less markedly so in another, and to have a negative effect in the third. The last result indicates that major portions of the molecular surfaces are involved in the interactions and that the entire molecules need to be considered, in contrast to the first two examples, in which certain segments appear to be of primary importance. This initial exploratory study shows that segmental analysis can provide insight into the nature of the process being investigated, as well as possibly enhancing the predictive capability.

Keywords Molecular surface electrostatic potentials Segmental analysis · Enzyme inhibition

Introduction

The electrostatic potential $V(\mathbf{r})$ that is created in the space around a molecule by its nuclei and electrons, defined by Eq. (1), is well established as a guide to molecular interactive behavior. [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

T. Brinck
Department of Physical Chemistry,
Royal Institute of Technology,
100 44 Stockholm, Sweden

P. Jin · Y. Ma · J. S. Murray · P. Politzer (✉)
Department of Chemistry,
University of New Orleans,
New Orleans, LA, 70148, USA
e-mail: ppolitze@uno.edu

$$V(\mathbf{r}) = \sum_A \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}')\mathrm{d}\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|} \quad (1)$$

In Eq. (1), Z_A is the charge on nucleus A, located at $\mathbf{R}_A$, and $\rho(\mathbf{r})$ is the electronic density function of the molecule. $V(\mathbf{r})$ is a physical observable, which can be determined experimentally, by diffraction methods, as well as computationally. [4, 7] Its sign at any point in space depends upon which of the two terms on the right side of Eq. (1) dominates; the first describes the contribution of the nuclei and is positive, while the second reflects the effect of the electrons and is therefore negative. The electrostatic potential is most effective in indicating the favored initial path of approach of an electrophile, and in analyzing noncovalent interactions or the early stages of processes that may eventually involve bond-breaking/forming; the separations in such situations are sufficient to minimize complications due to polarization and/or charge transfer. [3, 5, 10, 11, 12, 13] For these purposes, attention has increasingly focused upon the potential computed on the molecular surface, $V_S(\mathbf{r})$, since this is what other reactants initially encounter.

This of course poses the question of how to define a molecular surface, for which there is no rigorous basis. One approach involves intersecting spheres centered on the nuclei, having van der Waals or other suitable radii. [14, 15, 16] We normally prefer to follow Bader et al. [17] in taking the surface to be some outer contour of the electronic density, e.g. $\rho(\mathbf{r})$=0.001 or 0.002 electrons bohr^{-3}. It then reflects the specific features of the particular molecule, such as lone pairs or strained bonds.

We have shown that the most negative and most positive values of the surface potential, $V_{S,min}$ and $V_{S,max}$, correlate with empirically developed scales of hydrogen bond basicity and acidity, respectively. [18, 19, 20] However, while $V_{S,min}$ and $V_{S,max}$ are certainly key features of $V_S(\mathbf{r})$, they are site-specific, and cannot possibly convey all the information contained in it. Accordingly, we have sought to develop mechanisms for more adequately describing and quantitatively characterizing the electrostatic potential over the *entire* molecular

surface. We have found that this can be achieved through the introduction of several statistically defined global quantities that explicitly reflect the magnitude of $V_S(\mathbf{r})$ at each point on the surface. [21, 22, 23] These will be introduced in the next section.

In a series of studies, reviewed elsewhere, [23, 24, 25] we have shown that different subsets of these site-specific and global quantities can be used to develop analytical representations of good accuracy for a variety of solution, liquid and solid phase properties that depend upon noncovalent interactions. These properties include heats of fusion, vaporization and sublimation, boiling points, critical constants, solubilities and solvation energies, partition coefficients, liquid and solid densities, surface tensions, viscosities, diffusion coefficients, lattice energies and impact sensitivities. Our procedure is to utilize a statistical analysis package, e.g. SAS, [26] to find the subset of our computed quantities to which can best be fit an experimental database for the particular property.

We have now begun to extend this approach to interactions in biological systems, focusing initially upon two classes of HIV (human immunodeficiency virus) enzyme inhibitors. These target two enzymes that are essential for HIV formation and activity. The first, reverse transcriptase (RT), is required for the reverse transcription of viral RNA into double-stranded DNA, [27, 28] a key step in HIV replication. The RT inhibitors can be divided into two categories: nucleoside and nonnucleoside, which differ in that the former are substrate analogues and the latter are not. The second enzyme, HIV protease, promotes the conversion of polypeptide precursors into the smaller protein fragments required for packaging the budding virions. [28] Its inhibition in vitro produces immature and noninfectious progeny virions. [29, 30] Thus, the inhibition of either RT or HIV protease is an attractive therapeutic strategy.

Our initial studies treated three families of nonnucleoside RT inhibitors, [31, 32] and followed our usual procedure, which involves computing the site-specific and global quantities over the entire surfaces of the molecules of interest and using these to obtain expressions for anti-HIV activity. The results were quite satisfactory; the correlation coefficients, R, were between 0.930 and 0.952.

In the present paper, we introduce a new version of this approach. We will divide the molecules into several portions, or segments, and treat each of these separately to establish relationships for anti-HIV potency. If one of them gives a better correlation than does the whole molecule, then it may correspond to the reactive region for the particular enzyme interaction. The reason for the improvement would presumably be that the relevant site-specific and/or global quantities now reflect only the key segments of the molecules; their values are not distorted by the contributions of inactive regions. We will present the results of these analyses for two groups of RT inhibitors

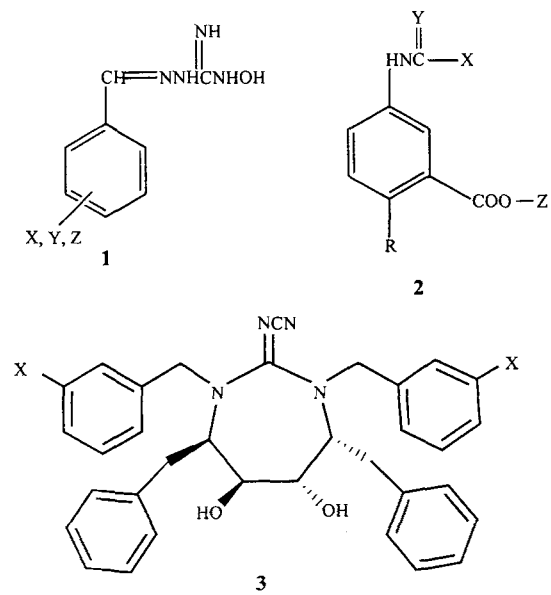


Fig. 1 Chemical structure for *N*-hydroxy-N^1-aminoguanidine derivatives (**1**), carboxanilide derivatives (**2**) and cyanoguanidine derivatives (**3**)

Table 1 Calculated and experimental data for *N*-hydroxy-N^1-aminoguanidine derivatives. Calculated properties are for the phenyl-X,Y,Z group, 4. The $\log(1/TC_{50})$ values are from Table 38 of Garg et al. [28]. A_S^+ and A_S^- are positive and negative surface areas, in Å^2; $A_S^+ + A_S^- = A_S$. All remaining quantities are defined in text. $\overline{V}_S^+$, $\overline{V}_S^-$ and Π are in kcal mol^{-1}; σ_+^2, σ_-^2 and $\sigma^{2,tot}$ are in (kcal mol^{-1})^2; ν is dimensionless

X,Y,Z	$\log(1/TC_{50})$	A_S^+	A_S^-	$\overline{V}_S^+$	$\overline{V}_S^-$	Π	σ_+^2	σ_-^2	$\sigma^{2,tot}$	ν
3-(OCH_3),4-(OCH_3),6-Br	7.30	138.1	48.7	8.8	−10.3	7.7	19.2	55.9	75.1	0.190
3-(OCH_3),4-(OCH_3),6-NO_2	7.20	142.9	48.5	13.1	−17.5	11.8	22.1	80.9	103.0	0.169
3-F,4-OCH_3,H	7.05	103.0	37.7	7.9	−7.4	6.6	22.0	62.2	84.3	0.193
3-F,H,H	6.93	90.9	19.1	6.9	−7.5	5.4	23.4	29.2	52.6	0.247
2-(OCH_3),3-CH_3,4-(OCH_3)	6.80	134.5	53.7	8.8	−7.6	7.4	23.7	69.8	93.5	0.189
H,H,H	6.71	79.2	28.5	7.1	−2.3	4.8	20.7	2.6	23.3	0.099
3-OH, 4-OCH_3,H	6.71	98.7	48.5	10.5	−11.9	10.3	90.4	127.2	217.6	0.243
3-(OH),4-(OH),H	6.69	77.1	46.2	13.0	−8.9	11.1	129.0	93.4	222.4	0.244
3-(OCH_3),4-(OCH_3),H	6.59	129.0	41.0	7.0	−9.9	6.6	16.2	96.9	113.2	0.123
3-OCH_3,4-OH,6-Cl	6.57	102.9	56.2	13.4	−9.8	11.3	65.6	70.5	136.1	0.250
2-Cl,3-OH,4-OCH_3	6.57	124.4	37.1	8.5	−12.9	8.1	28.2	121.1	149.3	0.153
2-(OCH_3),4-(OCH_3),H	6.49	118.3	50.7	12.0	−7.5	9.4	48.5	54.7	103.2	0.249
3-OCH_3,4-OH,H	6.23	97.7	48.3	11.6	−8.6	9.7	52.2	94.7	146.8	0.229

Table 2 Calculated and experimental data for carboxanilide derivatives. Calculated properties are for the entire molecule, **2**. A_S^+ and A_S^- are positive and negative surface areas, in Å^2; $A_S^+ + A_S^- = A_S$. All remaining quantities are defined in text. $\overline{V}_S^+$, $\overline{V}_S^-$ and Π are in kcal mol^{-1}; σ_+^2, σ_-^2 and σ_{tot}^2 are in (kcal mol^{-1})^2; ν is dimensionless. The log(1/EC$_{50}$) values are from Table 39 of Garg et al. [28]

X	Y	R	Z	log(1/EC_{50})	A_S^+	A_S^-	$\overline{V}_S^+$	$\overline{V}_S^-$	Π	σ_+^2	σ_-^2	σ_{tot}^2	ν
$OCH(CH_3)_2$	S	4-Cl	$CH(CH_3)_2$	10.54	207.0	128.9	6.0	−9.0	7.13	17.8	94.9	112.7	0.133
$OCH(CH_3)_2$	S	4-Cl	$CH_2CH(CH_3)_2$	10.52	209.1	144.4	5.7	−8.0	6.62	18.0	72.7	90.7	0.159
$OCH(CH_3)_2$	S	4-Cl	C_3H_7	10.50	197.5	141.6	5.8	−8.3	6.86	18.2	76.1	94.3	0.156
$OC(CH_3)C_2H_5$	S	4-Cl	$CH(CH_3)_2$	10.22	207.3	144.9	5.6	−8.9	7.01	20.5	84.5	105.1	0.157
$OCH(CH_3)_2$	S	4-Cl	C_2H_5	10.00	191.0	127.6	6.0	−9.1	7.22	18.4	83.9	102.3	0.148
$OCH(CH_3)_2$	S	4-CH_3	C_2H_5	9.97	184.8	136.1	5.0	−7.6	6.19	12.8	64.2	77.0	0.139
$OCH(CH_3)_2$	S	4-Cl	C_4H_9	9.92	206.8	152.0	5.7	−7.9	6.60	17.7	73.5	91.2	0.156
$OCH(CH_3)_2$	S	4-SCH_3	$CH(CH_3)_2$	9.91	207.7	149.7	5.0	−8.2	6.45	11.5	63.7	75.2	0.130
$OCH(CH_3)_2$	S	4-Cl	$CH_2CH{=}CH_2$	9.90	181.5	148.0	6.4	−8.5	7.39	23.6	67.4	91.1	0.192
$OCH(CH_3)_2$	S	4-SCH_3	C_2H_5	9.72	198.5	140.3	5.3	−8.6	6.76	11.7	66.2	77.9	0.128
OC_2H_5	S	4-Cl	$CH(CH_3)_2$	9.58	181.2	135.6	5.9	−8.6	7.09	19.9	80.3	100.2	0.159
$OCH(CH_3)_2$	S	4-Cl	$CH_2C{\equiv}CH$	9.03	185.4	138.6	6.8	−9.4	7.91	22.2	73.7	95.9	0.178
OC_5C_{11}	S	4-Cl	$CH(CH_3)_2$	8.80	231.3	146.7	5.4	−8.0	6.34	17.6	77.7	95.3	0.151
$OCH(CH_3)_2$	S	4-Cl	CH_3	8.76	178.9	117.6	6.4	−9.3	7.54	19.6	85.4	104.9	0.152
$OCH(CH_3)_2$	O	4-Cl	$CH(CH_3)_2$	8.74	188.8	140.3	6.9	−10.7	8.59	32.5	105.9	138.4	0.180
OCH_3	S	4-Cl	$CH(CH_3)_2$	8.18	163.8	132.8	6.2	−8.6	7.33	23.1	80.8	103.9	0.173

Table 3 Calculated and experimental data for cyclic cyanoguanidine derivatives. Calculated properties are for the benzyl-X groups, **7**. A_S^+ and A_S^- are positive and negative surface areas, in Å^2; A_S^+ and $A_S^- = A_S$. All remaining quantities are defined in text. $\overline{V}_S^+$, $\overline{V}_S^-$ and Π are in kcal mol^{-1}; σ_+^2, σ_-^2 and σ_{tot}^2 are in (kcal mol^{-1})^2; ν is dimensionless. The log(1/K_i) values are from Table 58 of Garg et al. [28]

X	log(1/K_i)	A_S^+	A_S^-	$\overline{V}_S^+$	$\overline{V}_S^-$	Π	σ_+^2	σ_-^2	σ_{tot}^2	ν
C(=NOH)H	11.00	56.8	57.7	11.4	−12.6	12.0	68.9	79.1	148.0	0.249
$C(=NOH)CH_3$	10.75	97.6	61.2	8.5	−13.0	10.2	43.3	89.9	133.2	0.219
$C(=NOH)C_2H_5$	10.51	131.6	71.6	6.8	−11.3	8.3	36.2	91.4	127.6	0.203
$C(=NOH)C_3H_7$	10.51	146.8	96.9	6.5	−8.8	7.3	34.1	79.6	113.7	0.210
$C(=O)CF_3$	10.43	54.6	84.9	11.4	−10.4	10.5	58.5	48.7	107.2	0.248
$C(=O)CH_3$	10.22	86.0	37.3	8.8	−19.1	11.8	20.1	134.3	154.4	0.113
$C(=O)C_2H_5$	9.68	114.7	53.4	7.2	−14.3	9.3	19.1	151.0	170.2	0.100
C(=O)H	9.36	41.6	35.0	10.4	−18.9	14.7	23.0	114.0	137.0	0.140
$C(=O)C_3H_7$	8.85	135.9	72.3	6.5	−10.7	7.8	19.2	131.4	150.6	0.111
$C(=O)C(CH_3)_3$	8.44	180.1	63.3	5.8	−12.0	6.9	13.1	139.7	152.8	0.078
$C(=NOH)CF_3$	8.41	74.0	100.9	12.9	−11.0	11.7	94.4	37.8	132.2	0.204

(13 aminoguanidines, series **1**, Table 1, and 16 carboxanilides, series **2**, Table 2) and one group of protease inhibitors (11 cyclic cyanoguanidines, series **3**, Table 3). The aminoguanidines (**1**) and the carboxanilides (**2**) are, respectively, nucleoside and nonnucleoside RT inhibitors (see chart shown in Fig. 1). These three series of enzyme inhibitors were selected simply for illustrative purposes.

Methodology

For each molecule of interest, the electrostatic potential was computed on the molecular surface, defined as the 0.001 electrons bohr^{-3} contour of $\rho(\mathbf{r})$ [10]. (Other low-value contours of $\rho(\mathbf{r})$, e.g. 0.002 electrons bohr^{-3}, could also be used. [33]) The calculations were carried out with Gaussian 98, [34] at the HF/STO-5G*//HF/STO-3G* level, which is generally quite satisfactory for analyzing $V_S(\mathbf{r})$. [6, 9]

We characterize the surface potential $V_S(\mathbf{r})$ by means of the site-specific quantities $V_{S,max}$ and $V_{S,min}$, mentioned earlier, plus several global ones: (a) the positive and negative average potentials, $\overline{V_S^+}$ and $\overline{V_S^-}$; (b) the average deviation, Π, which is defined in terms of $\overline{V_S}$, the average of $V_S(\mathbf{r})$ over the entire surface; (c) the positive, negative and total variances, σ_+^2, σ_-^2 and σ_{tot}^2; and (d) a balance parameter, ν. These are given by Eqs. (2, 3, 4, 5, 6, 7):

$$\overline{V_S^+} = \frac{1}{\alpha}\sum_{j=1}^{\alpha} V_S^+(\mathbf{r}_j) \tag{2}$$

$$\overline{V_S^-} = \frac{1}{\beta}\sum_{k=1}^{\beta} V_S^-(r_k) \tag{3}$$

$$\overline{V_S} = \frac{1}{n}\sum_{i=1}^{n} V_S(\mathbf{r}_i) \tag{4}$$

$$\Pi = \frac{1}{n}\sum_{i=1}^{n} \left|V_S(\mathbf{r}_i) - \overline{V_S}\right| \tag{5}$$

$$\sigma_{tot}^2 = \sigma_+^2 + \sigma_-^2 = \frac{1}{\alpha}\sum_{j=1}^{\alpha}\left[V_S^+(\mathbf{r}_j) - \overline{V_S^+}\right]^2 + \frac{1}{\beta}\sum_{k=1}^{\beta}\left[V_S^-(\mathbf{r}_k) - \overline{V_S^-}\right]^2 \tag{6}$$

$$\nu = \frac{\sigma_+^2 \sigma_-^2}{\left[\sigma_{tot}^2\right]^2} \tag{7}$$

$V_S(\mathbf{r}_i)$ is the electrostatic potential at any point $\mathbf{r}_i$ on the molecular surface. $V_S^-(\mathbf{r}_j)$ and $V_S^-(\mathbf{r}_k)$ are its values at any points $\mathbf{r}_j$ and $\mathbf{r}_k$ in, respectively, the regions of positive and negative potential.

Π is viewed as a measure of the local polarity, or internal charge separation, that exists even in molecules with zero dipole moments, such as *para*-dinitrobenzene. The quantities σ_+^2, σ_-^2 and σ_{tot}^2 reflect the variabilities, or ranges, of the positive, negative and total surface potentials, and thus the tendencies for interaction through each of these; the effects of the extrema, $V_{S,max}$ and $V_{S,min}$, are particularly emphasized, due to the terms being squared in Eq. (5). Finally, the balance parameter ν was introduced as an indicator of the degree of balance between the positive and negative surface potentials. The more similar are the magnitudes of σ_+^2 and σ_-^2 the higher is the value of ν, reaching a maximum of 0.250 when $\sigma_+^2 = \sigma_-^2$ and the better able is the molecule to interact through both its positive and negative regions.

In using this approach to develop an analytical expression for some macroscopic property of interest, we begin by computing $V(\mathbf{r})$ and the various quantities defined by Eqs. (2, 3, 4, 5, 6, 7) (plus the positive and negative surface areas, A_S^+ and A_S^-) for as many molecules as possible for which the property is known experimentally. Then we use a statistical analysis program [26] to optimally fit the database to some subset of the computed quantities, as few as will permit an accurate representation.

In the present work, in addition to treating the entire molecular surface, we also divide it into segments corresponding to different chemical groups, and compute our site-specific and global quantities on each of these separately. To achieve this, we generate the molecular surface by means of an iterative process. [35, 36] An initial reasonable estimate is made of the radius of each atom in the molecule. Using this set of radii, a fused-sphere surface is then created. A grid of uniformly distributed points is defined for each atom sphere. A Newton–Raphson algorithm is used to find the radial distance of each of these points from the contour of 0.001 electrons bohr^{-3} electron density. A new radius for each atom is then obtained by adding the average of all of these radial corrections to the original radius. The process is repeated with successive sets of new atomic radii until a grid is obtained for each atom in which all the points are arbitrarily near the 0.001 contour. A special sorting routine is finally applied to remove points that are too close to each other at the interface between atoms, ensuring a final nearly uniform distribution over the molecular surface. Since this approach represents a return to the intersecting sphere model, the resulting surface can readily be divided into segments corresponding to individual atoms or groups, and $V_S(\mathbf{r})$ determined on each of these.

For each of the three series of inhibitors included in this study, our first step was to compute the surface properties for the entire molecule and then use the SAS program [26] to search for relationships between inhibitory activity and subsets (typically three) of the calculated quantities. We then divided the molecules into two or three segments and repeated the procedure for each of these. It was not assumed that the subsets used for the segments and the molecules should be the same.

Results

Aminoguanidines

We divided the aminoguanidines into two segments: the phenyl-X,Y,Z, **4**, and the remainder of the molecule, **5** (see chart shown in Fig. 2). X, Y and Z are identified in Table 1. The best results were achieved with phenyl-X,Y,Z. Our computed properties for these portions of the aminoguanidines are given in Table 1 along with the experimental values of $\log(1/TC_{50})$, where TC_{50} is the concentration causing a 50% reduction in cell growth. Our most effective three-term expression is Eq. (8), shown in Fig. 3:

$$\log(1/TC_{50}) = 4.657\Pi^{0.5} - 0.6493\ \overline{V}_S^+ - 0.01967\sigma_-^2 + 1.193 \tag{8}$$

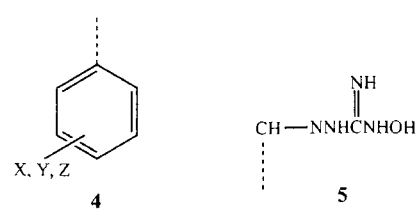
Correlation coefficient (R)=0.890Root mean square error ($RMSE$)=0.16

When the entire molecule is considered, our best three-term relationship has R=0.832.

For the same group of compounds, *but dropping two outliers*, Garg et al. obtained R=0.946 and $RMSE$=0.12, using three descriptors. [28] (When we omitted just one outlier, our R increased to 0.933.)

Carboxanilides

Three segments were considered for the carboxanilides: X, Z (see Table 2) and the remainder of the molecule, **6** (see chart in Fig. 4). The most successful three-term



4 5

Fig. 2 Two segments of the aminoguanidine derivatives: the phenyl group (**4**) and the remainder (**5**)

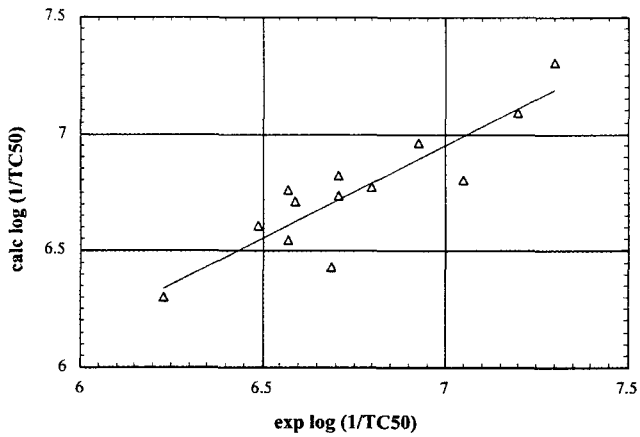


Fig. 3 Plot of calculated versus experimentally determined log(1/TC_{50}) for series **1**

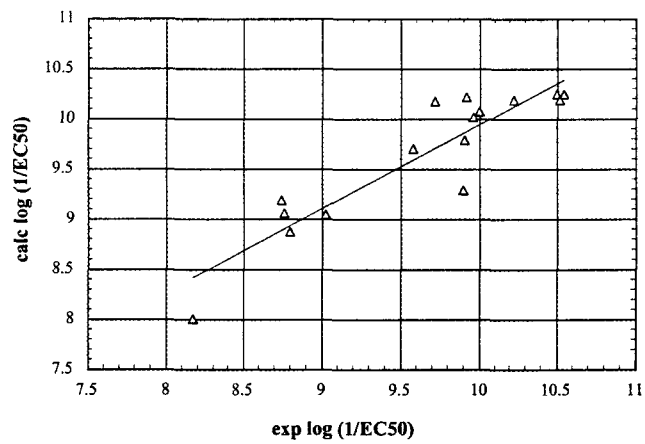
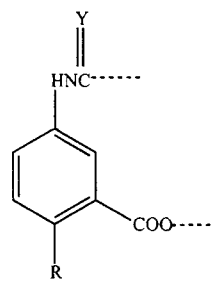
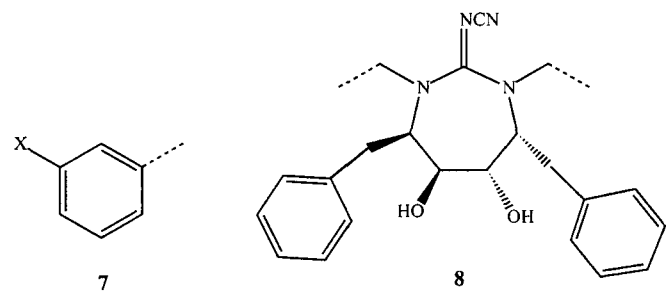


Fig. 5 Plot of calculated versus experimentally determined log(1/EC_{50}) for series **2**



6

Fig. 4 One segment of the carboxanilide derivatives: the remainder (**6**). The other two segments (X and Z) are not shown explicitly



7 **8**

Fig. 6 Two segments of the cyclic cyanoguanidine derivatives: the benzyl groups (**7**) and the remainder (**8**)

correlation with inhibiting power (EC_{50}, the concentration that reduces HIV-1-induced giant cell formation by 50%) was obtained for the whole molecules, Eq. (9) (Fig. 5). Our computed quantities for these are given in Table 2, along with the observed log(1/EC_{50}) values.

$$\log(1/EC_{50}) = 3.720\left(A_S^+\overline{V}_S^+\right)^{0.5} - 7.807 \times 10^{-4}\left(A_S^+\right)^2 - 10.99\overline{V}_S^+ - 21.47 \quad (9)$$

$R = 0.914$ $RMSE = 0.33$

The best three-term expression for any segment was for X, with $R = 0.756$. For 30 carboxanilides (including those in Table 2), from which three outliers were omitted, Garg et al. report R=0.914 and $RMSE$=0.27, for a four-descriptor relationship. [28]

Cyclic cyanoguanidines

For the cyclic cyanoguanidines, two segments were tested: the two benzyl-X, **7**, as one portion (see Table 3), and the rest of the molecule, **8** (see chart in Fig. 6). Comparing these to the intact molecule, the most effective three-term relationship to HIV protease inhibition, K_i, was found for the benzyl-X, Eq. (10) (Fig. 7). The data are in Table 3.

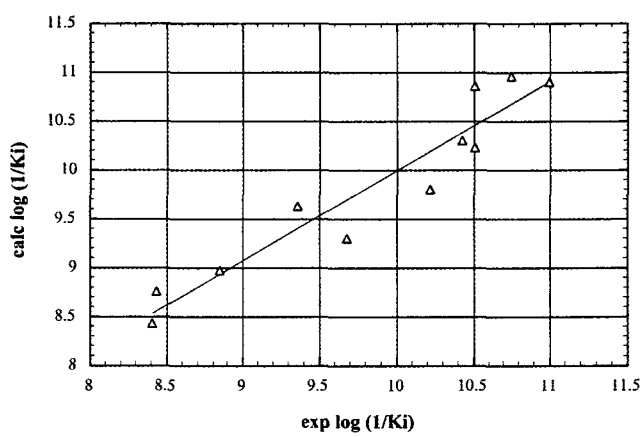


Fig. 7 Plot of calculated versus experimentally determined log(1/K_i) for series **3**

$$\log(1/K_i) = 34.98\nu - 1.665\nu\Pi - 0.04399A_S^- + 9.706 \quad (10)$$

$$R = 0.955 \quad RMSE = 0.33$$

For the entire molecule, our best three-term correlation had R=0.823. For the same compounds minus two outliers, Garg et al. had R=0.934 and $RMSE$=0.36, using two descriptors. [28]

Discussion and summary

Our objective in this work has not been to develop the best possible representations of these compounds' inhibitory activities, but rather to introduce the segmental analysis technique and to make an initial assessment of its feasibility and potential value. The results are encouraging.

We have shown that the use of molecular segments can produce a considerably better correlation (and predictive capability) than is obtained for the whole molecules – but it will not do so in every instance. Either outcome can lead to greater insight into the interactions that are involved. Thus, for the cyclic cyanoguanidines, **3**, segmental analysis is clearly beneficial. The relationship for the benzyl-X portions, with R=0.955, is much better than for the whole molecules, R=0.823. This focuses attention upon the benzyl-X groups as playing a key role in HIV protease inhibition. In contrast, for RT inhibition by the carboxyanilides, **2**, segmental analysis produces worse results than does treating the entire molecules, indicating that much of the molecular surface takes part in the interaction.

Our results for the aminoguanidines, **1**, are more ambiguous. The phenyl-X,Y,Z portions do yield a distinctly better correlation than do the whole molecules, but the change in R is only from 0.832 to 0.890 (although, as was pointed out, the removal of one outlier gives R=0.933). Part of the problem may be that the range of experimental values is the smallest among the three databases used in this work, only 1.07 compared to 2.36 for series **2** and 2.59 for **3**.

Judicious selection of the segments is of course a key factor in the procedure that has been presented, and improvement in this respect, by testing other options, may often be possible. Correlations can also sometimes be significantly enhanced by the introduction of an additional term. For example, we know from earlier work that the inclusion of Π dependence in Eq. (9) for the carboxanilides increases R from 0.914 to 0.939 and lowers the $RMSE$ from 0.33 to 0.29. [32] On the whole, however, our present results, intended to demonstrate the approach, are quite satisfactory, as is shown by the comparison to those of Garg et al., [28] especially considering that we did not drop any outliers, as they did consistently.

Finally, it may be that a modified version of our present segmental analysis technique would prove to be more effective. For example, in calculating Π, σ_+^2 and σ^2 via Eqs. (5) and (6), we have used $\overline{V}_S$, $\overline{V}_S^+$ and $\overline{V}_S^-$ that had been determined only for the particular segment of interest. Another option would be to take the $\overline{V}_S$, $\overline{V}_S^+$ and $\overline{V}_S^-$ corresponding to the entire molecular surface. This would amount to changing the reference points for Π, σ_+^2 and σ_-^2. We intend to explore the consequences of such a modification of the procedure.

References

1. Scrocco E, Tomasi J (1973) Top Curr Chem 42:95
2. Scrocco E, Tomasi J (1978) Adv Quantum Chem 11:115–193
3. Politzer P, Daiker KC (1981) In: Deb BM (ed) The force concept in chemistry. Van Nostrand Reinhold, New York, chapter 6
4. Politzer P, Truhlar DG (eds) (1981) Chemical applications of atomic and molecular electrostatic potentials. Plenum, New York
5. Politzer P, Laurence PR, Jayasuriya K (1985) Env Health Persp 61:191–202
6. Politzer P, Murray JS (1991) In: Lipkowitz KB, Boyd DB (eds) Reviews in computational chemistry, vol 2. VCH, New York, chapter 7
7. Naray-Szabo G, Ferenczy GG (1995) Chem Rev 95:829–847
8. Murray JS, Sen K (eds) (1996) Molecular electrostatic potentials: concepts and applications. Elsevier, New York
9. Murray JS, Politzer P (1998) In: Sapse A-M (ed) Molecular orbital calculations for biological systems. Oxford University Press, New York, chapter 3
10. Murray JS, Politzer P (2002) In: Carloni P, Alber F (eds) Medicinal quantum chemistry. Wiley-VCH, Weinheim, in press
11. Brinck T, Murray JS, Politzer P (1993) Int J Quantum Chem 48:73–88
12. Murray JS, Politzer P (1998) In: Parkanyi C (ed) Theoretical organic chemistry. Elsevier, Amsterdam, ch 7
13. Politzer P, Murray JS, Concha MC (2002) Int J Quantum Chem 88:19–27
14. Du Q, Arteca GA (1996) J Comput Chem 17:1258–1268
15. Brickmann J, Exner T, Keil M, Marhofer R, Moeckel G (1998) In: Schleyer PvR (ed) Encyclopedia of computational chemistry, vol 3. Wiley, New York, p 1678
16. Connolly ML (1998) In: Schleyer PvR (ed) Encyclopedia of computational chemistry, vol 3. Wiley, New York, p 1698
17. Bader RFW, Carroll MT, Cheeseman JR, Chang C (1987) J Am Chem Soc 109:7968–7979
18. Murray JS, Politzer P (1992) J Chem Res, Synop 110–111
19. Taft RW, Murray JS (1994) In: Politzer P, Murray JS (eds) Quantitative treatments of solute/solvent interactions. Elsevier, Amsterdam, chapter 3
20. Hagelin H, Murray JS, Brinck T, Berthelot M, Politzer P (1995) Can J Chem 73:483–488
21. Murray JS, Politzer P (1994) In: Murray JS, Politzer P (eds) Quantitative treatments of solute/solvent interactions. Elsevier, Amsterdam, chapter 8
22. Murray JS, Brinck T, Lane P, Paulsen K, Politzer P (1994) J Mol Struct (Theochem) 307:55–64
23. Murray JS, Politzer P (1998) J Mol Struct (Theochem) 425:107–114
24. Politzer P, Murray JS (1999) Trends Chem Phys 7:157–168
25. Politzer P, Murray JS (2001) Fluid Phase Equil 185:129–137
26. SAS, SAS Institute, Cary, NC 27511, USA
27. Fauci AS (1988) Science 239:617–622

28. Garg R, Gupta SP, Gao H, Babu MS, Debnath AK, Hansch C (1999) Chem Rev 99:3525–3601
29. Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RAF, Skolnick EM, Sigal IS (1988) Proc Natl Acad Sci USA 85:4686–4690
30. Peng C, Ho BK, Chang TW, Chang NT (1989) J Virol 63:2550–2556
31. Gonzalez OG, Murray JS, Peralta-Inga Z, Politzer P (2001) Int J Quantum Chem 83:115–121
32. Politzer P, Murray JS, Peralta-Inga Z (2001) Int J Quantum Chem 85:676–684
33. Murray JS, Brinck T, Grice ME, Politzer P (1992) J Mol Struct (Theochem) 256:29–45
34. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratman RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu C, Liashenko A, Piskorz P, Komaromi, I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson BG, Chen W, Wong MW, Andres JL, Gonzales C, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98. Gaussian, Pittsburgh Pa.
35. Francl MM, Hout RF, Hehre WJ (1984) J Am Chem Soc 106:563–570
36. Arteca GA, Grant ND (1999) J Comput-Aided Mol Des 13:315–324