

## QSAR study and VolSurf characterization of anti-HIV quinolone library

Enrica Filipponi<sup>a</sup>, Gabriele Cruciani<sup>b</sup>, Oriana Tabarrini<sup>a</sup>, Violetta Cecchetti<sup>a</sup> & Arnaldo Fravolini<sup>a,\*</sup>

<sup>a</sup>Dipartimento di Chimica e Tecnologia del Farmaco, Università di Perugia, via del Liceo 1, 06123 Perugia, Italy;

<sup>b</sup>Dipartimento di Chimica, Università di Perugia, via Elce di Sotto 9, 06123 Perugia, Italy

Received 26 May 2000; accepted 9 October 2000

**Key words:** antiviral activity, Principal Component Analysis (PCA), Projection onto Latent Structure (PLS), quinolones, VolSurf/GRID characterization

### Summary

Antiviral quinolones are promising compounds in the search for new therapeutically effective agents for the treatment of AIDS. To rationalize the SAR for this new interesting class of anti-HIV derivatives, we performed a 3D-QSAR study on a library of 101 6-fluoro and 6-desfluoroquinolones, taken either from the literature or synthesized by us. The chemometric procedure involved a fully semiempirical minimization of the molecular structures by the AMSOL program, which takes into account the solvation effect, and their 3D characterization by the VolSurf/GRID program. The QSAR analysis, based on PCA and PLS methods, shows the key structural features responsible for the antiviral activity.

### Introduction

Despite the success obtained with highly active anti-retroviral therapy (HAART) [1], the search for new anti-HIV derivatives is still a major task due to drug-resistant viral mutants and to toxicity induced by long term therapy with combined anti-reverse transcriptase and anti-protease drugs [2, 3].

In this continuing effort, fluoroquinolone derivatives have shown an interesting anti-HIV activity, in both acute and chronic infections [4–7].

Different hypotheses have been proposed for the mechanism of action of fluoroquinolones as anti-HIV agents, although no definitive data have been reported. Ciprofloxacin (Figure 1) seems to be involved in reducing Tumor Necrosis Factor- $\alpha$  (TNF- $\alpha$ ) levels and decreasing the activation of Nuclear Transcription Factor-kB (NF-kB), as well as the LTR-driven gene expression [4]. The fluoroquinolone K12 [5] (Figure 1) causes a remarkable reduction in viral mRNA synthesis, suggesting an inhibitory effect toward viral transcription, while conflicting results were obtained

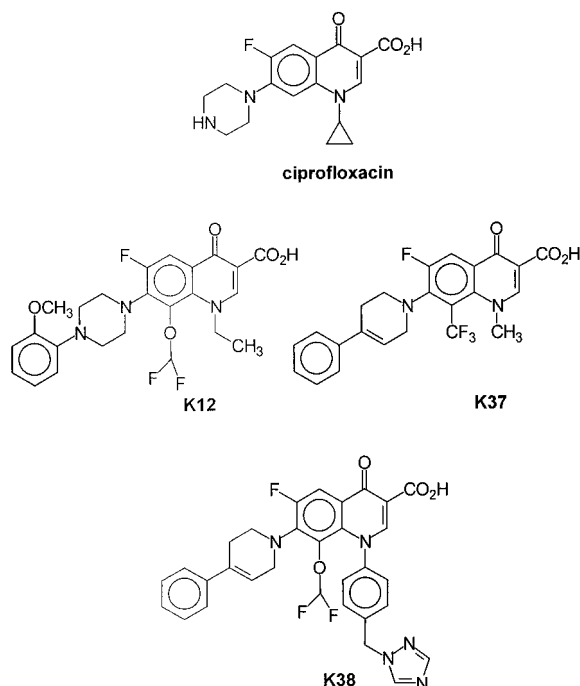
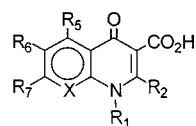


Figure 1. Anti-HIV quinolones.

\*To whom correspondence should be addressed.

E-mail: fravo@unipg.it

Table 1.



	X	R <sub>1</sub>	R <sub>2</sub>	R <sub>5</sub>	R <sub>6</sub>	R <sub>7</sub>	EC <sub>50</sub> ( $\mu$ M)
1 <sup>a</sup>	C-OCHF <sub>2</sub>	<i>c</i> -Pr	H	H	F		0.029
2	C-OCHF <sub>2</sub>	<i>c</i> -Pr	H	H	F		0.063
3	C-OCHF <sub>2</sub>	<i>c</i> -Pr	H	H	F		0.038
4	C-OCHF <sub>2</sub>	<i>c</i> -Pr	H	H	F		0.437
5	C-OCHF <sub>2</sub>	<i>c</i> -Pr	H	H	F		0.199
6	C-OCHF <sub>2</sub>	<i>c</i> -Pr	H	H	F		0.963
7	C-OCHF <sub>2</sub>	<i>c</i> -Pr	H	H	F		0.147
8	C-OCHF <sub>2</sub>	<i>c</i> -Pr	H	H	F		0.133
9	C-OCHF <sub>2</sub>	<i>c</i> -Pr	H	H	F		0.072
10	C-OCHF <sub>2</sub>	CH <sub>3</sub>	H	H	F		0.043
11	C-OCHF <sub>2</sub>	CH <sub>3</sub>	H	H	F		0.156
12	C-OCHF <sub>2</sub>	<i>i</i> -Pr	H	NH <sub>2</sub>	F		0.096
13	C-OCHF <sub>2</sub>	C <sub>2</sub> H <sub>5</sub>	H	H	F		0.496
14	C-OCHF <sub>2</sub>	C <sub>2</sub> H <sub>4</sub> F	H	H	F		0.137
15	C-OCHF <sub>2</sub>	C <sub>2</sub> H <sub>4</sub> -CO-CH <sub>3</sub>	H	H	F		0.112
16	C-OCHF <sub>2</sub>	NHCH <sub>3</sub>	H	H	F		0.646
17	C-OCHF <sub>2</sub>	C <sub>2</sub> H <sub>5</sub>	H	H	F		0.020
18	C-OCHF <sub>2</sub>	C <sub>2</sub> H <sub>5</sub>	H	H	F		0.173
19	C-OCHF <sub>2</sub>	CH <sub>2</sub> CH <sub>2</sub> F	H	H	F		1.056
20			H	H	F		1
21			H	H	F		0.074
22	N	C <sub>2</sub> H <sub>5</sub>	H	H	F		2.008
23	C-OCHF <sub>2</sub>	C <sub>2</sub> H <sub>5</sub>	H	H	F		1.047
24	C-H	<i>t</i> -Bu	H	H	F		0.94
25	C-OCHF <sub>2</sub>	C <sub>2</sub> H <sub>5</sub>	H	H	H		2.245
26	C-H			H	F		0.4
27 <sup>b</sup>	C-H	<i>t</i> -Bu	H	H	NH <sub>2</sub>		8.44
28	C-H	<i>t</i> -Bu	H	H	NH <sub>2</sub>		9.48
29	C-H	<i>t</i> -Bu	H	H	F		7.5
30	C-H	<i>t</i> -Bu	H	H	NH <sub>2</sub>		9.5

Table 1. Continued



























































31	C-H	CH <sub>3</sub>	H	H	NH <sub>2</sub>		0.1
32	C-H	<i>c</i> -Pr	H	H	NH <sub>2</sub>		0.7
33	C-H	<i>c</i> -Pr	H	H	NH <sub>2</sub>		3.2
34 <sup>c</sup>	C-CH <sub>2</sub> -N(CH <sub>3</sub> ) <sub>2</sub>	CH <sub>3</sub>	H	H	F		0.18
35 <sup>b</sup>	C-CH <sub>3</sub>	CH <sub>3</sub>	H	H	NH <sub>2</sub>		7.5
36	C-H	<i>t</i> -Bu	H	H	NH <sub>2</sub>		2
37	C-H	CH <sub>3</sub>	H	H	F		1.3
38	C-H	CH <sub>3</sub>	H	H	NH-CH <sub>3</sub>		3.5
39	C-H	<i>c</i> -Pr	H	H	F		5
40 <sup>e</sup>	C-CF <sub>3</sub>	CH <sub>3</sub>	H	H	F		0.020
41 <sup>d</sup>	C-H		H	H	H		0.7
42	C-H		H	H	H		0.1
43	C-H		H	H	H		0.2
44	C-H		H	H	H		0.13
45	C-H		H	H	H		0.12
46	C-H		H	H	H		0.25
47	C-H		H	H	H		0.3
48	C-H		H	H	H		0.09
49	C-H		H	H	H		0.7
50	C-H		H	H	H		0.08
51	C-H		H	H	H		0.08
52	C-H		H	H	H		0.3
53	C-H		H	H	H		0.3
54 <sup>c</sup>	C-H		H	H	F		3
55	C-H		H	H	F		0.7
56	C-H		H	H	F		0.7
57	C-H		H	H	F		0.25
58	C-H		H	H	F		0.2
59	N		H	H	F		0.5
60	N		H	H	F		0.4
61	N		H	H	F		0.2
62	C-H		H	H	F		0.7
63	N		H	H	F		0.6
64	C-H		H	H	F		2

Table 1. Continued

65	C-F		H	H	F		0.7
66 <sup>f</sup>	C-H		H	H	F		0.16
67	C-H		H	H	F		0.16
68	C-H		H	H	F		0.08
69	C-H		H	H	F		0.08
70	C-H		H	H	F		0.7
71	C-H		H	H	F		0.15
72	C-H		H	H	F		0.3
73	C-H		H	H	F		0.2
74	C-H		H	H	F		0.1
75	C-H		H	H	F		0.08
76	C-Cl		H	H	F		0.2
77 <sup>g</sup>	C-H		H	H	F		0.3
78	C-H		H	H	F		0.3
79	C-H		H	H	F		0.3
80	C-H		H	H	F		0.35
81	C-H		H	H	F		0.6
82	C-Cl		H	H	F		0.5
83	N		H	H	F		0.6
84	C-H		H	H	F		0.45
85	C-H		H	H	F		1
86	C-F		H	H	F		0.6
87	C-H		H	H	F		0.3
88	C-Cl		H	H	F		0.45
89	C-Cl		H	H	F		0.4
90	C-Cl		H	H	F		0.3
91	C-Cl		H	H	F		0.7

Table 1. Continued

92	C-Cl		H	H	F		0.4
93 <sup>b</sup>	N		H	H	F		0.3
94	N		H	H	F		0.3
95	C-H		H	H	F		0.002
96	C-H		H	H	H		0.3
97	C-Cl		H	H	F		0.03
98	C-Cl		H	H	F		0.015
99	C-Cl		H	H	F		0.03
100	C-Cl		H	H	F		0.3
101 <sup>i</sup>	C-OCHF <sub>2</sub>		H	H	F		0.004

<sup>a</sup>For compounds 1-26 see reference 11a). <sup>b</sup>For compounds 27-33 and 35-39 see reference 10. <sup>c</sup>For compounds 34 and 40 see reference 7. <sup>d</sup>For compounds 41-53 see reference 11b). <sup>e</sup>For compounds 54-65 see reference 11c). <sup>f</sup>For compounds 66-76 see reference 11d). <sup>g</sup>For compounds 77-92 see reference 11e). <sup>h</sup>For compounds 93-100 see reference 11f). <sup>i</sup>For compound 101 see reference 11g).

about its inhibitory activity against Tat-dependent transactivation [5, 6]. The recently reported analogues of the K12 derivative (Figure 1) K37 [7] and K38 [7] have a greater activity against chronically infected cells than the parent compound. They seem to be involved in the inhibition of Tat-induced HIV-1 LTR-driven gene expression, as well as in the suppression of TNF- $\alpha$  and the production of IL-6 in mononuclear blood cells, suggesting that they could act through an inhibition of cellular factors cooperating with Tat [7].

Recently, a large series of 6-aminoquinolones, synthesized as antibacterial agents by our research group [8], was screened for anti-HIV activity and the 1-*tert*-butyl-7-[4-(2-pyridyl)-1-piperazinyl] derivative was identified as a lead compound for the development of new anti-HIV agents [9].

A new series of 6-aminoquinolones, together with some 6-fluoroquinolones, were synthesised and tested as antiviral agents [10]. Many of these compounds showed good antiviral activity. Some important structural features were identified and a mechanism of action, entailing the interaction with TAR-RNA, was proposed for this interesting new class of anti-HIV compounds.

In order to rationalize the structure activity relationship (SAR) in-depth, a QSAR study was performed on a library of anti-HIV quinolones previously synthesized by us [10], together with literature compounds [5, 11]. This study, based on Principal Component Analysis (PCA) and Projection onto La-

tent Structures<sup>1</sup> (PLS) methods, follows a preliminary QSAR of antiviral 6-aminoquinolones [10] and it was performed in order to obtain a more satisfactory statistical model. The new data cover the experimental space of the majority of possible chemical modifications in the quinolone skeleton. As a result, those properties which are related mainly to the anti-HIV-1 activity of the quinolones considered have been identified.

## Materials and methods

### Training set molecules

The considered molecules are listed in Table 1. No experimental information on the three-dimensional (3D) structure of available antiviral quinolones is reported in the literature. Therefore, the molecular structures were built using the SYBYL molecular modeling program [12] and minimized by means of molecular mechanics TRIPOS force field [12]. Conformational analysis was performed on **R-1**, **R-7** and **X** substituents which have a certain degree of conformational flexibility, in order to identify, for each molecule, an energy minimum which could be useful for the subsequent structural optimization.

<sup>1</sup>Projection onto Latent Structure is an alternative definition of PLS model respect to the most commonly used Partial Least Squares. In our opinion it is better responding to the basic assumption of the method, that is the existence of a few 'underlying' latent variables.

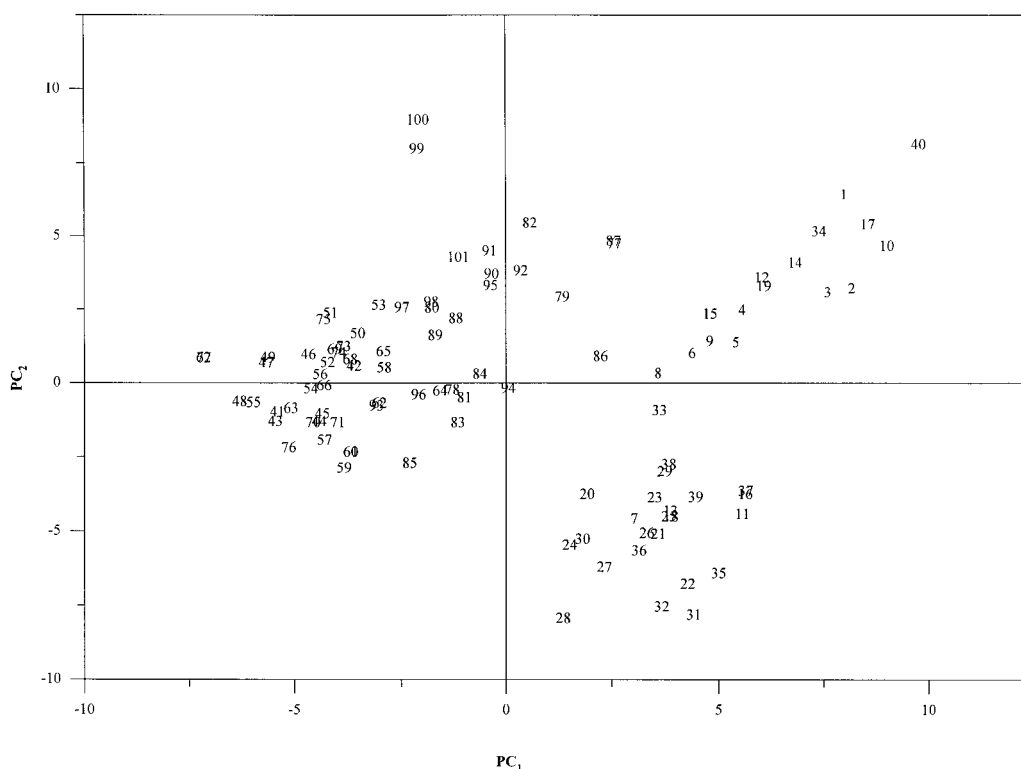


Figure 2. Score plot of the first vs. the second component for PCA model on the whole data set.

The 3D structures were imported into the AM-SOL program [13] and fully minimized using the AM1 semiempirical method, including solvation effects.

#### Descriptor variables

The molecular descriptors were derived by using the VolSurf/GRID [14] program, which is a computational procedure for producing and exploring the physico-chemical property space of a molecule, starting from 3D interaction energies grid maps between the target molecule and different chemical probes. In our study 3D-maps for water probe H<sub>2</sub>O and hydrophobic DRY (grid space of 0.5 Å) were used. As a result Vol-Surf generated the 56 chemical descriptors reported in Table 2.

The basic concept of VolSurf is to compress the information present in 3D grid maps into a few quantitative two-dimensional (2D) numerical descriptors that are very simple to understand and interpret. The principal advantage of these descriptors is that they do not require structural superimposition for a 3D-QSAR analysis, as is usually required when working with grid-field variables [15], and their numerical values are related to conformers submitted to computation.

#### Chemometric tools for classification analysis

##### Principal Component Analysis (PCA) [16]

PCA is a multivariate statistical analysis method, which allows the systematic information contained in a data matrix (in which each object is characterized by a series of descriptors) to be extracted and the original number of variables to be reduced to a few factors called Principal Components (PCs). The mathematical expression of a PCA model describes each element  $x_{ik}$  of the data matrix by Equation (1),

$$x_{ik} = \bar{x}_k + \sum_{a=1}^A t_{ia} p_{ak} + e_{ik} \quad (1)$$

where  $t_{ia}$  are called *scores* and represent the projection of the objects onto the model.

A PC can be regarded as a linear combination of the original variables and the *loadings*  $p_{ak}$  are the coefficients of the variables in the linear combination.

The results of the analysis are given in a few informative diagrams (score plot and loading plot), which permit a simple, straightforward interpretation of the problem under investigation. In our study, PCA

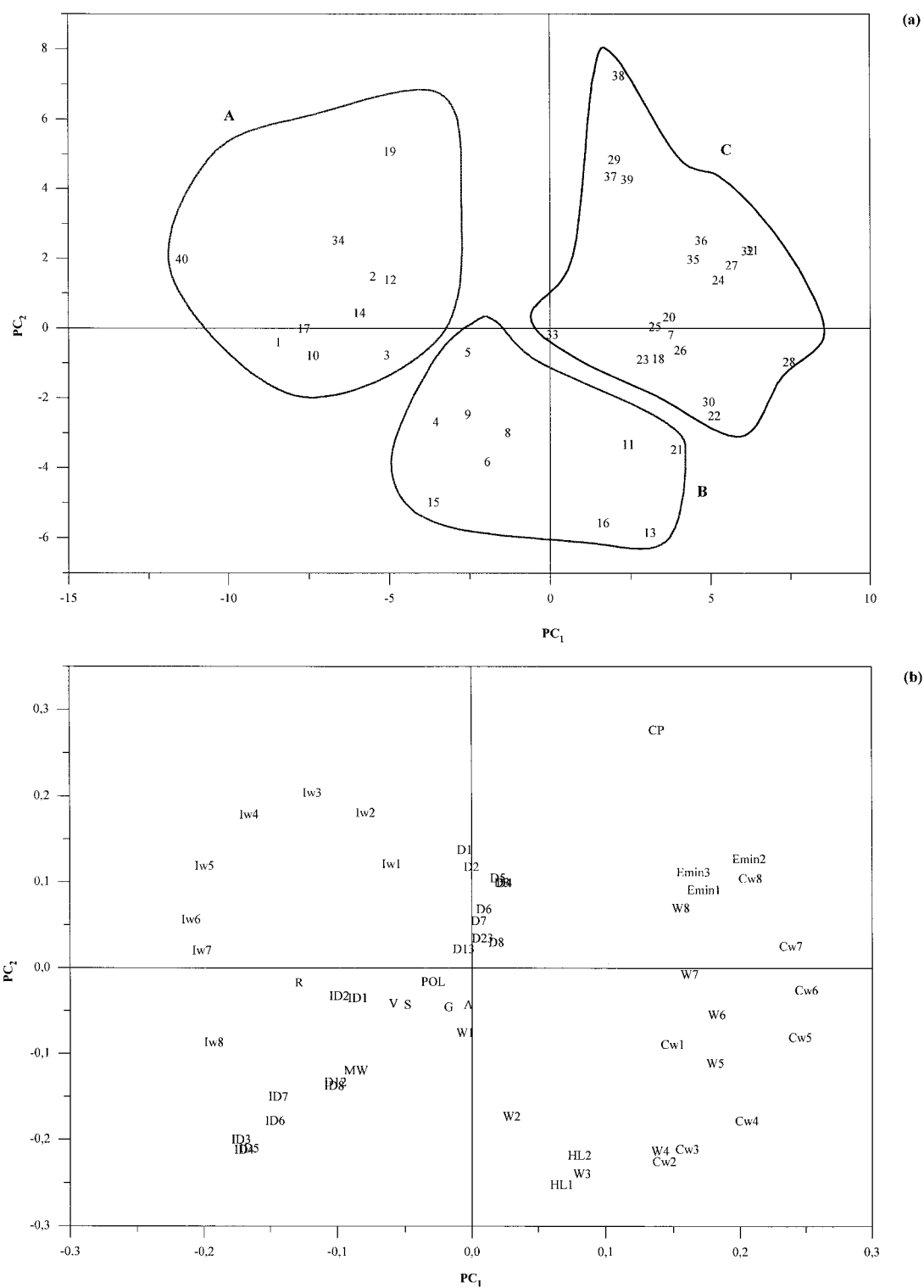


Figure 3. (a) Score plot of the first vs. the second component for PCA on the first subset. (b) Loading plot of the first vs. the second component for PCA on the first subset.

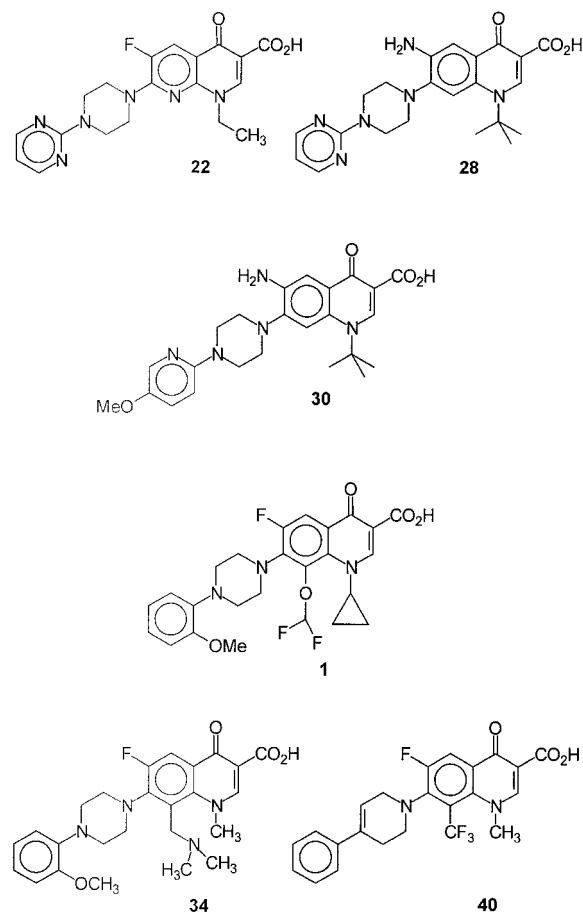


Figure 4.

was performed as implemented in the GOLPE [17] program.

#### Chemometric tools for QSAR analysis

##### Projection onto Latent Structure (PLS)

PLS<sup>2</sup> [18] has been shown to be one of the most appropriate regression methods to derive QSAR models. The objective of the analysis is to provide a relationship between the **Y** vector (anti-HIV activity data of Table 1) and the **X** matrix so that the biological behaviour of a series of molecules can be explained. A PLS model describes the **X** matrix by a Principal Component-like model (1) and the **Y** vector as a predictive relationship with the Principal Components, here called latent variables, under the constraint of maximizing the correlation between *y* and *t*.

<sup>2</sup>Since only one **Y** vector is considered, the regression method used is PLS1.

The results of a PLS model allow the relative importance of the structural features affecting the biological activity to be ranked on the basis of the relative importance of the *p<sub>ak</sub>* values, the loadings, which describe how much each original variable (here structural feature) participates in the definition of the latent variables. PLS analysis was also performed as implemented in the GOLPE [17] program.

## Results and discussion

### PCA on the whole data set of molecules

The descriptors obtained from the VolSurf calculation on 3D-GRID maps were rearranged in an **X** data matrix made of 101 rows (the considered molecules) and 56 column variables (the calculated descriptors). PCA on the autoscaled matrix gave a model of three PCs (according to random group cross-validation tests [19]) which explained 67% of the variance of the data. The information derived from the model is rather obvious. In fact, the first component divides the considered molecules into two different subsets (Figure 2). The first group, located in the positive scores region on the first PC, mainly contains molecules from **1** to **40** which are characterized by the presence of an alkyl group at N-1 position, while the second group, in the negative scores region on the first PC, contains the remaining molecules which bear bulky aryl groups at the same position.

This relevant structural difference exerts the most influence on the variance of the data. Since the first PC is not strongly related to the antiviral activity, we decided to perform calculations on the two subsets of molecules.

### PCA on the first subset of molecules

PCA on a reduced matrix (compounds **1–40**) gave a three PC model [19] which explains 63% of the variance of the data, the first two PCs being the most informative. Figures 3a and 3b show the score plot of the second *versus* the first component and the corresponding loading plot.

It is important to recall that all the descriptors used have positive numerical values and therefore, a positive loading means that an increase of the variable determines an increase in the considered PC toward positive score values, while a negative loading means that an increase of the variable determines a decrease in the corresponding PC.



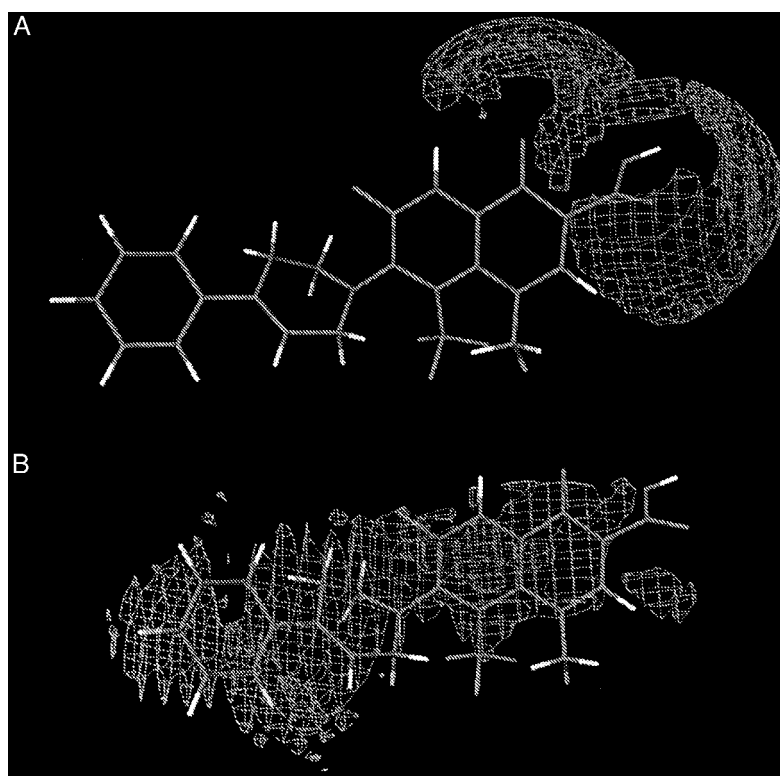


Figure 5. (a) Distribution of hydrophilic regions for molecule **40** (water probe). (b) Distribution of hydrophobic regions for molecule **40** (DRY probe).

Table 2. Used descriptors calculated by VolSurf programme

No.	Code	Description
1	<b>V</b>	Molecular volume
2	<b>S</b>	Molecular surface
3	<b>R</b>	Ratio volume/surface
4	<b>G</b>	Molecular globularity
5–12	<b>W1-W8</b>	Hydrophilic regions at eight energy levels
13–20	<b>IW1-IW8</b>	Integy moments
21–28	<b>CW1-CW8</b>	Capacity factors
29–36	<b>D1-D8</b>	Hydrophobic regions at eight energy levels
37–44	<b>ID1-ID8</b>	Hydrophobic integy moments
45–47	<b>Emin1-Emin3</b>	Local interactions energy minima
48–50	<b>d12-d13-d23</b>	Local interaction energy minima distances
51–52	<b>HL1-HL2</b>	Hydrophilic-liphophilic balance
53	<b>A</b>	Amphiphilic moment
54	<b>CP</b>	Crytical packing parameter
55	<b>POL</b>	Polarizability
56	<b>MW</b>	Molecular weight

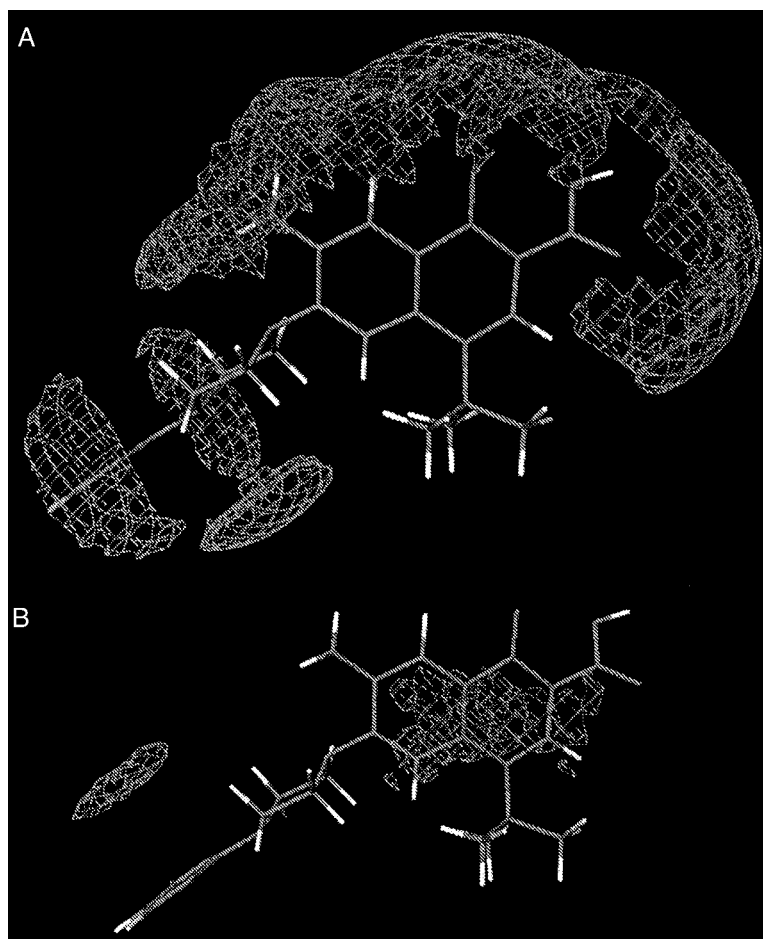


Figure 6. (a) Distribution of hydrophilic regions for molecule **28** (water probe). (b) Distribution of hydrophobic regions for molecule **28** (DRY probe).

The first PC mainly classifies molecules according to the distribution of hydrophilic regions around them. Molecules with negative scores are influenced most by integrity moments **Iw4-Iw8** and by hydrophobic integrity moments **ID3-ID7**. Integrity moments measure the imbalance between the centre of mass of a molecule and the position of hydrophilic or hydrophobic regions around it. If the integrity moment is high, there is a clear concentration of hydrated or hydrophobic regions in only one part of the molecular surface. If the integrity moment is small, the polar or hydrophobic moieties are either close to the centre of mass or they are at opposite ends of the molecule.

Conversely, at the positive extremity, the first PC is influenced most by capacity factors **CW1-CW8**, which define the amount of hydrophilic regions per surface unit and by hydrophilic regions **W5-W8** which

define the molecular envelope accessible to the solvent water molecule.

Therefore, the molecules characterized by negative PC1 scores show two localized hydrophobic and hydrophilic regions, while molecules with positive PC1 scores show hydrophilic regions distributed over greater surface areas and the hydrophobic regions do not seem to be particularly relevant.

This differentiation can be easily detected by looking at the main structural differences of the molecules at the extremes of the first PC (Figure 4).

For example, molecules **22**, **28**, and **30** with positive scores show an extensive distribution of substituents characterized by a certain hydrophilicity at the **X** position (nitrogen, for molecule **22**) and **C-6** position (amino, for molecules **28** and **30**) but they have a bulky substituent with a strong hydrophilic character, a 4-(2-pyrimidinyl)-1-piperazinyl side chain

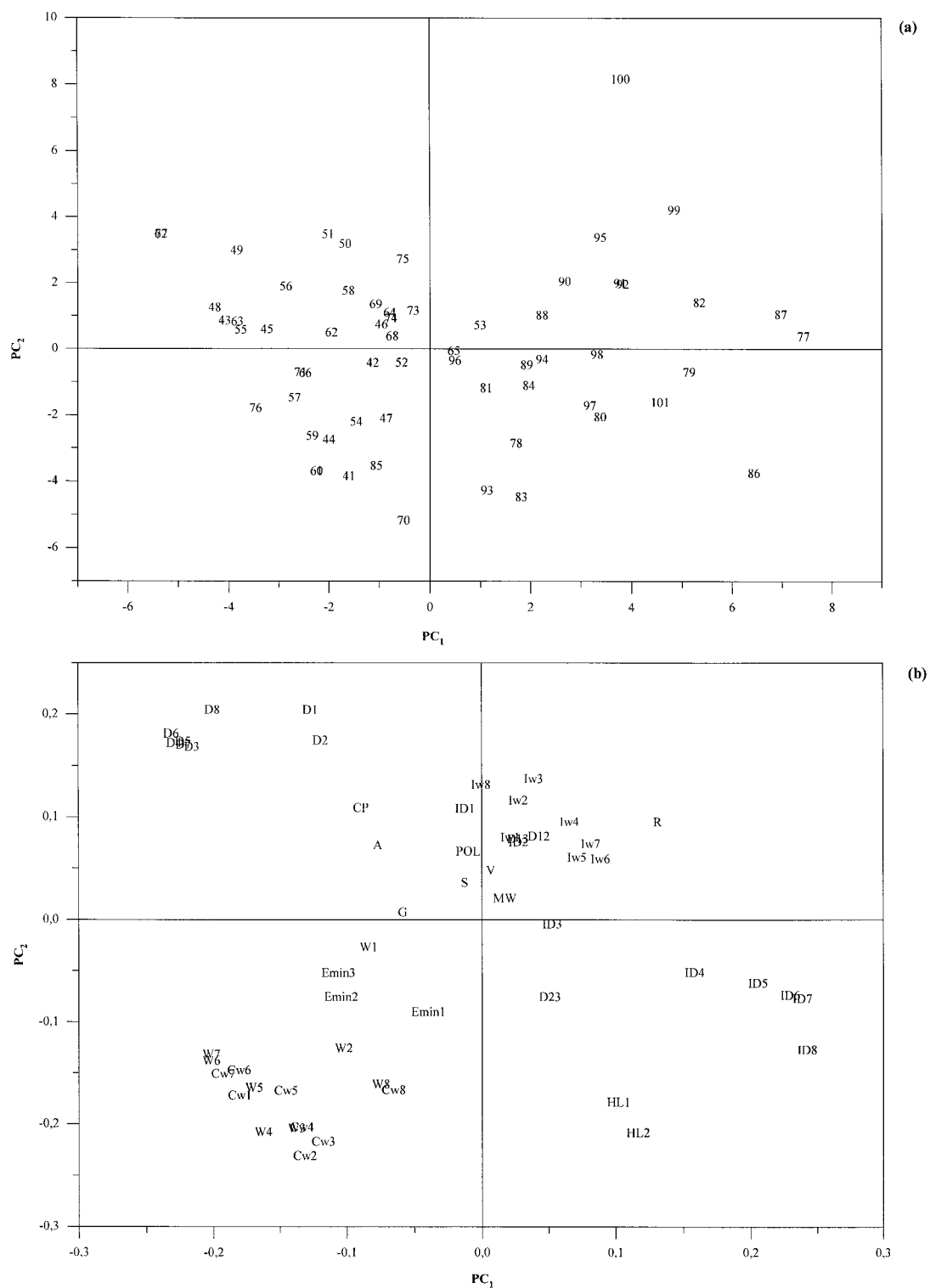


Figure 7. (a) Score plot of the first vs. the second component for PCA on the second subset. (b) Loading plot of the first vs. the second component for PCA on the second subset.

for molecules **22** and **28** or a 4-(5-methoxy-2-pyridyl)-1-piperazinyl side chain for molecule **30** at the **C-7** position. Conversely, molecules **1**, **34** and **40** with negative scores have a substituent characterized by a reduced hydrophilicity at **C-7** position due to the absence of an aromatic nitrogen with a free lone pair.

Figures 5a, 5b, 6a and 6b show the distribution of the hydrophilic (a) and hydrophobic (b) regions for molecule **40** and molecule **28**, respectively.

At the considered energy level ( $-4$  Kcal), molecule **40** has a distribution of hydrated regions only around the 4-keto-3-carboxylic moiety (which is common to all the other quinolone derivatives), while molecule **28** is also characterized by the presence of other important hydrated regions which cover the aromatic nitrogens of pyrimidine, the two heterocyclic nitrogens of piperazine and the amino group at **C-6** position. Concerning the distribution of the hydrophobic region, besides having the hydrophobic area around the quinolone nucleus, molecule **40** has another extensive hydrophobic region localized around the substituent at **C-7** (almost diametrically opposite the 4-keto-3-carboxylic moiety) at an energy level of  $-1$  Kcal. In contrast molecule **28** shows a very reduced hydrophobic region in this position at the same energy level. In other words the only significant hydrophobic region, which is present around molecule **28**, is that common to all the quinolone molecules and therefore it is not considered as statistically relevant in the calculation.

Turning our attention to the loading plot (Figure 3b), the second PC in the negative scores region is also influenced most by hydrophobic integrity moments **ID3-ID7** and by **W2-W4**, **CW2-CW4**, **HL1-HL2**. These last three groups of descriptors define the hydrophilic regions in terms of molecular envelope accessible by solvent water molecules (in particular **W2-W4** account for polarizability and dispersion forces), the capability factors which represent the amount of hydrophilic regions per surface unit and the ratio between hydrophilic and hydrophobic regions, respectively.

In the positive score value region the second PC is determined most by **CP**, the critical packing parameter, that is the ratio between the hydrophobic and hydrophilic parts of a molecule. **CP** only refers to the molecular shape and is a good parameter for predicting the molecular packing.

Practically, the second component also measures the extent of hydrophobic and hydrophilic parts of the molecule but in this component the hydrophilic-

hydrophobic balance seems to be more important, as demonstrated by the position of the **CP**, **HL1** and **HL2** descriptors.

Roughly, the score plot of Figure 3a can be divided in three parts according to the relative influences of the variables, taking into account that molecules with intermediate score values of PC1 and PC2 are influenced to a lesser extent by the above-mentioned variables. Therefore, molecules belonging to the **A** group are characterized by a concentration of hydrophilic regions in correspondence to the 4-keto-3-carboxylic moiety and a concentration of hydrophobic regions around the substituent at **C-7**. Molecules belonging to group **B** are characterized by a reduced importance of hydrophobic regions and an increasing dispersion of hydrophilic ones. Finally, molecules belonging to group **C** are characterized by a dispersion of hydrophilic regions, while the hydrophobic regions do not seem to assume particular relevance (at least on the first PC).

It is interesting to note that the majority of the molecules with particularly high anti-HIV-1 activity ( $EC_{50} \leq 0.1 \mu M$ ) are in the first group (**A**), the majority of molecules with a intermediate activity ( $EC_{50} \leq 1 \mu M$ ) are in the second group (**B**), while the majority of molecules with low activity ( $EC_{50} > 1 \mu M$ ) are in the last group (**C**). Therefore, it can be hypothesized that for molecules to have high activity, they should be characterized by an important hydrophobic region around the substituent at **C-7** position, which is diametrically opposed to the hydrophilic keto-carboxylic moiety at **C-3** and **C-4** positions. The classification obtained is of particular interest since PCA does not explicitly consider the activity of the considered structures nor training was explicitly made as in Neural Network. However, PCA was able to divide the molecules in such a way that the biological activity was recognized in a correct ranking order around the series of compounds.

#### *PCA on the second subset of molecules*

PCA on the second reduced matrix (compounds **41-101**) gave a model in which the first three PCs explained 62% of the total variance of the data. The number of components was chosen according to the random group cross-validation test [19]. Figure 7a shows the score plot of the second *versus* the first PC and Figure 7b shows the corresponding loading plot for this second model. The loading plot shows which combination of the variables plays a role in de-

termining the component trends. The first component is mainly a combination of **Cw**, **W** and **D** parameters at negative scores, **ID** parameters (and **Iw** to a lesser extent) at positive scores. Similarly, the second component is a combination of **Iw** and **D** parameters at positive scores, and of **Cw**, **W** and **HL** parameters at negative scores.

Therefore, molecules localized at positive scores on the first PC should be mainly characterized by a relative imbalance in the position of hydrophobic regions with respect to the centre of mass, while those localized at negative score values on the first PC should present a more symmetrical distribution of both hydrophobic and hydrophilic regions.

The second component describes the extent of the hydrophobic and hydrophilic regions, ranking molecules with larger hydrophobic regions and a concentration of hydrophilic ones at positive scores in the second PC and molecules with larger hydrophilic regions at negative scores.

In this model it was not easy to identify a clear correlation between the activity of the considered molecules and their relative localization in the score plot. Effectively, the molecules considered here have a homogeneous activity and therefore the relationships to chemical structure are more difficult to interpret.

However, it can be noted that the most active molecules are localised in the upper half of the score plot (first and second quadrants). In this region, molecules are influenced by a combination of hydrophilic integrity moments (**Iw1-Iw8**) and hydrophobic regions **D1-D8**, even if the opposition of hydrophilic (**Iw**) and hydrophobic (**ID**) localized regions maintains a certain importance, at least in the first quadrant. The increased importance of **D1-D8** variables in the second subset can be attributed to the positive effect of bulky substituents at N-1 position, which is the most significant difference in discriminating the two groups.

#### *PLS on the first subset of molecules*

The PCA proposed an interesting classification of the considered molecules, but it does not explicitly consider the biological activity.

In the following, PLS was applied in order to elaborate a QSAR model for both of the considered subsets of molecules.

The analysis on the first subset of molecules (**1-40**) led to a three-component model which explains 75% of the variance of the data. The number of components was chosen according to random group

cross-validation criteria [19]. Figure 8a and Figure 8b, respectively, show the results obtained in terms of the recalculated *versus* experimental activity values for the three-component model and the loading plot of the second *versus* the first component.

According to PLS analysis, variables **ID1-ID8** (which indicate a concentration of hydrophobic regions around the substituent at **C-7**) are strongly related to high antiviral activity in both the first and second component, and a relevant correlation is found on the first component with **Iw5-Iw8** (which indicates a concentration of hydrophilic regions around the keto-carboxylic moiety). The results are in agreement with those obtained by means of PCA on the first subset of molecules and they can be easily interpreted looking at the field distribution for molecule **40** in Figures 5a and 5b. In addition, PLS quantitatively correlates the present variables with the antiviral activity.

#### *PLS on the second subset of molecules*

The PLS analysis on the second subset of molecules (**41-101**) also gave a three-component model which explained 42% of the variance of the data according to random group cross-validation procedure [19]. On the first component, high activity values seem to be more related to hydrophobic integrity moments **ID4-ID8** and the **R** parameter, while the second component shows that **D3-D8** hydrophobic regions are important for the activity. The results are substantially in agreement with those of previous PLS analysis with respect to the first component, even if a reduced importance of the **Iw** parameters can be noted. A certain discrepancy was found on the second component which identifies an inverse correlation of the **D** and **ID** parameters in comparison with the second component of the first PLS model.

The second PLS classification analysis probably shows the effect of the substitution of an aryl group at **N-1** position, which determines a more extensive distribution of hydrophobic regions around the **N-1** position without causing a decrease in activity.

This effect was not indicated by the PLS analysis on the first subset of molecules which did not contain any structure with an aryl group bound to the **N-1** position.

#### *Pharmacophoric model*

The results of the chemometric study allow us to propose a revised pharmacophoric model for the antiviral

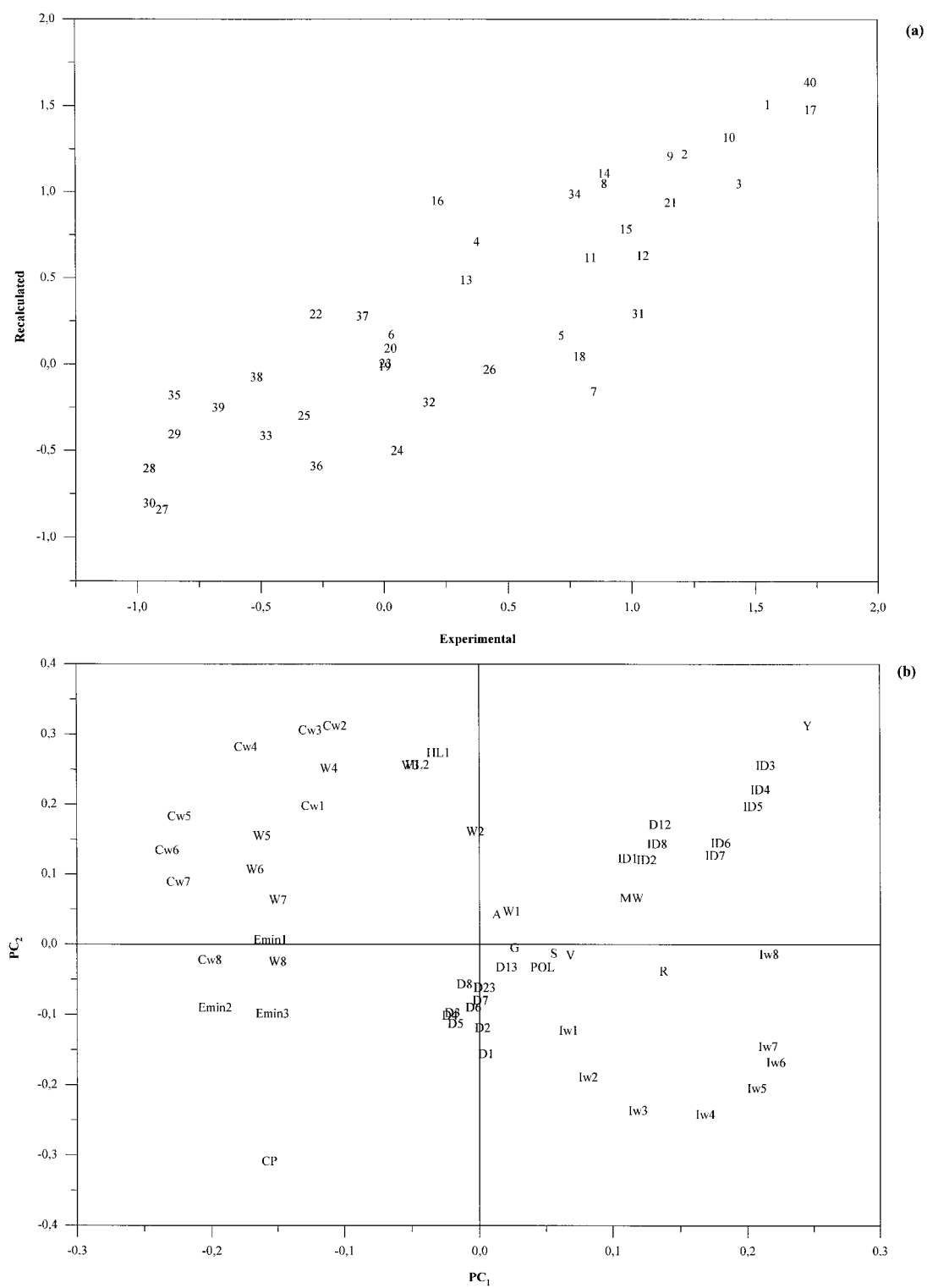


Figure 8. (a) Recalculated vs. experimental activity values for the three-component PLS model on the first subset. (b) Loading plot of the first vs. the second component for PLS model on the first subset.

activity of quinolones. In our previous paper [10], the presence of a carboxylic acid at the **C-3** position was found to be a fundamental requirement for antiviral activity. This result was confirmed and the nature of the substituents in the other positions was also studied in-depth, for the whole class of anti-HIV quinolones.

According to the PCA and PLS results, a high antiviral activity should be determined by a concentration of hydrophilic regions around the 4-keto-3-carboxylic moiety and a concentration of a suitable hydrophobic region around the substituent at **C-7** position. High antiviral activity seems to be related to the presence of a hydrophobic substituent at **N-1** position, while the nature of the substituent at **C-6** position does not seem to assume particular relevance.

New anti-HIV quinolones have been planned taking into account the proposed pharmacophoric model. Synthesis and biological evaluations are in due course and will be reported in future works.

## References

1. Palella, F., Delaney, K., Moorman, A., Loveless, M., Fuhrer, J., Satten, G., Aschman, D., Holmberg, S. and Investigators, T.H.O.S., *New Engl. J. Med.*, 338 (1998) 853.
2. Perrin, L. and Telenti, A., *Science*, 280 (1998) 1871.
3. Hecht, F.M., Grant, R.M., Petropoulos, C. J., Dillon, B., Chesney, M.A., Tian, H., Hellmann, N.S., Bandrapalli, N.I., Digilio, L., Branson, B. and Kahn, J. O., *New Engl. J. Med.*, 339 (1998) 307.
4. Gollapudi, S., Kim, C., Roshanravan, B. and Gupta, S., *AIDS Res. Human Retrovirus*, 14 (1998) 499.
5. Baba, M., Okamoto, M., Makino, M., Kimura, Y., Ikeuchi, T., Sakaguchi, T. and Okamoto, T., *Antimicrob. Agents Chemother.*, 41 (1997) 1250.
6. Witvrouw, M., Daelemans, D., Pannecouque, C., Neyts, J., Andrei, G., Snoeck, R., Vandamme, A., Balzarini, J., Desmyter, J., Baba, M. and De Clercq, E., *Antivir. Chem. Chemother.*, 5 (1998) 403.
7. Baba, M., Okamoto, M., Kawamura, M., Makino, M., Higashida, M., Takashi, T., Kimura, Y., Ikeuchi, T., Tetsuka, T. and Okamoto, T., *Mol. Pharmacol.*, 53 (1998) 1097.
8. a. Cecchetti, V., Clementi, S., Cruciani, G., Fravolini, A., Pagella, P.G., Savino, A. and Tabarrini, O., *J. Med. Chem.*, 38 (1995) 973.  
b. Cecchetti, V., Fravolini, A., Lorenzini, M.C., Tabarrini, O., Terni, P. and Xin, T., *J. Med. Chem.*, 39 (1996) 436.
9. Cruciani, G., Fravolini, A., Cecchetti, V., Filippini, E., Tabarrini, O., Palù, G. and Del Pup, L., XIII Conv. Naz. Div. Chim. Farm., Paestum, Italy, September 23–27, 1996, 71.
10. a. Cecchetti, V., Tabarrini, O., Fravolini, A., Palù, G., Gatto, B. and Palumbo, M., Italian, Hungarian, Polish, Joint Meeting on Medicinal Chemistry, Giardini Naxos-Taormina, Italy, September 28–October 1, 1999  
b. Cecchetti, V., Parolin, C., Moro, S., Pecere, T., Filippini, E., Callistri, A., Tabarrini, O., Gatto, B., Palumbo, M., Fravolini, A. and Palù, G. *J. Med. Chem.*, 43 (2000), 3799.
11. a. EP 572259 A1. Kimura, T. and Katsube, T., C.A. 121: 57343.  
b. DE 4425660 A1. Bender, W., Roeben, W., Paesens, A. and Bartel, S., C.A. 124: 261076s.  
c. DE 4303657 A1. Bartel, S., Kleefeld, G., Schulze, T., Paesens, A., Neumann, R., Reefsclaeger, J. and Streissle, G., C.A. 121: 179620u.  
d. DE 4425648 A1. Bender, W., Roeben, W., Paesens, A. and Bartel, S., C.A. 124: 289570x.  
e. DE 4425647 A1. Bender, W., Roeben, W., Paesens, A. and Bartel, S., C.A. 124: 289571y.  
f. DE 4425649 A1. Bender, W., Roeben, W., Paesens, A. and Paesens, A., C.A. 124: 261077t. g. DE 4425659 A1. Bender, W., Roeben, W., Paesens, A., Paesens, A. and Bartel, S., C.A. 124: 289573a.
12. SYBYL Molecular Modeling Software, Version 6.4, December 1997, Tripos Inc.
13. Cramer, C.J., Hawkins, G.D., Lynch, G.C., Giesen, D.J., Rossi, I., Storer, W.J., Thrular, D.G. and Liotard, D.A., *QCPE Bull.*, 15 (1995) 41.
14. a. Cruciani, G., Crivori, P., Carrupt, P.A. and Testa, B., *Teochem*, 53 (2000) 17.  
b. VolSurf, version 2.0.1, Multivariate Infometric Analysis.
15. Kubiny, H., *Drug Discovery Today*, 2 (1997) 457.
16. Wold, S., Esbensen, K. and Geladi, P., *Chemom. Intell. Lab. Syst.*, 2 (1987) 37.
17. Baroni, M. and Pastor, M., GOLPE, Version 6.4.5, 1999, Multivariate Infometric Analysis.
18. Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory Methods and Applications. ESCOM, Leiden, 1993.
19. Wold, S., *Technometrics*, 20 (1978) 397.