Journal of Computer-Aided Molecular Design, 10 (1996) 567–582
ESCOM



J-CAMD 367

# A 3D QSAR CoMFA study of non-peptide angiotensin II receptor antagonists

Laura Belvisi[a], Gianpaolo Bravi[a], Giovanna Catalano[a], Massimo Mabilia[b], Aldo Salimbeni[c] and Carlo Scolastico[a,*]

[a]*Organic and Industrial Chemistry Department, C.N.R. (National Research Council) Centre for the Study of Organic and Natural Compounds, University of Milan, Via Venezian 21, I-20133 Milan, Italy*
[b]*S.IN – Soluzioni Informatiche s.a.s., Via Salvemini 9, I-36100 Vicenza, Italy*
[c]*Medicinal Chemistry Department, Institute LusoFarmaco d'Italia S.p.a., Via Carnia 26, I-20132 Milan, Italy*

Received 15 May 1996
Accepted 8 July 1996

*Keywords:* Renin–angiotensin system; Binding requirements; Steric and electrostatic interaction energy; PLS

## Summary

A series of non-peptide angiotensin II receptor antagonists was investigated with the aim of developing a 3D QSAR model using comparative molecular field analysis descriptors and approaches. The main goals of the study were dictated by an interest in methodologies and an understanding of the binding requirements to the $AT_1$ receptor. Consistency with the previously derived activity models was always checked to contemporarily test the validity of the various hypotheses. The specific conformations chosen for the study, the procedures invoked to superimpose all structures, the conditions employed to generate steric and electrostatic field values and the various PCA/PLS runs are discussed in detail. The effect of experimental design techniques to select objects (molecules) and variables (descriptors) with respect to the predictive power of the QSAR models derived was especially analysed.

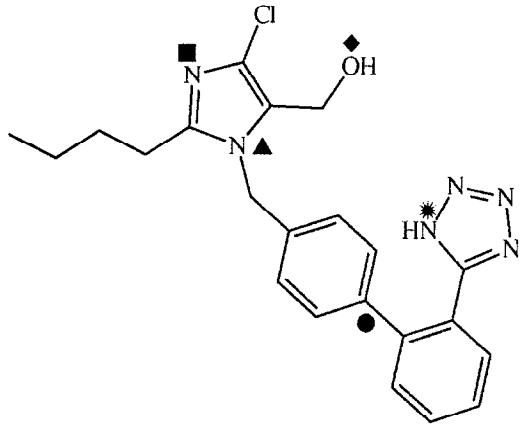
## Introduction

Angiotensin II (A II) is a potent vasoconstrictor octapeptide hormone and, in particular, is the primary active component of the renin–angiotensin system (RAS). The inhibition of this system, which plays a central role in the regulation of blood pressure, has been the target of considerable drug design efforts to control hypertension [1]. The direct antagonism of A II binding at the receptor has received extensive consideration over the last few years: in fact, it should represent the most effective and specific way of RAS modulation [2]. Furthermore, non-peptide A II ($AT_1$ subtype) receptor antagonists offer a real promise for the therapy of hypertension and heart failure. A profusion of pharmaceutical research has been initiated in this area, as demonstrated by the variety of non-peptide structures recently reported in the literature by numerous laboratories [3].

We had already performed a three-dimensional quantitative structure–activity relationships (3D QSAR) study on non-peptide A II ($AT_1$) receptor antagonists, with the aim of identifying the 3D geometrical requirements for receptor binding. By the combined use of conformational analysis and chemometrics, we carried out a comparative analysis of the 3D conformational preferences of 13 antagonists showing different levels of binding affinity and defined the minimal pharmacophoric requirements [4].

Interatomic distances between atoms belonging to relevant functional groups common to all the training set molecules were adopted as 3D geometrical molecular descriptors. The pharmacophoric points are illustrated in Fig. 1 for compound DuP-753 (losartan), a potent orally active A II receptor antagonist discovered by the DuPont group and already registered in some countries as an antihypertensive agent [5]. The selected points are (i) the atom (B, ▲) that binds the aromatic spacer; (ii) the -N= type nitrogen atom (N, ■) as a hydrogen-bond acceptor, both belonging to the same heterocyclic ring; (iii) the heteroatom hydrogen-bond acceptor (A, ◆) placed sideways to the heterocycle; (iv) the carbon atom (C, ●) of

*To whom correspondence should be addressed.




**Dup-753**

Fig. 1. Compound DuP-753 (losartan). The labelled atoms (▲, ■, ◆, ●, ✷) have been considered in the definition of conformational descriptors.

the aromatic linker; and (v) the heteroatom of the acid group (E, ✷) present as a substituent on the aromatic system. This pharmacophoric model is in fairly good agreement with the working hypothesis put forward by other authors in a recent study on the same class of compounds [6].

The relative spatial dispositions of these functionalities, adopted by the A II antagonists showing the highest levels of receptor affinity, were first investigated and then considered to characterize the 'active conformation(s)', that is the conformation(s) in which the molecules interact at the receptor site. Three significantly different putative active geometries were identified. Two of them describe different spatial dispositions of the acid moiety with respect to the heterocyclic ring and its substituents in DuPont antagonists containing the biphenylmethylimidazole moiety. The third likely active geometry is specific to A II receptor antagonists characterized by heterocyclic rings other than imidazole. This geometry differs from the previous ones mainly in terms of the distances involving atoms N (■), A (◆) and B (▲) previously defined. The difference between this last geometry and the other two is primarily due to structural diversity instead of conformational reasons. Such diversity in the type of heterocyclic ring in any case allows the molecular structures to adopt the relative spatial arrangements of the acid moiety and the heterocyclic ring as typified by the two DuPont-type putative active geometries [4].

Our 3D QSAR approach turned out to be a promising and powerful tool to suggest models for the receptor-bound geometries of ligands, when structural information on the receptor active site is lacking and the ligands are conformationally very flexible. However, improvements of our geometrical model for activity, possibly by extending the theoretical approach to other molecular descriptors, were deemed necessary to better rationalize the activity of the non-peptide A II receptor antagonists. A first improvement was achieved by performing a visual comparative analysis of the molecular electrostatic potential (MEP) distributions generated by relevant molecular fragments of several non-peptide A II antagonists by means of ab initio calculations carried out with GAUSSIAN '90 [7]. In this way, a deeper insight into the electrostatic requirements for binding was also gained.

We established that the electrostatic behaviour of the region surrounding the heterocycle is relevant to receptor interaction. A positive long-range MEP, in the region of space surrounding the lipophilic side chain, as well as strongly electrophile attracting regions bulging out of the hydrogen-bond acceptor moieties seem to be required for the recognition process. In addition, the electrostatic behaviour of the terminal aromatic ring of the biarylic portion was found to be crucial for receptor binding. Terminal heterobiaryl derivatives are characterized by long-range MEP distributions that could explain their reduced affinity compared to that of the biphenylic analogues. In fact, additional negative potential contours, in the plane of the terminal aromatic ring and surrounding the heteroatom, might give rise to unfavourable interactions with a hypothetic lipophilic pocket at the receptor site [8]. In any case, the possibility of electrostatic discrimination among antagonists showing different levels of binding affinity is related more to the overall topology of the electrostatic potential distributions in the regions of interest than to the depth (i.e. absolute value) of the MEP minima.

On the basis of these studies, non-peptide A II receptor antagonists turned out to pose quite a complex SAR problem, where both the geometrical and electronic properties of the compounds must be taken into account to fully rationalize their activity. The present work reports a further development of our initial indirect theoretical study of this promising pharmacological class. The aim of the research is the quantitative evaluation of the steric and electrostatic factors contributing to binding differences by means of comparative molecular field analysis (CoMFA) descriptors and approaches [9,10]. The basic assumption of all these methods is that structurally diverse compounds bind in conformations presenting a common steric and electrostatic shape or pattern to a common receptor site and that a suitable sampling of the steric and electrostatic potentials around a set of aligned ligands or drug molecules might provide all the information necessary for understanding their biological properties.

The use and validation of computationally determined models for the bioactive geometries of A II non-peptide antagonists, the reliability of the conclusions derived from the comparative analysis of their MEP distributions and a better understanding of the molecular requirements for

optimal drug–receptor interactions represent the basic aspects and objectives of this study. Considerable efforts have also been made in order to critically analyse most of the relevant methodological aspects of CoMFA-like analyses. Potential problems arising at each step of the methodology were investigated and, in some cases, alternative solutions were adopted for comparison purposes.

Specific topics that were dealt with include the difficulty of managing multiple energetically acceptable conformations, the choice of which procedure(s) to employ to superimpose all structures, the operative conditions to generate steric and electrostatic potential values for subsequent regression analyses and the cross-validation techniques to check the predictive ability of the models derived. The molecular descriptors employed in CoMFA are probe–ligand interaction energies computed in the space surrounding the molecules and may range from hundreds to thousands. Finally, specific attention was also paid to the significance and usefulness of molecule and variable selection procedures. While the latter is widely recognized as being crucial for treating large data matrices such as CoMFA ones [11,12], so far molecule selection has received less attention.

## Methods

*Training set molecules and relative binding affinities*

The CoMFA methodology has been applied to a starting training set of 28 structurally different and conformationally flexible antagonists (Fig. 2) to investigate the relationships between chemical structures and binding affinities to the $AT_1$ subtype A II receptor available from the literature [5,13–23].

As already pointed out for compound DuP-753, the following key structural elements can be easily identified: a heterocyclic ring, a lipophilic side chain, a hydrogen-bond acceptor group, a suitably positioned acid moiety, ionized at physiological pH, and an aromatic spacer connecting the acid group to the heterocycle. The structures that make up the training set present different alternatives to all these key structural features, including the lack of a given feature or the substitution of a cyclic unit by an acyclic one.

The selection of these compounds from the literature was performed following two basic criteria. The first one is geometrical in nature and is based on the interatomic distances derived from our 3D QSAR model, previously described. Only those compounds showing a specific spatial layout of the acid moiety and heterocyclic ring, in agreement with geometrical activity models, were considered for the CoMFA study. On this premise, a variety of different heterocyclic rings, or mimic structural units, were accepted and included in the study.

The lack of given pharmacophoric points in some compounds did not prevent us from checking the 3D spatial relationship between the functionalities of interest. In such cases, the values of a smaller number of interatomic distances were analysed. The inclusion of compounds lacking specific pharmacophoric points is perfectly in line with the overall CoMFA-like approach and is instrumental in getting more information, taking advantage of structural diversity.

This first selection criterion required the preliminary study of the conformational properties of each candidate compound (described later). The second criterion derives from the requirement for homogeneity and consistency of biological activity data. The binding properties of the antagonists at the $AT_1$ A II receptor site are tested in vitro by assays that evaluate their ability to displace different radioligands from a source of $AT_1$ receptors. Incidentally, the relative binding of antagonists can be expressed by different parameters.

The most common binding affinity values reported in the literature are those obtained from assay experiments with labelled A II as the radioligand and expressed as $IC_{50}$ values, which represent the concentrations of antagonists to displace 50% of the specifically bound radiolabelled A II. Therefore, we decided to consider for the CoMFA study only those antagonists for which this type of experimental determination is available. In spite of this selection criterion, difficulties can be encountered when putting together activity data coming from different sources. The operative assay conditions are seldom perfectly identical. As a consequence, the $IC_{50}$ values determined by different research groups for the same compound and using the same (at least 'on paper') test can be different. Luckily enough, the use of compound DuP-753 as a reference standard in all the binding assays performed by the various pharmaceutical companies makes it possible to approach such a crucial problem.

In the present work, the experimental binding affinities were expressed as the negative logarithm of $IC_{50}$ (M) ($pIC_{50}$). In particular, two different dependent variables were considered: (i) the negative logarithm of the original $IC_{50}$ (M) values reported by the various sources without considering the differences observed in the value of the reference compound DuP-753 ($pIC_{50}$ or Y-1); and (ii) the negative logarithm of the $IC_{50}$ (M) values proportionated to the value of the reference compound DuP-753 reported by DuPont Merck ($IC_{50} = 19$ nM) in their original paper ($pIC_{50}$ proportionated or Y-2).

The dependent property values of the 28 antagonists used in the CoMFA training set are listed in Table 1. $pIC_{50}$ values, which range from approximately 4 to 8, have been obtained from six different sources.

*Molecular modelling and selection of conformations*

Starting 3D models for all compounds were generated using the model building tools available within the interactive molecular modelling system MacroModel v. 3.1X

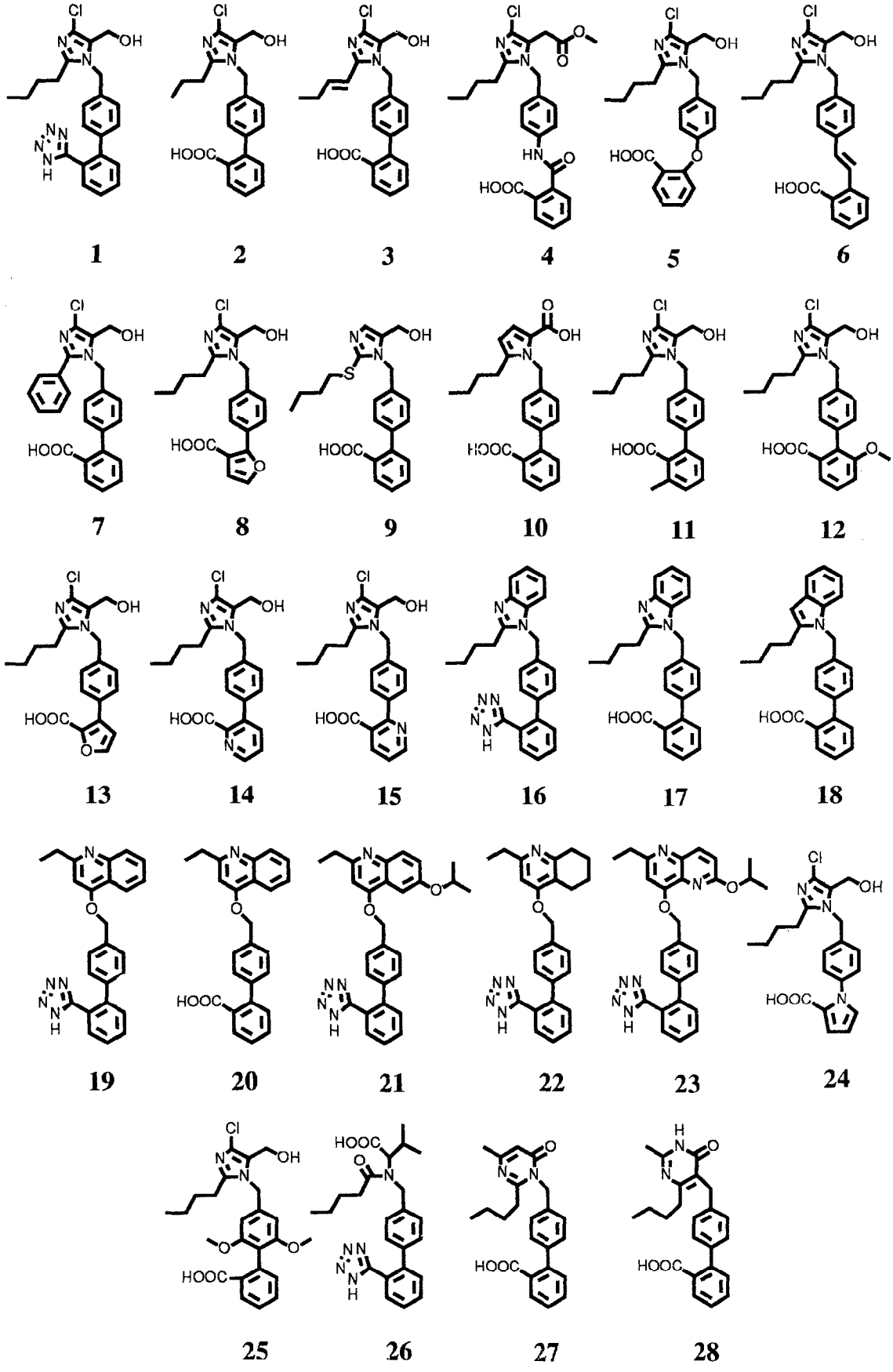



Fig. 2. Chemical structures of the 28 molecules used in the CoMFA training set.

[24]. Molecular mechanics calculations in vacuo were performed with the batch-mode program BatchMin v. 3.1, interfaced to MacroModel on a Silicon Graphics workstation.

The geometry of each molecule was minimized using the MM2* force field [25] and the TNCG energy minimization algorithm [26] as implemented in this integrated software system. A 0.01 kcal $Å^{-1}$ $mol^{-1}$ energy gradient convergence criterion and a distance-dependent dielectric ($\varepsilon = r$, where r is the distance in Å) Coulombic treatment of electrostatics were used. MacroModel MM2* partial atomic charges are derived from the original MM2 dipoles through the classical definition of a bond dipole*.

Starting from each minimized geometry, a conformational analysis was carried out, leading to the identifications of all stable conformers within 8 kcal $mol^{-1}$ of the global minimum for each compound. The torsional space of each molecule was randomly varied with the Chang–Guida–Still [27] Monte Carlo conformational search (1000 steps for variable torsion angle) as implemented in the BatchMin program. The structures thus obtained by the random conformation generator were subjected to energy minimization using the same conditions described above. The nature of the stationary points individuated was tested by computing the eigenvalues of the second-derivative matrix. The minimum-energy conformations were considered identical unless the least-squares superimposition of all heavy atoms showed one or more pairs of equivalent atoms separated by more than 0.25 Å.

Although an energy window was used, the total number of minima individuated for each compound was high (from about 100 to more than 1000). The first and probably most important step in this CoMFA study was the selection of which conformation(s) to use for each flexible molecule. The choice was made according to the interatomic distance 3D QSAR model previously derived [4].

On the basis of that work, two likely active geometries with respect to the relative spatial arrangement of the acid moiety and the heterocyclic ring were identified. One, here called **g**-type active geometry, corresponds also to the global minimum-energy conformation for the standard compound, DuP-753, as determined by MM2* implemented in MacroModel. The other putative active geometry, indicated by **x**, is compatible with the X-ray-determined structure for the same molecule.

In vacuo MM2* geometry-optimized conformations, corresponding to these two active geometries and featuring an extended lipophilic side chain, were selected for each molecule. Two different training sets were thus obtained (sets **g** and **x**, respectively) and then separately studied to investigate the dependence of the final model on the conformational preference.

TABLE 1
EXPERIMENTAL BINDING AFFINITIES OF THE 28 ANTAGONISTS TO THE $AT_1$ A II RECEPTOR

| Compound | $IC_{50}$ (nM) | Ref. | p($IC_{50}$ M) Y-1 | p($IC_{50}$ M prop.) Y-2 |
|---|---|---|---|---|
| **1** (DuP-753) | 19 | 5 | 7.72 | 7.72 |
| **2** | 160 | 5 | 6.80 | 6.80 |
| **3** | 80 | 5 | 7.10 | 7.10 |
| **4** | 140 | 13 | 6.85 | 6.85 |
| **5** | 400 | 13 | 6.40 | 6.40 |
| **6** | 5400 | 13 | 5.27 | 5.27 |
| **7** | 24 000 | 5 | 4.62 | 4.62 |
| **8** | 4200 | 5 | 5.38 | 5.38 |
| **9** | 5600[a] | 14 | 5.25 | 5.56 |
| **10** | 12 000[a] | 15 | 4.92 | 5.23 |
| **11** | 1900 | 5 | 5.72 | 5.72 |
| **12** | 3000 | 5 | 5.52 | 5.52 |
| **13** | 3600 | 16 | 5.44 | 5.44 |
| **14** | 2151 | 16 | 5.66 | 5.66 |
| **15** | 4395 | 16 | 5.35 | 5.35 |
| **16** | 96[b] | 17 | 7.02 | 6.99 |
| **17** | 2300[b] | 17 | 5.64 | 5.61 |
| **18** | 51 000[b] | 17 | 4.29 | 4.27 |
| **19** | 31[b] | 18 | 7.51 | 7.49 |
| **20** | 600[b] | 18 | 6.22 | 6.20 |
| **21** | 26[b] | 18 | 7.59 | 7.56 |
| **22** | 50[b] | 19 | 7.30 | 7.28 |
| **23** | 6[b] | 20 | 8.22 | 8.20 |
| **24** | 870[c] | 21 | 6.06 | 6.34 |
| **25** | 1900[c] | 22 | 5.72 | 6.00 |
| **26** | 9[d] | 23 | 8.05 | 8.19 |
| **27** | 454 | 16 | 6.34 | 6.34 |
| **28** | 195 | 16 | 6.71 | 6.71 |

[a] $IC_{50}$ (DuP-753) = 39 nM.
[b] $IC_{50}$ (DuP-753) = 18 nM.
[c] $IC_{50}$ (DuP-753) = 36 nM.
[d] $IC_{50}$ (DuP-753) = 26.2 nM.

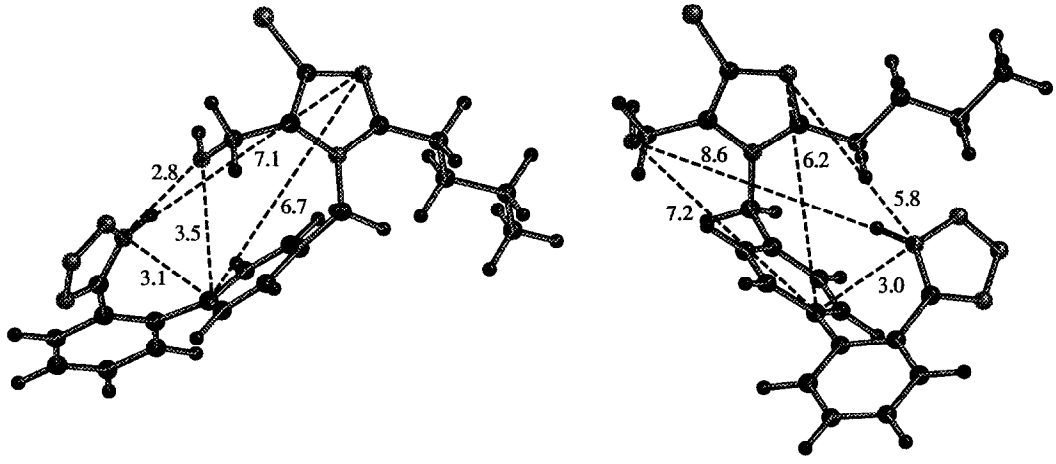
*Molecular alignment*

All the molecular structures, within each conformational series, were aligned with respect to the lead compound DuP-753, maximizing their steric and electrostatic overlap by means of the SEA molecular superimposition program [28]. SEA allows a rapid pairwise rigid comparison of different molecules employing an empirical function that comprises a double sum over all possible atom pairs and is based on a Monte Carlo routine followed by an optimization algorithm. A similarity score $A_F$ is computed as follows:

$$A_F = -\sum_i \sum_j w_{ij} \exp\{-\alpha r_{ij}\}^2 \qquad (1)$$

where $r_{ij}$ represents the distance between atom i of the

*The MM2 parameter set supplied with MacroModel (MM2*) is very similar, but not identical, to that provided by Allinger in 1987. MM2* differs from the authentic field by the use of a Coulomb's law treatment of electrostatics, a torsional barrier treatment of conjugation and an improper torsion treatment of out-of-plane bending. Included in the MacroModel field are a number of additional parameters to handle substructures other than those parameterized by Allinger.




**g conformation** **x conformation**

Fig. 3. Type **g** and **x** active conformations of compound **1** (DuP-753) selected for CoMFA studies.

first structure and atom j of the second one, and α determines the attenuation range of this distance dependence. The $w_{ij}$ pre-exponential factor is computed as a function of atomic partial charges and van der Waals radii:

$$w_{ij} = w_E\, q_i\, q_j + w_S\, v_i\, v_j \qquad (2)$$

The electrostatic and steric contributions ($w_E$ and $w_S$) were set to have equal weighting. The best match for each pair of compounds is usually characterized by the best similarity score out of all the possible alignments produced by the Monte Carlo search procedure.

To test how the results from SEA depend on different methods to calculate atomic charges, a few preliminary runs were executed by using both Gasteiger–Marsili (from SYBYL v. 6.0) [29,30] and MNDO molecular electrostatic potential fitted partial atomic charges (ESP fit routine of a modified version of MOPAC 5.0) [31,32]. Since extremely similar alignments were obtained, all subsequent SEA molecular comparisons were carried out by means of Gasteiger–Marsili empirical charges, whose calculation requires much less computational time.

*Molecular potential energies calculation*

Steric potential energies were computed by means of the Lennard-Jones 6-12 potential function within the SYBYL QSAR module, using the TRIPOS force field [30]. A probe of the size of either an $sp^3$ carbon or of a chlorine atom was used. Grid points having an energy value greater than +30 kcal $mol^{-1}$ were treated as 'sterically bad points' and set to a default value of 30.

Electrostatic potential energies were calculated with the SYBYL QSAR module using a Coulomb function with Gasteiger–Marsili partial atomic charges, a probe atom with a positive unitary charge (i.e. a proton) and a constant dielectric constant ($\varepsilon = 1$). An energy cutoff of ±30 kcal $mol^{-1}$ and the 'DROP NO' option in SYBYL (i.e. electrostatic energies at sterically bad points were computed normally) were employed. Alternatively, molecular electrostatic potential values were obtained directly by the MNDO semiempirical wave function using a modified version of MOPAC 5.0 [32]. In this case an energy cutoff of +30 kcal $mol^{-1}$ was applied.

A regularly spaced (1.0 Å) 3D grid was used, extending 4.0 Å in all directions beyond the van der Waals envelope of all molecules, thus yielding a total of 9108 ($23 \times 22 \times 18$ Å) and 9614 ($19 \times 23 \times 22$ Å) points for set **g** and set **x**, respectively.

*Chemometric analysis*

*PLS regression models* Statistical analyses of the relationships between interaction energies and the property of interest were performed by the partial least squares (PLS) algorithm with both the QSAR module of SYBYL [30] and the program GOLPE (generating optimal linear PLS estimations) [33]. The optimal number of components in each PLS model was determined by cross-validation procedures: the leave-one-out technique was invoked in SYBYL, whereas five random groups were used in GOLPE. The latter protocol was repeated up to 100 times and the associated parameters represent mean values. The predictive power of each statistical model was evaluated by means of $Q^2$ ($R^2_{CV}$ in SYBYL) and SDEP parameters computed as follows:

$$Q^2 = 1 - \sum(y_{pred} - y_{obs})^2 / \sum(y_{obs} - y_{mean})^2 \qquad (3)$$

$$SDEP = [\sum(y_{pred} - y_{obs})^2 / n]^{1/2} \qquad (4)$$

where n represents the number of compounds.

The corresponding $R^2$ and SDEC parameters, referred to calculated values ($y_{calc}$) for the molecules actually used

in modelling (instead of the predictions ($y_{pred}$) of molecules left out in cross-validation – Eqs. 3 and 4), were employed to describe the modelling or fitting power of the PLS equations.

The resulting PLS final models were also validated by a 'Y scrambling' technique: in practice, while the CoMFA X matrix remains the same, the Y vector, containing the binding data, is randomly altered so that the positions of $y_{obs}$ in the vector become scrambled. As a consequence, the $pIC_{50}$ values are no longer associated with the original structures. If the original PLS model is really sound, no PLS models should be found using a leave-one-out cross-validation procedure on the original X matrix and the new Y 'scrambled' vector. This is a reasonable and convenient scheme to establish the risk for chance correlation. It is worth pointing out that only Y scrambled vectors characterized by small correlation coefficients (typically <0.4 in absolute value) with respect to the original Y vector are employed.

*Potentials pretreatment* In order to minimize the influence of 'noise', all potential values, both steric and electrostatic, less than 0.1 kcal $mol^{-1}$ in absolute value, were set to 0 and a minimum standard deviation of either 1.0 or 2.0 kcal $mol^{-1}$ was then applied (i.e. columns having a lower standard deviation were ignored). Both steric and electrostatic columns corresponding to 'sterically bad' grid points for at least one molecule were either excluded or included by means of in-house suitably developed programs. The CoMFA standard scaling option was applied in SYBYL, whereas no scaling was performed with GOLPE.

*Molecule selection* The combined use of principal component analysis (PCA) [34] and D-optimal design [35] was applied to obtain a better balanced (or at least a less unbalanced) training set. D-optimal experimental designs can be recommended when the data set is discrete, as it happens with molecular structures and descriptions, and in particular when the variable space is not well sampled. A PCA was performed on the CoMFA data matrix (also called the X matrix) and the first few principal component scores were used by the DESDOP program to select an informative training set of molecules. The goodness of the designs is typically evaluated in terms of the determinant of the matrix |X'X|, where X represents the $n \times p$ data matrix, X' is its transpose matrix, n is the number of selected compounds and p is the number of latent variables obtained by PCA. An experimental design can be defined as D-optimal when such a determinant is maximum.

*Variable selection* The GOLPE variable selection procedures were applied to both the whole and 'experimental design-reduced' training sets. For comparison purposes, two different strategies were adopted using GOLPE: (i) a variable pre-preselection using a D-optimal design, followed by a final selection based on fractional factorial design (FFD); and (ii) a variable selection using FFD only. Using the D-optimal design the original variables were typically reduced by 50%. To perform FFD calculations, a 2:1 ratio of combinations/variables and a 3:1 ratio of true/dummy variables were adopted. SDEP values were computed by means of five random groups made in 20 different ways to achieve a good compromise between computation time and stability of statistical results. The FFD variable selection procedure works in an iterative manner, keeping fixed those variables with a positive effect on the predictivity and excluding variables with a negative effect at each iteration. At the end, variables with an uncertain effect on predictivity were excluded from the PLS regression model .

*PLS design through contour maps* The overall results of a CoMFA analysis are twofold. One outcome is a regression equation with electrostatic and steric relevant contributions. Most of the QSAR studies end at this step: the equation obtained can be employed for predicting the activity, or property value, of additional molecules unknown to the training set.

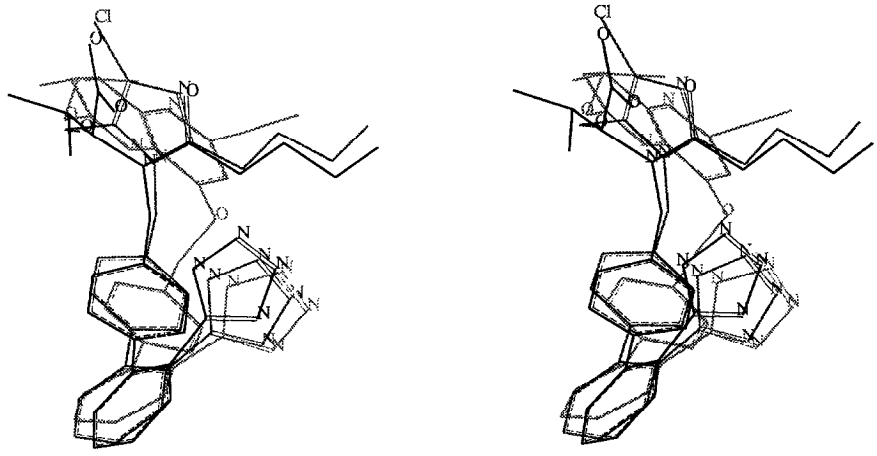


Fig. 4. Alignment of compounds **1**, **23** and **26** within set **x** after adjustment with SEA.

TABLE 2
PLS TRIALS

| Analysis | Steric probe | Electrostatic | Inside points | | SYBYL PLS[a] | | | GOLPE PLS[b] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Steric | Electrostatic | Sigma | Scaling | $R^2_{CV}$ | SDEV | Scaling | $Q^2$ |
| **Conformation g** | | | | | | | | | | |
| 1 | $Csp^3$ | MNDO MEP | In | In | 2.0 | CoMFA std | 0.48 (5)[c] | | | |
| 2 | $Csp^3$ | MNDO MEP | In | In | | | | 2.0 | None | 0.41 (3) |
| 3 | $Csp^3$ | MNDO MEP | In | Out | | | | 1.0 | None | 0.37 (5) |
| 4 | $Csp^3$ | MNDO MEP | Out | Out | | | | 1.0 | None | 0.25 (2) |
| 5 | $Csp^3$ | Coulombic | In | In | 2.0 | CoMFA std | 0.46 (3) | | | |
| 6 | $Csp^3$ | Coulombic | In | In | | | | 2.0 | None | 0.38 (3) |
| 7 | $Csp^3$ | Coulombic | In | Out | | | | 1.0 | None | 0.32 (5) |
| 8 | $Csp^3$ | Coulombic | Out | Out | | | | 1.0 | None | 0.24 (2) |
| 9 | Cl | MNDO MEP | In | In | | | | 2.0 | None | 0.40 (3) |
| **Conformation x** | | | | | | | | | | |
| 10 | $Csp^3$ | MNDO MEP | In | In | 2.0 | CoMFA std | 0.52 (3) | | | |
| 11 | $Csp^3$ | MNDO MEP | In | In | | | | 2.0 | None | 0.48 (3) |
| 12 | $Csp^3$ | MNDO MEP | In | Out | | | | 1.0 | None | 0.34 (3) |
| 13 | $Csp^3$ | MNDO MEP | Out | Out | | | | 1.0 | None | 0.20 (2) |
| 14 | $Csp^3$ | Coulombic | In | In | 2.0 | CoMFA std | 0.52 (6) | | | |
| 15 | $Csp^3$ | Coulombic | In | In | | | | 2.0 | None | 0.40 (3) |
| 16 | $Csp^3$ | Coulombic | In | Out | | | | 1.0 | None | 0.34 (3) |
| 17 | $Csp^3$ | Coulombic | Out | Out | | | | 1.0 | None | 0.26 (3) |
| 18 | Cl | MNDO MEP | In | In | | | | 2.0 | None | 0.47 (3) |

[a] Cross-validation: leave-one-out.
[b] Cross-validation: five random groups, 100 SDEP.
[c] The optimal number of components is in parentheses.

In contrast, a CoMFA QSAR is characterized by the fact that each numerical coefficient in the equation corresponds directly to a specific grid point [36]. The CoMFA PLS model can therefore be graphically represented as 3D contour maps. Such maps represent the second kind of outcome of a CoMFA study and can be of assistance to identify regions of space surrounding the molecules where steric and electrostatic differences significantly affect biological responses.

Pseudo-coefficients, potential values for each compound and the corresponding products (the so-called 'activity contributions') represent a useful visual aid that, coupled with the knowledge and expertise of the user–researcher, may lead to the design of new or modified structures.

## Results and Discussion

A training set of 28 structurally heterogeneous non-peptide A II antagonists, showing different levels of binding affinity, has been investigated by means of a CoMFA-like approach. Starting geometries, for each molecule, were selected according to two likely active conformations previously identified. Thus, two different conformational sets, here called set **g** and set **x**, were considered and separately analysed. As an example, the **g**- and, respectively, **x**-type conformations chosen for the reference compound, DuP-753, are illustrated in Fig. 3, along with the values of some relevant interatomic distances. The **g**-type conformation, stabilized by an intramolecular H-bond between the tetrazole and the alcoholic function placed sideways to the imidazole ring, corresponds to the global minimum-energy conformation, whereas the **x**-type conformation is compatible with the X-ray-determined structure for this compound.

DuP-753, within each conformational set, has been used as a template and all the molecular structures were superimposed on it by means of SEA [28] to maximize their steric and electrostatic alignments. The 3D orientations so obtained, for each one of the two conformational sets, show a good spatial correspondence between those functionalities identified as key structural elements.

As an example, the superposition used to generate the CoMFA potentials for set **x** is shown in Fig. 4 for compounds **23** and **26**, i.e. the compounds showing the highest $pIC_{50}$ values in the series. The tetrazole acid moieties placed on the biphenyl chains turn out perfectly superimposed; the imidazole ring of DuP-753 has been replaced by one of the two fused pyridine rings in compound **23**, whereas in **26** the amide oxygen is a bioisostere of the imidazole -N= type nitrogen atom. The alcoholic function of DuP-753 is close to the nitrogen atom of the second pyridine ring as well as to the isopropyloxy moiety with regard to compound **23**, whereas in **26** the same spatial region is occupied by the carboxylic group. Finally, the extended lipophilic chains of DuP-753 and **26** are well aligned, whereas in **23** the ethyl moiety still points in the same direction.

With all the molecules aligned (for each set separately), interaction potential energies were computed to explore the effect of different electrostatic treatments and the size of the steric probe. Multiple PCA/PLS analyses were carried out to find the optimal settings. In particular, the role of sterically bad points was accurately investigated. Similar conditions were applied to both conformational sets with the aim of finding out if the CoMFA approach can discriminate between the two likely active geometries previously found. We should immediately point out that we did not believe that this could be the case. At any rate, the lack of a good quantitative correlation could be tentatively attributed to the selection of a biologically meaningless conformation. On the other hand, it should also be emphasized that a good $Q^2$ value can be considered as only a necessary condition, and not a sufficient one, to put forward a 3D pharmacophore model [36]. Furthermore, molecule and variable selections were carried out to derive the final PLS models, whose reliability was tested by visually checking the coefficient contour maps.

*PLS trials*

A few cross-validated preliminary PLS runs were executed with GOLPE using both Y-1 and Y-2 (see the Methods section) dependent variable vectors together in order to select the response variable that best contributes to the structure–activity correlation. This choice was made by first looking at Y loadings for each significant component and then at SDEP and $Q^2$ values (data not shown). No significant differences were detected between the Y-1 and Y-2 vectors; however, since the former is characterized by slightly higher loading values, all subsequent PLS calculations were performed by taking only Y-1 into account.

$Q^2$ and the optimal number of components obtained for both set **g** and set **x** are reported in Table 2, where each row represents a different PLS run. Both CoMFA and GOLPE cross-validated results are shown in the same table, although a direct comparison between the two programs is not very significant since the cross-validation techniques applied are quite different. As an example, the CoMFA cross-validation parameters, obtained with a leave-one-out procedure, simply appear to be 'numerically' better (but 'in practice' they are not) than the GOLPE ones with random groups (see analyses 1 versus 2, 5 versus 6, 10 versus 11 and 14 versus 15).

The following alternative solutions have been adopted with regard to sterically bad points: (i) both steric and electrostatic energy potentials included; (ii) steric potential included whereas electrostatic potential excluded; and (iii) both potentials excluded. In the first case, a minimum standard deviation of 2 kcal $mol^{-1}$ was applied, reducing the number of effective variables from about 20 000 down to two or three thousand, with an almost unitary ratio between steric and electrostatic descriptors. In the second

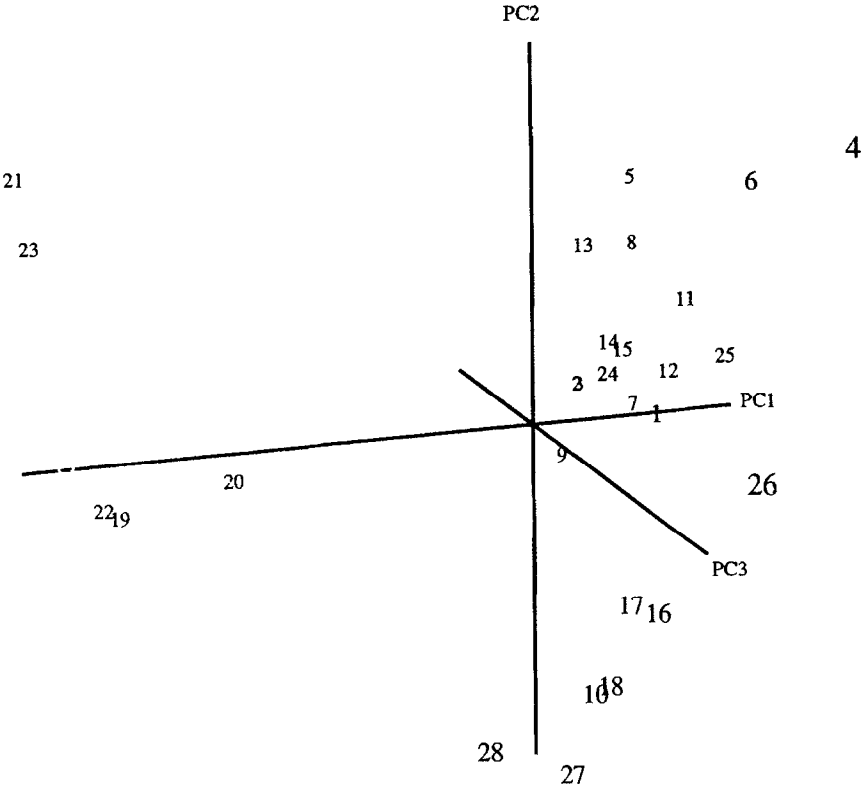



Fig. 5. PCA score plot for the 28-molecule training set (analysis 11 of Table 2).

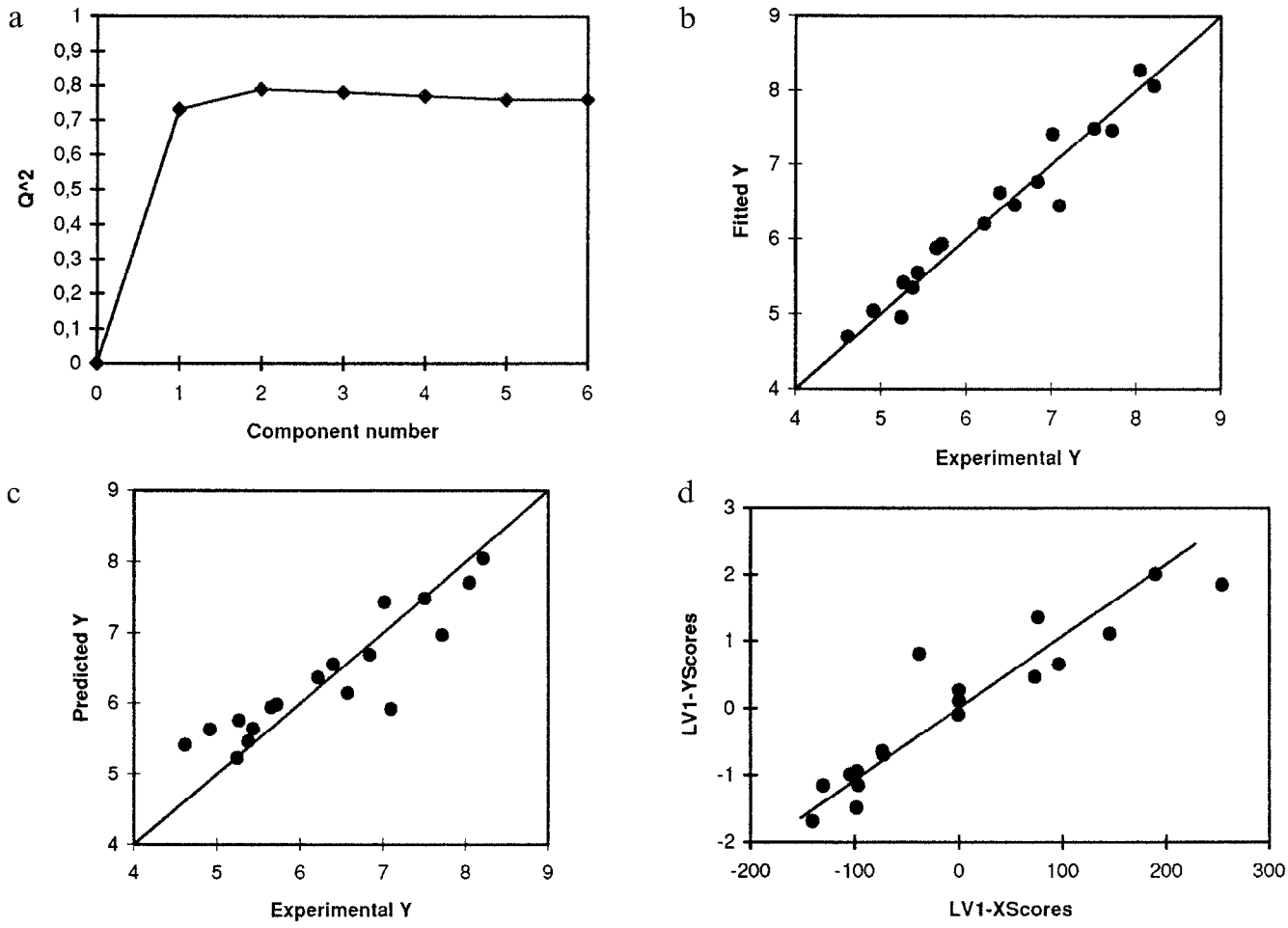



Fig. 6. Final PLS model derived from the 18-molecule training set after variable selection (optimal plus FFD). (a) $Q^2$ values; (b) fitted versus experimental $pIC_{50}$ values (comp.: 2, SDEC = 0.24, $R^2 = 0.95$); (c) predicted versus experimental $pIC_{50}$ values (comp.: 2, SDEP = 0.5, $Q^2 = 0.79$); (d) PLS X–Y score plot, comp.: 1.

and third cases, a lower minimum standard deviation, 1 kcal $mol^{-1}$, was applied to keep that ratio constant. The reported results show that the highest $Q^2$ values are achieved by including both the energy potentials, whereas the lowest ones are obtained by excluding both of them. The trend appears to be independent of the nature of the electrostatic potential (analyses 2,3,4 and 6,7,8) and of the specific conformational set (analyses 11,12,13 and 15,16,17) analysed. In addition, the major drop in terms of $Q^2$ is due to the exclusion of steric potential energies: it may be plausible that, since the Lennard-Jones potential function is highly distance-dependent, a grid stepsize of 1 Å is still too large to adequately sample the steric potential and, consequently, the inclusion of sterically bad points becomes necessary to improve the steric description. By the same token, it should be said that stepsizes smaller than 1 Å would be extremely costly in terms of computational time.

Only slight improvements were obtained on moving from a Coulombic description to MEP values computed at the MNDO level. Anyway, looking at the results from both CoMFA (see analyses 1,5 and 10,14) and GOLPE (see analyses 2,6 and 11,15), the best $Q^2$ values, within each conformational set, were achieved by using the latter, theoretically superior, electrostatic description. Finally, analyses 2,9 and 11,18 of Table 2 clearly show that a larger steric probe, e.g. a chlorine atom instead of a $Csp^3$, does not produce any significant change.

Therefore, on the basis of these preliminary PLS trials, the following preferred settings and conditions have been identified: $Csp^3$ steric probe, MNDO electrostatic potential values and inclusion of sterically bad points. As expected, the PLS runs reported in Table 2 do not allow a reasonable discrimination between the two proposed pharmacophore models. Nevertheless, any further calculations were done only on set **x**, which is characterized by the highest $Q^2$ values, both considering CoMFA and GOLPE results.

*Exploratory data analysis (PCA)*

Before performing a PLS run, we always analysed the nature of the CoMFA X data matrix resulting from speci-

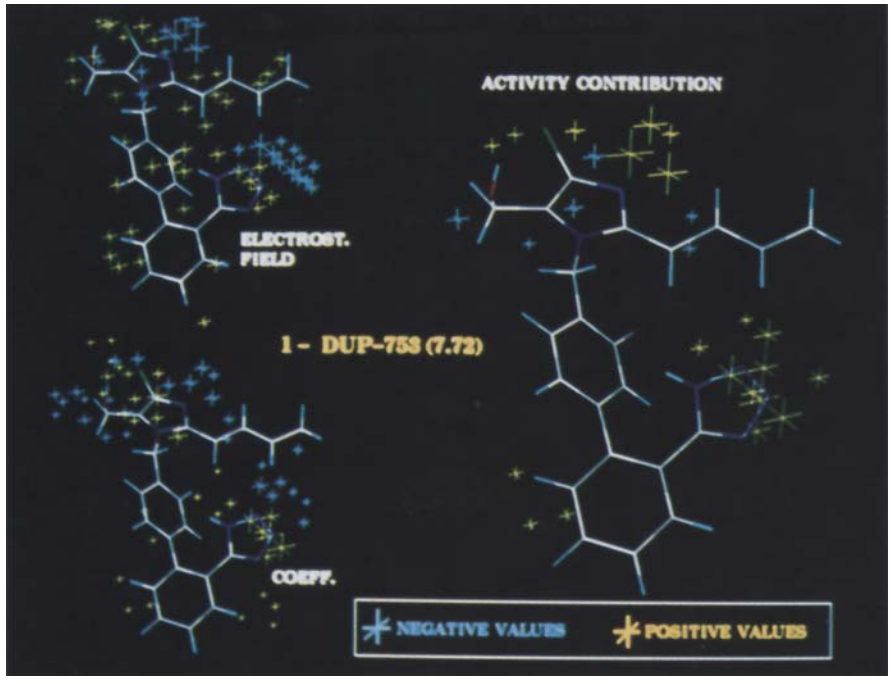



Fig. 7. Electrostatic potential, pseudo-coefficient and activity contribution maps for compound **1**. The symbol colours, for all grid plots, are light-blue for the negative values and yellow for the positive ones. The parameters are visualized in terms of crosses, whose magnitude depends on the absolute numerical values. Only parameters with an absolute value higher than a cutoff value are visualized.

fic settings and conditions. The distribution of steric and electrostatic values (X distribution) and their standard deviations were visualized by means of the GOLPE graphic tools. A principal component analysis (PCA) was also carried out within GOLPE on each X matrix, in order to get an idea of the space of the objects (i.e. molecules) corresponding to different operative conditions.

PCA enables one to view most of the information with just a few latent variables or principal components (PCs). This is particularly suited when the X matrix contains a large number of columns as in the case of multivariate characterization employing CoMFA or CoMFA-like molecular descriptors. The first few PCs (usually three or four, explaining about 45–60% of the total variance of the

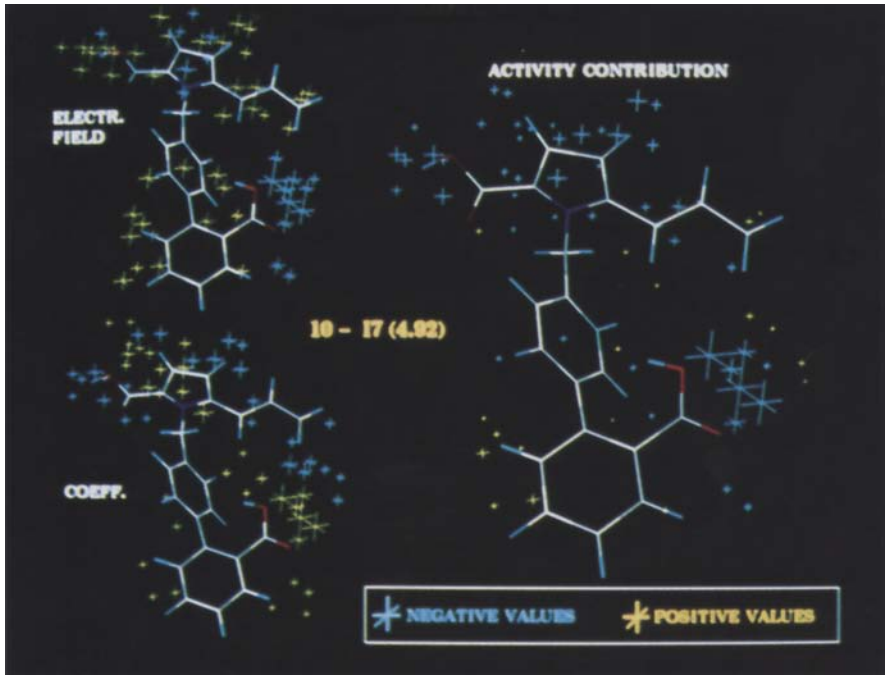



Fig. 8. Electrostatic potential, pseudo-coefficient and activity contribution maps for compound **10**. Graphical details as in Fig. 7.

TABLE 3
PLS ANALYSIS ON THE 18-MOLECULE TRAINING SET (CONFORMATION x)

| Analysis | Steric probe | Electrostatic | Inside points | | GOLPE PLS[a] | | |
|---|---|---|---|---|---|---|---|
| | | | Steric | Electrostatic | SDEV | Scaling | $Q^2$ |
| 19 | $Csp^3$ | MNDO MEP | In | In | 2.0 | None | 0.28 (2)[b] |

[a] Cross-validation: five random groups, 100 SDEP.
[b] The optimal number of components is in parentheses.

data matrix) were, in each case, extracted from the steric and electrostatic potentials, and a 3D plot of the first three PCs scores, at least, was generated to visually analyse the distribution of the objects.

The projection of the molecules in the space of the first three PCs enabled us to realize that the starting training set is not well distributed. As an example, the score plot relative to the optimal parameter settings (analysis 11 of Table 2) is reported in Fig. 5. Looking at this plot for the 28-compound training set, it can be seen that the first component separates the five compounds on the left from all the others and that most of the PCs space is not sampled.

A close inspection of the score plot in Fig. 5 enabled us to understand why some compounds are closer in space than others, and to ascertain the relevance of common structural features. For instance, all molecules close to compound **1** (DuP-753) share the biphenyl moiety and the imidazole ring. Compounds **8** and **13** have a tetrahydrofuran ring. Compounds **4**, **5** and **6** have a *trans*-amide, an oxygen atom and a trans double bond, respectively, to bridge the two aromatic rings. Compound **26** has an open chain and the amide oxygen is a bioisostere for the imidazole -N= type nitrogen atom. Compound **17** looks like **16** with a carboxylic function instead of a tetrazole. Compounds **10** and **18** lack the same pharmacophoric nitrogen atom and compounds **27** and **28** have a pyrimidinone ring instead of the imidazole one. Compounds **21** and **23** are very much alike, the only difference being a carbon atom instead of a nitrogen one. Finally, compound **19** is like **21** without the *i*-propyloxy chain, while **22** is like **19** but with a partially saturated ring.

*Molecule selection*

The first three latent variables extracted by PCA from the X matrix of analysis 11 (Table 2) were employed to carry out a D-optimal experimental design. One of the reasons for using a design strategy is to obtain a well-balanced or at least a better-balanced set compared to the original training set. Another reason is to verify if it is possible to get similar or possibly better PLS models with a smaller training set. Several applications of experimental design in the parametric space of latent variables generated by PCA have been reported [12] to solve the crucial problem of selecting a balanced and informative set of molecules for SAR studies.

The DESDOP calculations were repeated several times, each time starting from randomly selected initial designs: very similar values of the determinant were obtained. The lead compound DuP-753 was always constrained in the design. By D-optimal design, 18 out of the original 28 compounds were selected. The 10 leftover molecules are compounds **2**, **12**, **15**, **17**, **18**, **21**, **22**, **24**, **25** and **28**, which could be used as an external set, although we fully agree that this would be a biased set and anyway not a real test set.

A PLS analysis was then performed on this reduced 18-molecule training set using GOLPE and the parameter settings previously chosen ($Csp^3$ steric probe, MNDO electrostatic potential values, inclusion of sterically bad points and type **x** active conformation). The corresponding results are reported in Table 3. A reduction of the $Q^2$ value with respect to that obtained in the corresponding analysis for the original 28-molecule training set can be observed: this is expected of a smaller and better-balanced set when a cross-validation procedure with random groups is employed.

*Variable selection and final PLS models*

The GOLPE variable selection procedure (D-optimal plus fractional factorial design) was applied to both the original 28-molecule and the DESDOP-reduced 18-molecule training sets (X matrices of analyses 11 and 19, respectively). The cross-validated results are reported in Table 4.

Similar trends were obtained for both analyses: the initial 19 228 descriptor variables were reduced, step by step, to 469 and 444, respectively, while the $Q^2$ values were increased to 0.80 and 0.79 for the whole and reduced sets, respectively. The large improvement obtained by applying the variable selection procedure can be especially appreciated by looking at the SDEP parameter, which is directly related to the uncertainty of predictions. So, for the 18-molecule training set, the SDEP value decreased by almost 50%, passing from 0.92 to 0.50.

The final $Q^2$ and SDEP values obtained for the reduced set are comparable to those from the analysis carried out on the whole original set. Thus, as pointed out from the PCA score plot of Fig. 5, the 28-molecule training set is redundant (i.e. several clusters of 3–4 molecules can be detected), and the information provided by the 18 molecules chosen by DESDOP can be successively correlated with the biological response.

The scatterplot of $Q^2$ values versus the number of components corresponding to the PLS-variable selection analysis derived from the DESDOP-reduced training set is illustrated in Fig. 6a. Fitted and predicted $pIC_{50}$ versus experimental data are shown in Figs. 6b and c, respectively. Finally, the X–Y plot, i.e. scores of predictor matrix versus scores of response matrix along the first PLS component (since only Y-1 has been shown, the Y scores correspond to the centred activity data), is represented in Fig. 6d. From this plot no high-leverage points or isolated groups of objects can be detected; moreover, Fig. 6d outlines a linear correlation between binding data and predictors.

For comparison purposes, an additional PLS-variable selection analysis was carried out on the reduced training set (X matrix of analysis 19) by executing the FFD step only. As is evident from the values in Table 4, the FFD selection clearly represents the crucial step in order to improve $Q^2$ and SDEP, while no significant improvements can be detected after the D-optimal preselection. This is expected since the D-optimal criterion is applied in order to retain all relevant information while employing a much smaller number of variables, making the subsequent FFD step much less time-consuming. In other words, during the D-optimal preselection only variables characterized by high loadings, or partial weights, are kept. However, it cannot be excluded that variables held out on the basis of the D-optimality criterion could play a role when searching for a correlation with the biological response. Moreover, the D-optimal algorithm is susceptible to converging to a local maximum, and repeating the whole procedure on the same data set would not yield exactly the same results. For these and other reasons that will be discussed later, the use of the D-optimal variable preselection is still in debate and, as outlined by the authors of GOLPE, the procedure still needs to be refined. When using GOLPE, it is currently recommended to employ FFD directly whenever possible.

The results obtained are reported in case (3) of Table 4. The number of final retained variables is now slightly higher, 679 instead of 444, whereas, on the basis of $Q^2$ and SDEP values, no relevant differences can be detected by applying the FFD selection only. In order to assess the likelihood of chance correlation, new biological response Y scrambled vectors were generated as previously described. No PLS models were found using a leave-one-out cross-validation procedure: the 'best' model found has a $Q^2$ of 0.15.

*PLS design*

After the final PLS regression models were derived through variable selection and cross-validation in GOLPE, 3D visualization tools were employed to further investigate SARs and to assess the design capabilities of our CoMFA-like approach. In addition to steric or electrostatic 3D contour maps and pseudo-coefficient maps, it is also possible to generate so-called activity contribution maps. These appear to be quite useful since they allow one to identify, for a given compound and a specific potential, the regions of space associated with positive or negative contributions to the desired biological response. Activity contributions can be easily obtained by multiplying, at each grid node, the steric or electrostatic potential value by its corresponding pseudo-coefficient (i.e. the PLS regression coefficient): the graphical output highlights the regions of the lattice most directly linked to the binding affinity for the compound under investigation [33].

It is worth recalling that, as stated by Clementi et al. [11], 'when transforming the PLS loadings into the pseudo-coefficients of the polynomial in the original variables, it is not appropriate to give too much credit to the indi-

TABLE 4
GOLPE-PLS MODELS[a]: EFFECTS OF VARIABLE SELECTION

| Action | No. var. | No. comp. | $Q^2$ | SDEP | $R^2$ | SDEC |
|---|---|---|---|---|---|---|
| **(1) 28-molecule training set, settings of analysis 11** | | | | | | |
| | 19 228 | | | | | |
| SDEV = 2 | 3222 | 3 | 0.48 | 0.75 | 0.93 | 0.28 |
| D-optimal design | 1611 | 3 | 0.54 | 0.71 | 0.93 | 0.28 |
| FFD | 469 | 2 | 0.80 | 0.47 | 0.92 | 0.29 |
| **(2) 18-molecule training set, settings of analysis 19** | | | | | | |
| | 19 228 | | | | | |
| SDEV = 2 | 3050 | 2 | 0.28 | 0.92 | 0.87 | 0.38 |
| D-optimal design | 1525 | 2 | 0.32 | 0.89 | 0.87 | 0.38 |
| FFD | 444 | 2 | 0.79 | 0.50 | 0.95 | 0.24 |
| **(3) 18-molecule training set, settings of analysis 19** | | | | | | |
| | 19 228 | | | | | |
| SDEV = 2 | 3050 | 2 | 0.28 | 0.92 | 0.87 | 0.38 |
| FFD | 679 | 2 | 0.78 | 0.51 | 0.94 | 0.26 |

[a] Cross-validation: five random groups, 100 SDEP.

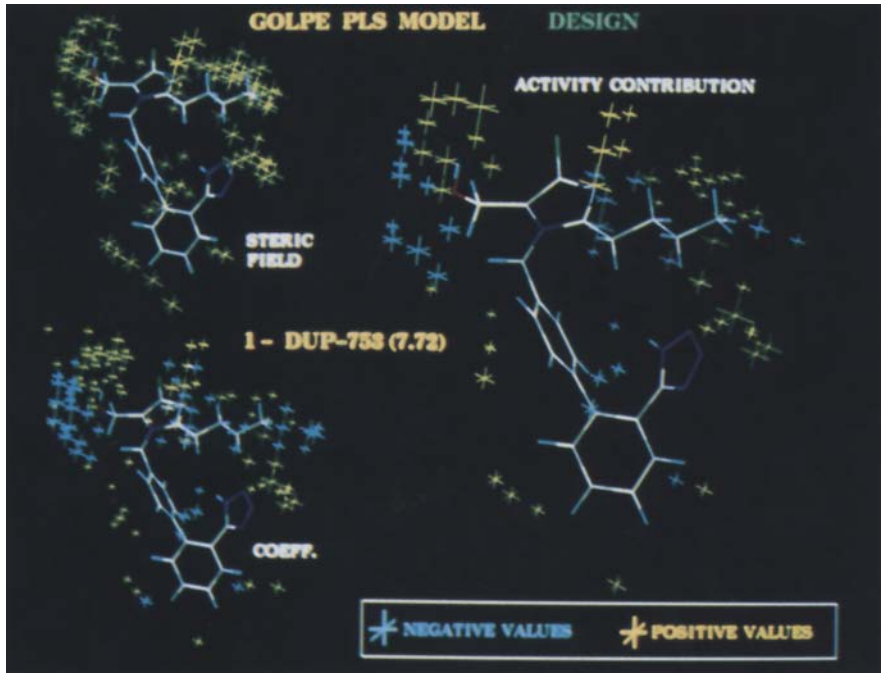



Fig. 9. Steric potential, pseudo-coefficient and activity contribution maps for compound **1**. Graphical details as in Fig. 7.

vidual coefficients: they suffer from the same instability as if they were computed directly by MLR' (multilinear regression analysis).

The pseudo-coefficient and activity contribution grid plots shown and discussed in the paper are referred to the final PLS model derived through variable selection from the starting 28-molecule training set (case (1) of Table 4), but analogous considerations can be drawn from the DESDOP-reduced 18-molecule training set. For the sake of precision, the model derived through the two-step variable selection procedure (D-optimal plus fractional factorial design) from the latter set (case (2) of Table 4) produces electrostatic grid plots slightly different from the corresponding ones of the starting training set. In particular, some electrostatic pseudo-coefficients in relevant grid points disappear (or become smaller) on passing from the

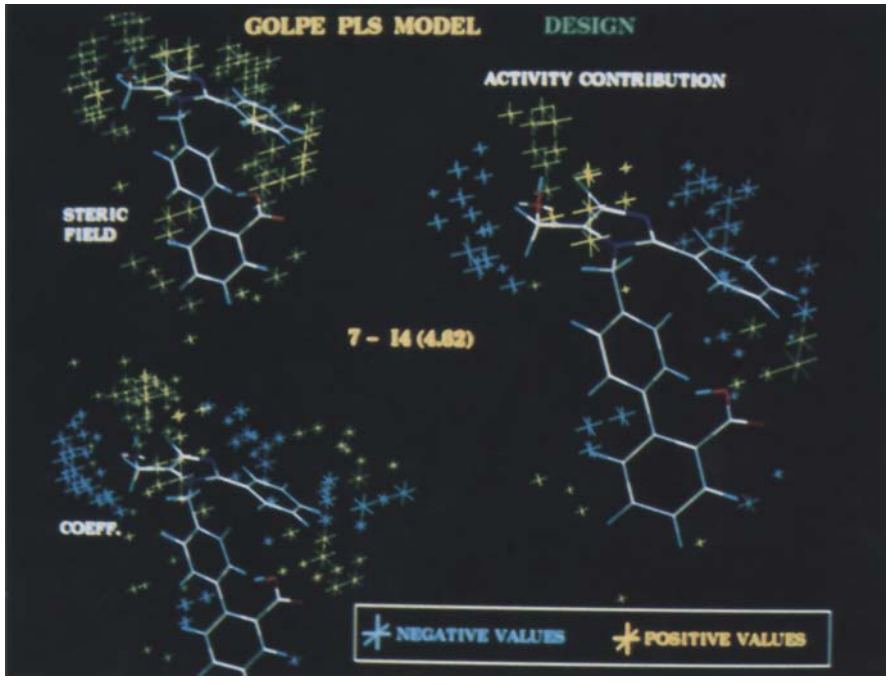



Fig. 10. Steric potential, pseudo-coefficient and activity contribution maps for compound **7**. Graphical details as in Fig. 7.

map of the original set to that of the reduced one, although no significative differences can be detected in $Q^2$ and SDEP values.

On the contrary, the model derived from the 18-molecule training set employing FFD directly (case (3) of Table 4) gives grid plots, as well as $Q^2$ and SDEP values, in very good agreement with those obtained from the original set. This fact could question the usefulness and the opportunity of D-optimal variable preselection, especially when the training set is small.

PLS electrostatic design is presented by means of compounds **1** and **10** (Figs. 7 and 8). On the top left side of each figure, the electrostatic potential for the compound under examination is shown, while the bottom left presents the pseudo-coefficients from the PLS model, although identical. The activity contributions are shown on the right side: yellow means that a given feature increases the $pIC_{50}$ values and blue means the opposite.

The examination of the electrostatic pseudo-coefficient map supports the known SARs and the conclusions derived from the visual comparative analysis of MEP distributions previously performed [7,8]. Electron-rich groups at the acid site and the H-bond acceptor moieties, and positive MEP values in the regions surrounding the lipophilic side chain (the absolute value of the relative positive pseudo-coefficients is smaller than the cutoff adopted) and the terminal aromatic ring are predicted to increase the potency.

It is worth noting how the requirement of negative electrostatic potential around the heterocyclic proton acceptor group gives rise to a negative region in the activity contribution map of the low affinity A II receptor antagonist **10** (Fig. 8), which, featuring a pyrrole ring, lacks this key structural element. The presence of large positive coefficients in correspondence to some atoms of the tetrazole ring and the consequent requirement of positive electrostatic potential in this region are the result of the major potency of the tetrazolyl analogues with respect to the corresponding carboxylic acids.

PLS steric design is exemplified by means of compounds **1** and **7** (Figs. 9 and 10). Negative, blue crosses in the steric pseudo-coefficient maps indicate steric boundaries for the lipophilic side chain, and below and at the left side of the site corresponding to the hydroxymethyl group of DuP-753, while yellow crosses indicate that the region above the chlorine and the hydroxymethyl groups of DuP-753 needs bulk for optimum affinity. The steric wall for the lipophilic side chain can be clearly shown by the activity contribution maps of compounds differing only in this position: an increase of the negative region around this site can be observed together with bulk addition and a decrease of binding affinity. The example of the phenyl substituent in the inactive compound **7** is reported (Fig. 10). By looking at the sign and size of the potentials and coefficients, one may get useful information for designing new compounds, although the limit due to the structural area sampled by the training set should always be taken into account.

The features of the CoMFA 3D contour plots derived in this study are consistent with those obtained by other authors for a different training set of A II antagonists based on 'Merck' heterocycles [6]. In the present work no efforts have been made to design new compounds on the basis of the PLS models developed so far, as the main goals were primarily dictated by an interest in methodologies and an understanding of drug–receptor interactions. However, a design phase would be essential to construct a 'true' test set of new derivatives to be used for the validation of the predictive power of the model.

## Conclusions

CoMFA descriptors and approaches were applied to study 3D QSARs in a series of structurally different and conformationally flexible non-peptide A II antagonists using binding affinities to the A II receptor obtained from the literature. Most of the critical methodological aspects of a CoMFA analysis have been dealt with and, in some cases, alternative solutions were adopted for comparison purposes. The most crucial aspect improving the quality of the models resulted from the variable selection procedure in the GOLPE-PLS analysis. Potentials pretreatment also appeared to play a relevant role: better results (the highest $Q^2$ values) were obtained by including both steric and electrostatic columns corresponding to sterically bad points. The conformation used, the nature of the electrostatic field and the type of steric probe appeared to be less important aspects, since they affect the models to a much smaller extent.

The overall PLS procedure suggested in this study can be summarized as follows: (i) PCA on the data matrix; (ii) object selection by D-optimal design; (iii) variable selection by fractional factorial design; (iv) cross-validation with five random groups and 100 SDEP calculations; (v) final model and inspection of the X–Y plot; and (vi) Y scrambling. In particular, object selection proved to be a very useful tool that allowed one to define a reduced, better-balanced and more statistically informative training set. It can be recommended each time a set of molecules is not or cannot be planned by experimental design techniques.

Variable selection is, as expected, a crucial step in any CoMFA-PLS approach, given the size of the data matrix with thousands of variables left even after different filtering and 'noise' reduction procedures. Although similar results were obtained (in terms of $Q^2$ and SDEP values) both with and without a D-optimal preselection of variables, it is now generally recommended to skip the D-optimal procedure and directly perform the FFD variable selection. Finally, a graphical examination of the results

through pseudo-coefficients and activity contribution maps leads to a better understanding of the molecular requirements for optimal drug–receptor interactions and offers a useful visual aid for designing new potentially interesting structures.

## Acknowledgements

We wish to thank Professor Sergio Clementi and his group for helpful discussions, ideas and suggestions during the work, and Professor W. Clark Still for his permission to use MacroModel in this research. The Istituto LusoFarmaco d'Italia (Menarini group) is gratefully acknowledged for the scholarship granted to L.B. The work was financially supported by a European Community Grant (Human Capital and Mobility Program-A2, Contract No. ERBCHRXCT 920027).

## References

1 Reid, J.L. and Rubin, P.C., Physiol. Rev., 67 (1987) 725.
2 Valloton, M.B., Trends Pharmacol. Sci., 8 (1987) 69.
3 Greenlee, W.J. and Mantlo, N.B. (Eds.) Recent Advances in the Design and Synthesis of Angiotensin II Receptor Antagonists, Bioorg. Med. Chem. Lett., 4 (1994), Symposia-in-Print 9.
4 Belvisi, L., Bravi, G., Scolastico, C., Vulpetti, A., Salimbeni, A. and Todeschini, R., J. Comput.-Aided Mol. Design, 8 (1994) 211.
5 Carini, D.J., Duncia, J.V., Aldrich, P.E., Chiu, A.T., Johnson, A.L., Pierce, M.E., Price, W.A., Santella III, J.B., Wells, G.J., Wexler, R.R., Wong, P.C., Yoo, S.E. and Timmermans, P.B.M.W.M., J. Med. Chem., 34 (1991) 2525.
6 Prendergast, K., Adams, K., Greenlee, W.J., Nachbar, R.B., Patchett, A.A. and Underwood, D.J., J. Comput.-Aided Mol. Design, 8 (1994) 491.
7 Belvisi, L., Bonati, L., Bravi, G., Pitea, D., Scolastico, C. and Vulpetti, A., J. Mol. Struct. (THEOCHEM), 281 (1993) 237.
8 Salimbeni, A., Canevotti, R., Paleari, F., Bonaccorsi, F., Renzetti, A.R., Belvisi, L., Bravi, G. and Scolastico, C., J. Med. Chem., 37 (1994) 3928.
9 Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
10 Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, The Netherlands, 1993.
11 Cruciani, G., Clementi, S. and Baroni, M., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, The Netherlands, 1993, pp. 551–564.
12 Clementi, S., Cruciani, G., Baroni, M. and Costantino, G., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, The Netherlands, 1993, pp. 567–582.
13 Carini, D.J., Duncia, J.V., Johnson, A.L., Chiu, A.T., Price, W.A., Wong, P.C. and Timmermans, P.B.M.W.M., J. Med. Chem., 33 (1990) 1330.
14 Carini, D.J. and Duncia, J.V., Eur. Pat. 253 310, 1988 (Chem. Abstr. 109, 129008g, 1988).
15 Carini, D.J., Wells, G.J. and Duncia, J.V., Eur. Pat. 323 841, 1989 (Chem. Abstr. 112, 55881z, 1990).
16 Unpublished results of LusoFarmaco Laboratories, related to A. Menarini – Industrie Farmaceutiche Riunite.
17 Thomas, A.P., Allott, C.P., Gibson, K.H., Major, J.S., Masek, B.B., Oldham, A.A., Ratcliffe, A.H., Roberts, D.A., Russell, S.T. and Thomason, D.A., J. Med. Chem., 35 (1992) 877.
18 Bradbury, R.H., Allott, C.P., Dennis, M., Fisher, E., Major, J.S., Masek, B.B., Oldham, A.A., Pearce, R.J., Rankine, N., Revill, J.M., Roberts, D.A. and Russell, S.T., J. Med. Chem., 35 (1992) 4027.
19 Bradbury, R.H., Allott, C.P., Dennis, M., Girwood, J.A., Kenny, P.W., Major, J.S., Oldham, A.A., Ratcliffe, A.H., Rivett, J.E., Roberts, D.A. and Robins, P.J., J. Med. Chem., 36 (1993) 1245.
20 Bradbury, R.H., Edwards, M.P., Luke, R.W.A., Pearce, R.J., Roberts, D.A., Major, J.S. and Oldham, A.A., 202nd ACS Meeting, New York, NY, U.S.A., 25–30 August, 1991, Abstr. No. 102.
21 Bovy, P.R., Reitz, D.B., Collins, J.T., Chamberlain, T.S., Olins, G.M., Corpus, V.M., McMahon, E.G., Palomo, M.A., Koepke, J.P., Smits, G.J., McGraw, D.E. and Gaw, J.F., J. Med. Chem., 36 (1993) 101.
22 Bovy, P.R., Collins, J.T., Olins, G.M., McMahon, E.G. and Hutton, W.C., J. Med. Chem., 34 (1991) 2410.
23 Bühlmayer, P., Ostermayer, F. and Schmidlin, T., Eur. Pat. 443 983, 1991 (Chem. Abstr. 116, 151772t, 1991).
24 Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R., Caufield, C., Chang, G., Hendrickson, T. and Still, W.C., J. Comput. Chem., 11 (1990) 440.
25 Burkert, U. and Allinger, N.L., Molecular Mechanics, ACS Monograph Series, Vol. 177, American Chemical Society, Washington, DC, U.S.A., 1982.
26 Ponder, W.J. and Richards, F.M., J. Comput. Chem., 8 (1987) 116.
27 Chang, G., Guida, W.C. and Still, W.C., J. Am. Chem. Soc., 111 (1989) 4379.
28 Smith, G.A., Steric and Electrostatic Alignment Molecular Superposition Program, Merck Sharp & Dohme Research Laboratories, West Point, PA, U.S.A. Program obtained through Quantum Chemistry Program Exchange, QCPE 567, University of Indiana, Bloomington, IN, U.S.A.
29 a. Gasteiger, J. and Marsili, M., Tetrahedron, 36 (1980) 3219.
b. Marsili, M. and Gasteiger, J., Croat. Chem. Acta, 53 (1980) 601.
30 SYBYL, v. 6.0, Tripos Associates, St. Louis, MO, U.S.A., 1993.
31 Besler, B.H., Merz, K.M. and Kollman Jr., P.A., J. Comput. Chem., 11 (1990) 431.
32 Merz, K.M. and Besler, B.H., QCPE Bull., 10 (1990) 15.
33 Baroni, M., Costantino, G., Cruciani, G., Riganelli, D., Valigi, R. and Clementi, S., Quant. Struct.–Act. Relatsh., 12 (1993) 9.
34 Joliffe, I.T., Principal Component Analysis, Springer, New York, NY, U.S.A., 1986.
35 Baroni, M., Clementi, S., Cruciani, G., Kettaneh-Wold, N. and Wold, S., Quant. Struct.–Act. Relatsh., 12 (1993) 225.
36 Cramer III, R.D., DePriest, S.A., Patterson, D.E. and Hecht, P., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, The Netherlands, 1993, pp. 443–485.