

Autocorrelation of Molecular Electrostatic Potential Surface Properties Combined with Partial Least Squares Analysis as Alternative Attractive Tool to Generate Ligand-Based 3D-QSARs

Stefano Moro^{1*}, Magdalena Bacilieri¹, Cristina Ferrari¹, Giampiero Spalluto²

¹Molecular Modeling Section, Dipartimento di Scienze Farmaceutiche, Università di Padova, Via Marzolo 5, I-35131 Padova, Italy; ²Dipartimento di Scienze Farmaceutiche, Università degli Studi di Trieste, Piazzale Europa 1, I-34127 Trieste, Italy

Abstract: A database of 106 human A₃ adenosine receptor antagonists was used to derive two alternative PLS models: one starting from CoMFA descriptors and the other starting from the autocorrelation descriptors. The peculiarity of this work is the introduction of autocorrelation vectors as molecular descriptors for the PLS analysis. The autocorrelation allows comparing molecules (and their properties) with different structures and with different spatial orientation without any previous alignment. In particular, Molecular Electrostatic Potential (MEP) was the property computed and its information encoded in autocorrelation vectors. The 3D spatial distribution and the values of the electrostatic potential is in fact largely responsible for the binding of a substrate to its receptor binding site. Validation was done with an external test set and the results of the two models were compared. Interestingly, our preliminary results seem to indicate that this new alternative approach could robustly compete with the already well consolidated CoMFA approach. In particular, we have suggested that it could be a very interesting tool to filter large structural database in several virtual screening applications.

Keywords: 3D-QSAR, Molecular Electrostatic Potential (MEP), Comparative Molecular Field Analysis (CoMFA), Autocorrelation Vectors, Partial Least Squares (PLS); A₃ Adenosine Receptor Antagonists.

INTRODUCTION

It is generally accepted that receptor and corresponding ligands recognize each other at their molecular surfaces. Therefore, the binding force of a receptor-ligand complex depends on the shape of the substrate surface and on the distribution of certain properties on this surface. Any method attempting to model biological activity should take into account this information and try to correlate it to biological activity. This is the major objective of all three-dimensional quantitative structure-activity relationship approaches (3D-QSAR) [1].

Rational drug design with 3D-QSAR comprises several subsequent steps: conformational analyses, alignment of the molecules, generation of molecular descriptors and regression analysis. Optionally, one or more biological response(s) can be used as the independent variable(s) [1]. Since multicollinearity among the descriptor variables may affect the regression analysis detrimentally, Partial Least Square (PLS) algorithm is frequently used as the regression method in 3D-QSAR [2]. A very well known and consolidated 3D-QSAR approach is the Comparative Molecular Field Analysis (CoMFA) presented by Cramer *et al.* in 1988 [3]. Following CoMFA methodology, low energy

conformations of the molecules are aligned by superimposition of mutual and possible interaction points (steric and electrostatic) with the target receptor protein. Molecular superimposition is by far the most crucial step in order to achieve reliable 3D-QSAR models [3].

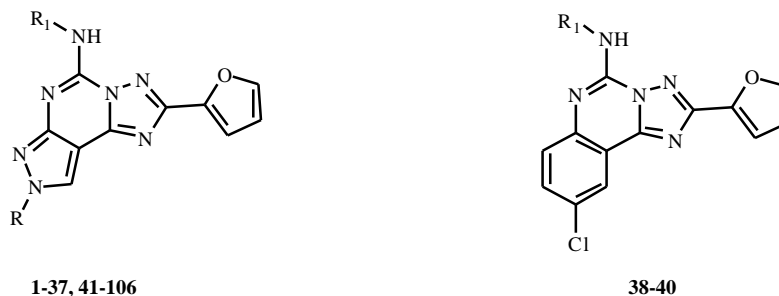
Few years ago, Gasteiger and collaborators investigated the molecular electrostatic potential (MEP) on a molecular surface as particularly useful method for rationalizing the interactions between molecules and molecular recognition processes [4]. In fact, values and spatial distribution of MEP are in fact responsible of the chemical behavior of an agent in a chemical reaction, as well as they strongly influence the binding of a substrate to its active site. However, the MEPs property values, as well as other molecular variables, strongly depend on the spatial orientation of the different molecules. It is not possible therefore to compare the property of a set of compounds without a previous alignment. The introduction of the autocorrelation vector allows to overcome this in convenience, making the MEP information invariant to the spatial rotation and translation of molecules [5]. Even if autocorrelation is potentially useful in medicinal chemistry and in particular in developing QSAR models, nowadays only few interesting studies have been reported [6-9]. The major conceptual limitation of using autocorrelation vectors is due to the difficulty of tracing back the obtained correlation coefficients into the starting 3D structure properties, and, consequently, it makes the chemical interpretation of the QSAR model, less convincing [6-9].

*Address correspondence to this author at the Molecular Modeling Section, Department of Pharmaceutical Sciences, University of Padova, Via Marzolo, 5 - 35100 Padova (ITALY); Tel: +39 049 8275704; Fax: +39 049 827 5366; E-mail: stefano.moro@unipd.it

In the present paper, we describe an interesting application of MEP autocorrelation vectors as molecular descriptors for the PLS analysis. A database of 106 human A₃ adenosine receptor antagonists (see Table 1) was used to

derive two alternative PLS models: one starting from CoMFA descriptors and the other starting from the autocorrelation descriptors.

Table 1. Structures and human A₃ (calculated and predicted by both 3D-QSAR models) binding affinities of new synthesized compounds (1-106).



#	R	R ¹	Observed (pK _i , nM)	Predicted (αMEP/PLS)	Predicted (CoMFA)
1	H	H	-2,54	-2.20	-2,23
2	H	4-MeO-Ph-NHCO	0,85	-0,59	0,66
3	H	3-Cl-Ph-NHCO	0,30	-1,05	0,16
4	CH ₃	H	-2,48	-0,58	-2,52
5	CH ₃	4-MeO-Ph-NHCO	0,70	-0,07	0,36
6	CH ₃	3-Cl-Ph-NHCO	0,39	-0,33	-0,10
7	C ₂ H ₅	H	-2,52	-0,34	-2,41
8	C ₂ H ₅	4-MeO-Ph-NHCO	0,22	-0,11	0,34
9	C ₂ H ₅	3-Cl-Ph-NHCO	-0,20	-0,53	-0,14
10	n-C ₃ H ₇	H	-2,60	0,06	-2,48
11	n-C ₃ H ₇	4-MeO-Ph-NHCO	0,10	0,06	0,40
12	n-C ₃ H ₇	3-Cl-Ph-NHCO	0,04	-0,09	-0,07
13	CH ₂ -CH=CH ₂	4-MeO-Ph-NHCO	0,32	0,07	0,22
14	n-C ₄ H ₉	H	-2,78	0,03	-2,81
15	n-C ₄ H ₉	4-MeO-Ph-NHCO	0,49	0,56	0,12
16	n-C ₄ H ₉	3-Cl-Ph-NHCO	0,22	0,39	-0,27
17	t-C ₄ H ₉	H	-3,06	0,51	-2,83
18	t-C ₄ H ₉	4-MeO-Ph-NHCO	0,10	0,51	-0,09
19	t-C ₄ H ₉	3-Cl-Ph-NHCO	-0,44	0,15	-0,51
20	(CH ₃) ₂ CH-CH ₂ -CH ₂	H	-2,85	-0,08	-3,59
21	(CH ₃) ₂ CH-CH ₂ -CH ₂ -	4-MeO-Ph-NHCO	-1,48	-1,16	-0,66
22	(CH ₃) ₂ CH-CH ₂ -CH ₂	3-Cl-Ph-NHCO	-1,60	-1,30	-1,12
23	(CH ₃) ₂ C=CH-CH ₂	H	-2,91	-0,23	-3,19
24	(CH ₃) ₂ C=CH-CH ₂	4-MeO-Ph-NHCO	-1,60	-1,17	-0,44
25	(CH ₃) ₂ C=CH-CH ₂	3-Cl-Ph-NHCO	-1,74	-1,46	-0,91
26	Ph-CH ₂ -CH ₂	H	-2,45	0,02	-2,60
27	Ph-CH ₂ -CH ₂	4-MeO-Ph-NHCO	0,00	0,44	-0,32
28	Ph-CH ₂ -CH ₂	3-Cl-Ph-NHCO	-0,90	-0,44	-0,81
29	Ph-CH ₂ -CH ₂ -CH ₂	H	-2,63	0,02	-3,13
30	Ph-CH ₂ -CH ₂ -CH ₂	4-MeO-Ph-NHCO	-1,60	-1,11	-1,16
31	Ph-CH ₂ -CH ₂ -CH ₂	3-Cl-Ph-NHCO	-1,78	-1,22	-1,59
32	2,4,5-Br ₃ -Ph-CH ₂ -CH ₂	H	-3,65	-0,43	-3,67
33	2,4,5-Br ₃ -Ph-CH ₂ -CH ₂	4-MeO-Ph-NHCO	-1,40	-0,02	-1,42
34	2,4,5-Br ₃ -Ph-CH ₂ -CH ₂	3-Cl-Ph-NHCO	-1,85	-0,06	-1,79
35	2-(1-naphthyl)ethyl	H	-3,53	-0,98	-3,29

(Table 1. cont.....)

#	R	R ¹	Observed (pK _i , nM)	Predicted (aMEP/PLS)	Predicted (CoMFA)
36	2-(-naphthyl)ethyl	4-MeO-Ph-NHCO	-1,20	-0,46	-1,32
37	2-(-naphthyl)ethyl	3-Cl-Ph-NHCO	-1,71	-1,01	-1,77
38	-	H	-1,93	-1,50	-1,55
39	-	4-MeO-Ph-NHCO	0,85	0,34	1,12
40	-	3-Cl-Ph-NHCO	0,72	0,84	0,67
41	CH ₃	Ph-NH-CO	0,80	0,02	0,34
42	CH ₃	3,4-Cl ₂ -Ph-NH-CO	-0,53	-0,63	-0,03
43	CH ₃	3,4-O-CH ₂ O--Ph-NH-CO	0,62	0,82	0,44
44	CH ₃	4-NO ₂ -Ph-NH-CO	0,37	0,11	-0,10
45	CH ₃	4-CH ₃ -Ph-NH-CO	0,51	0,10	0,47
46	CH ₃	4-Br-Ph-NH-CO	0,34	-0,02	0,53
47	CH ₃	4-F-Ph-NH-CO	0,47	-0,04	0,35
48	CH ₃	4-CF ₃ -Ph-NH-CO	0,13	-0,08	0,07
49	CH ₃	2-OMe-Ph-NH-CO	0,15	-0,41	0,41
50	CH ₃	3-OMe-Ph-NH-CO	0,10	-0,50	0,03
51	CH ₃	2-Cl-Ph-NH-CO	0,04	-0,39	0,30
52	CH ₃	4-Cl-Ph-NH-CO	0,54	0,22	0,45
53	C ₂ H ₅	Ph-NH-CO	0,74	0,27	0,33
54	C ₂ H ₅	3,4-Cl ₂ -Ph-NH-CO	-0,48	-0,34	-0,04
55	C ₂ H ₅	3,4-O-CH ₂ O--Ph-NH-CO	0,57	1,03	0,33
56	C ₂ H ₅	4-NO ₂ -Ph-NH-CO	0,19	0,40	-0,10
57	C ₂ H ₅	4-CH ₃ -Ph-NH-CO	0,85	0,66	0,46
58	C ₂ H ₅	4-Br-Ph-NH-CO	0,43	0,28	0,51
59	C ₂ H ₅	4-F-Ph-NH-CO	0,07	-0,05	0,33
60	C ₂ H ₅	4-CF ₃ -Ph-NH-CO	0,01	0,04	0,06
61	C ₂ H ₅	2-OMe-Ph-NH-CO	0,25	-0,02	0,41
62	C ₂ H ₅	3-OMe-Ph-NH-CO	0,07	-0,04	0,02
63	C ₂ H ₅	2-Cl-Ph-NH-CO	0,52	0,65	0,29
64	C ₂ H ₅	4-Cl-Ph-NH-CO	0,70	0,65	0,43
65	n-C ₃ H ₇	Ph-NH-CO	0,82	0,75	0,40
66	n-C ₃ H ₇	3,4-Cl ₂ -Ph-NH-CO	-0,40	0,09	0,02
67	n-C ₃ H ₇	3,4-O-CH ₂ O-Ph-NH-CO	0,52	1,26	0,39
68	n-C ₃ H ₇	4-NO ₂ -Ph-NH-CO	0,09	0,61	-0,03
69	n-C ₃ H ₇	4-CH ₃ -Ph-NH-CO	0,40	0,51	0,51
70	n-C ₃ H ₇	4-Br-Ph-NH-CO	0,35	0,54	0,57
71	n-C ₃ H ₇	4-F-Ph-NH-CO	0,54	0,64	0,40
72	n-C ₃ H ₇	4-CF ₃ -Ph-NH-CO	0,29	0,62	0,14
73	n-C ₃ H ₇	2-OMe-Ph-NH-CO	0,47	0,51	0,68
74	n-C ₃ H ₇	3-OMe-Ph-NH-CO	0,40	0,48	0,09
75	n-C ₃ H ₇	2-Cl-Ph-NH-CO	0,15	0,50	0,35
76	n-C ₃ H ₇	4-Cl-Ph-NH-CO	0,47	0,66	0,50
77	n-C ₄ H ₉	Ph-NH-CO	0,68	0,82	0,12
78	n-C ₄ H ₉	3,4-Cl ₂ -Ph-NH-CO	0,57	1,19	-0,25
79	n-C ₄ H ₉	3,4-O-CH ₂ O--Ph-NH-CO	0,30	1,05	0,11
80	n-C ₄ H ₉	4-NO ₂ -Ph-NH-CO	0,26	0,85	-0,30
81	n-C ₄ H ₉	4-CH ₃ -Ph-NH-CO	0,68	0,97	0,24
82	n-C ₄ H ₉	4-Br-Ph-NH-CO	0,04	0,37	0,30
83	n-C ₄ H ₉	4-F-Ph-NH-CO	0,10	0,41	0,13
84	n-C ₄ H ₉	4-CF ₃ -Ph-NH-CO	0,14	0,56	-0,14
85	n-C ₄ H ₉	2-OMe-Ph-NH-CO	0,24	0,41	0,19

(Table 1. cont.....)

#	R	R ¹	Observed (pK _i , nM)	Predicted (aMEP/PLS)	Predicted (CoMFA)
86	n-C ₄ H ₉	3-OMe-Ph-NH-CO	0,22	0,49	-0,18
87	n-C ₄ H ₉	2-Cl-Ph-NH-CO	0,07	0,63	0,08
88	n-C ₄ H ₉	4-Cl-Ph-NH-CO	0,37	0,75	0,22
89	CH ₃	Ph-CH ₂ -CO	0,09	-0,32	0,45
90	C ₂ H ₅	Ph-CH ₂ -CO	-0,01	-0,18	0,45
91	n-C ₃ H ₇	Ph-CH ₂ -CO	0,00	0,28	0,52
92	n-C ₄ H ₉	Ph-CH ₂ -CO	0,05	0,53	0,25
93	Ph-CH ₂ CH ₂	CONHCH(CH ₃) ₂	-0,95	-0,63	-1,22
94	Ph-CH ₂ CH ₂	CONHC(CH ₃) ₃	-0,69	0,08	-0,90
95	(CH ₃) ₂ CHCH ₂ CH ₂	CONHCH(CH ₃) ₂	-1,81	-1,66	-2,10
96	(CH ₃) ₂ CHCH ₂ CH ₂	CONHC(CH ₃) ₃	-1,59	-0,93	-1,73
97	CH ₃ CH ₂ CH ₂	CONHCH(CH ₃) ₂	-1,32	-1,56	-1,07
98	CH ₃ CH ₂ CH ₂	CONHC(CH ₃) ₃	-1,18	-0,85	-0,70
99	Ph-CH ₂ CH ₂ CH ₂	CONHCH(CH ₃) ₂	-1,74	-1,31	-1,78
100	Ph-CH ₂ CH ₂ CH ₂	CONHC(CH ₃) ₃	-1,81	-0,92	-1,46
101	Ph-CH ₂ CH ₂	CO(CH ₂) ₃ NH ₃ ⁺ Cl	-1,81	-1,04	-1,56
102	Ph-CH ₂ CH ₂	CO(CH ₂) ₃ NHCOOC(CH ₃) ₃	-0,17	1,04	0,12
103	(CH ₃) ₂ CHCH ₂ CH ₂	CO(CH ₂) ₃ NH ₃ ⁺ Cl	-1,95	-1,07	-2,41
104	(CH ₃) ₂ CHCH ₂ CH ₂	CO(CH ₂) ₃ NHCOOC(CH ₃) ₃	-0,39	0,78	-0,74
105	Ph-CH ₂ CH ₂	COCH ₂ NH ₃ ⁺ Cl	-2,21	-0,95	-1,98
106	Ph-CH ₂ CH ₂	CO(CH ₂) ₂ NH ₃ ⁺ Cl	-1,90	-0,59	-1,73

Validation was done with an external test set and the results of the two models were compared. Interestingly, our preliminary results seem to indicate that this new alternative approach could robustly compete with the already well consolidate CoMFA approach.

In the last five years, many efforts have been conducted for searching potent and selective human A₃ adenosine antagonists [10-13]. Indeed, this specific adenosine receptor subtype is under examination in relation to its potential therapeutic applications [10-13]. In particular antagonists for A₃ receptors seem to be useful for the treatment of inflammation, and regulation of cell growth. In this field many different classes of compounds have been proposed, possessing good affinity (nM range) and with a broad range of selectivity [10-13]. Recently, our group has synthesized a new series of pyrazolo-triazolo-pyrimidines bearing different substitutions at the N⁵ and N⁸ positions, which has been described as highly potent and selective human A₃ adenosine receptor antagonists [10-14]. The development of robust and validate 3D-QSAR model is a useful tool to rational design or screening new potential antagonists. In this paper, we showed that autocorrelation vectors derived from MEP surfaces can represent a fast and efficient alternative to the conventional CoMFA strategy.

RESULTS AND DISCUSSION

Autocorrelation MEP/PLS Model. 3D models of all 106 A₃ antagonists were obtained by using the 3D structure generator Corina [15]. Partial atomic charges were calculated by the PEOE method [16] and its extension to conjugated systems [17]. Points were randomly distributed on the van

der Waals surface and the electrostatic potential at each point was calculated by a classical Coulomb approach using a unit positive point charge and the partial charges on all atoms of the molecule (see Experimental Section for details). Autocorrelation vectors were calculated for each molecule for distance intervals of 1 Å, from 1 to 13 Å, by using the electrostatic potential as property on the van der Waals surface and a point density of 10 points/Å². As a first step toward a model for the binding affinities of all pyrazolo-triazolo-pyrimidine derivatives, the suitability of the surface property autocorrelation vectors as QSAR descriptors was investigated by a partial least square (PLS) to search for a relationship between autocorrelation vectors and biological activity. PLS was used in conjunction with cross-validation to obtain the optimal number of components to be used in the subsequent analyses. PLS analysis based on least squares fit gave a correlation with a cross-validated r^2_{cv} of 0.78 (standard deviation = 0.67), with the maximum number of components set equal to 4 (maximum number of components set equal to 3, 5, or 6 gave lower cross-validated r^2_{cv}) and the cross-validation groups set equal to the number of observations. The non-cross-validated PLS analysis was repeated with the optimum number of components, as determined by the cross-validated analysis, to give an r^2 of 0.82. The autocorrelation MEP/PLS model of pyrazolo-triazolo-pyrimidine derivatives exhibited a good cross-validated correlation, indicating that it was highly predictive. Cross-validation provides information concerning the predictive ability of the QSAR data set by minimizing the occurrence of chance correlations in the QSAR model. Predicted vs experimental pK_i values for autocorrelation

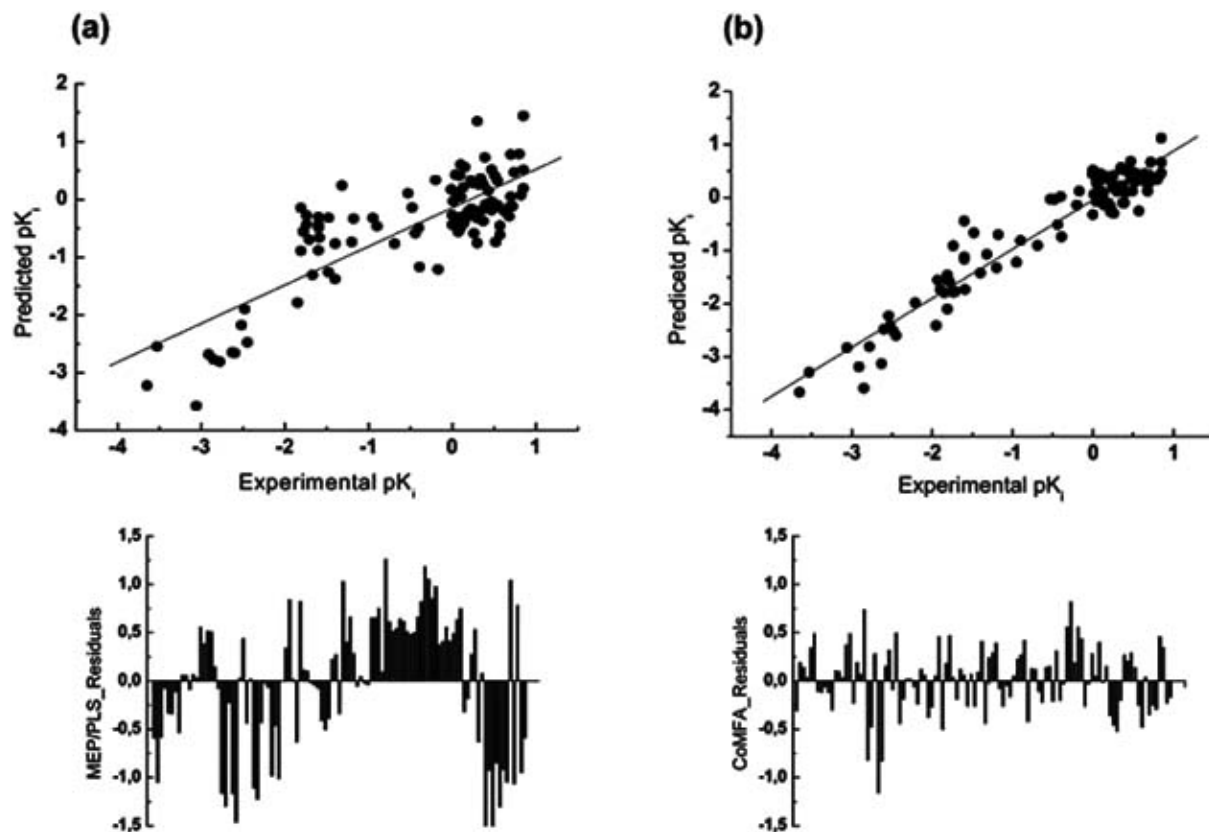


Fig. (1). Statistic of the Calibration Autocorrelation MEP/PLS (*panel a*) and CoMFA (*panel b*) Models. Experimental vs. Predicted pKi values and corresponding residual plots are shown.

MEP model of the human A₃ training set are shown in Fig. (1a).

Conventional CoMFA Approach. In parallel to the autocorrelation MEP/PLS model, a CoMFA 3D-QSAR methodology was applied to the data from human A₃ adenosine antagonist binding assay as a means of identifying the structural features of pyrazolo-triazolo-pyrimidine derivatives responsible for affinity. Also in this case, PLS was used in conjunction with cross-validation to obtain the optimal number of components to be used in the subsequent analyses. PLS analysis based on least squares fit gave a correlation with a cross-validated r^2_{cv} of 0.84 (standard deviation = 0.019), with the maximum number of components set equal to 6 (maximum number of components set equal to 4, 5, or 7 gave lower cross-validated r^2_{cv}) and the cross-validation groups set equal to the number of observations. The non-cross-validated PLS analysis was repeated with the optimum number of components, as determined by the cross-validated analysis, to give an r^2 of 0.92. To obtain statistical confidence limits, the non-cross-validated analysis was repeated with 10 bootstrap groups, which yielded an r^2 of 0.80 (optimum number of components was 6), standard deviation = 0.022, steric contribution = 0.571, and electrostatic contributions = 0.429. The CoMFA-derived QSAR of pyrazolo-triazolo-pyrimidine derivatives exhibited a good cross-validated correlation, indicating that it was highly predictive. In particular, the high bootstrapped r^2 value and a small standard deviation suggest a high degree of confidence in the analysis. This method determines the

confidence level of a predicted value when a small number of data points are present. An example would be a cross-validation study where the number of points in the training set must be large. This leaves an insufficient number of data points in the test set to generate a confidence level of the accuracy. Instead of using a single set of N data points, the Bootstrap Method creates a very large number of sets by randomly choosing N members of the set of data points (which obviously means that a single data point may be present more than once in a particular set of data). An average value can be determined from each of these sets and this large number of averages can be fitted by a gaussian function to generate an overall average and standard deviation.

Fitted vs measured pKi values for the CoMFA analysis of the human A₃ training set are shown in Fig. (1b).

The CoMFA coefficients corresponding to each sampled field point in the resulting correlation equation were graphically contoured. Contours corresponding to the steric (green and yellow) and electrostatic (blue and red) fields are plotted together with compound (5) docked inside the human A₃ binding cavity as shown in Fig. (2). We selected derivative 5 as reference potent a selective human A₃ antagonist. The polyhedra describe the regions of space where the steric and the electrostatic fields are predicted by the CoMFA model to have the greatest effect on binding affinity. The yellow and the blue polyhedra correspond to regions of the field that are predicted to decrease the A₃

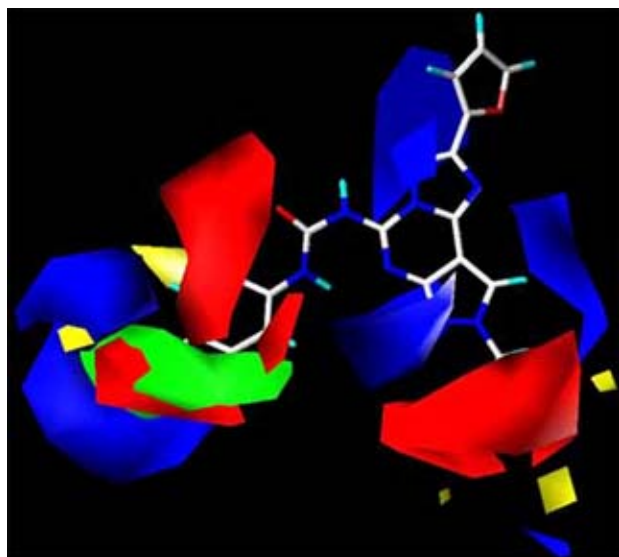
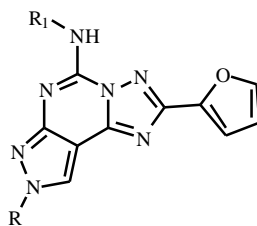


Fig. (2). CoMFA steric and electrostatic STDEV*COEFF contour plots from the analysis based on the A₃ receptor 3D-QSAR without cross-validation. Compound (**5**) shown inside the fields. Favoring activity: green, bulky group (contribution level 80%); yellow, less bulky; blue, positive charge (contribution level 70%); red, negative charge.

receptor affinity, while the green and the red regions are predicted to increase binding affinity.

Table 2. Structures and human A₃ (calculated and predicted by both 3D-QSAR models) binding affinities of new synthesized compounds (**107-123**).



#	R	R ¹	Observed (pK _i , nM)	Calculated (MEP/PLS)	Calculated (CoMFA)
107	n-C ₈ H ₁₇	H	-3,31	-2,82	-3,52
108	CH ₂ -cC ₆ H ₁₁	H	-3,24	-2,96	-3,14
109	CH ₃	CONH-cC ₆ H ₁₁	0,00	0,31	0,03
110	CH ₃	CONH-nC ₅ H ₁₁	-0,29	0,25	0,48
111	CH ₃	CONH-Ph-4-CH ₂ CO ₂ Et	0,09	0,88	0,35
112	CH ₂ CH ₂ CH(CH ₃) ₂	CONH-cC ₆ H ₁₁	-1,31	-0,62	-1,32
113	CH ₂ CH ₂ CH(CH ₃) ₂	CONH-nC ₅ H ₁₁	-0,91	-0,40	-0,87
114	CH ₂ CH ₂ CH(CH ₃) ₂	CONH-Ph-4-CH ₂ CO ₂ Et	-1,08	-0,02	-1,00
115	CH ₂ CH ₂ CH(CH ₃) ₂	CONH-Ph-4-CH ₃	-0,87	-0,45	-0,86
116	n-C ₈ H ₁₇	CONH-cC ₆ H ₁₁	-1,21	-0,95	-1,23
117	n-C ₈ H ₁₇	CONH-nC ₅ H ₁₁	-0,89	-0,90	-0,79
118	n-C ₈ H ₁₇	CONH-Ph-4-CH ₂ CO ₂ Et	-1,00	-0,48	-0,91
119	n-C ₈ H ₁₇	CONH-Ph-4-CH ₃	-0,65	-0,89	-0,49
120	CH ₂ -cC ₆ H ₁₁	CONH-cC ₆ H ₁₁	-0,83	-1,03	-0,86
121	CH ₂ -cC ₆ H ₁₁	CONH-nC ₅ H ₁₁	-0,60	-0,91	-0,41
122	CH ₂ -cC ₆ H ₁₁	CONH-Ph-4-CH ₂ CO ₂ Et	-0,84	-0,21	-0,54
123	CH ₂ -cC ₆ H ₁₁	CONH-Ph-4-CH ₃	-0,77	-0,81	-0,40

Validation Set. Based on both autocorrelation MEP/PLS model and CoMFA model, we have designed, synthesized and tested 17 new derivatives with a different spectrum of affinity at the human A₃ receptor (**107-123**, test set, shown in Table 2) [14,18]. Compounds **107-123** were used to evaluate the predictive power of both 3D-QSAR models.

As in the calibration step, a similar and good predictive ability was obtained ($r^2_{\text{pred}}=0.87$ and $r^2_{\text{pred}}=0.79$ for autocorrelation MEP/PLS model and CoMFA respectively). Table 2 and Figure 3 show that the affinities of all the examined compounds were predicted within 0.25 log unit across a range of 2.00 log units. Impressively, the predicted pK_i were very closed to the experimental values (residual plot is shown in Fig. (3)). In particular, derivative (**111**) (4-[3-(2-furan-2-yl-8-methyl-8H-pyrazolo[4,3-e][1,2,4]triazolo[1,5-c]pyrimidin-5-yl)-ureido]-phenylacetic acid ethyl ester) is active in a sub-nanomolar range (K_i = 0.82 nM) as correctly predicted by both regression model. This is a clear evidence that autocorrelation MEP/PLS model is a robust and efficient alternative to a classical CoMFA approach.

EXPERIMENTAL SECTION

Synthesis and biological evaluation of analyzed A₃ receptor antagonists. All pyrazolo-triazolo-pyrimidine derivatives (**1-123**) were synthesized according to a well known procedure for the synthesis of the pyrazolo[4,3-e]-1,2,4-triazolo[1,5-c]pyrimidines, previously reported [14,18]. Binding of [³H]MRE3008-F20 to CHO cells transfected with the human recombinant A₃ adenosine receptors was

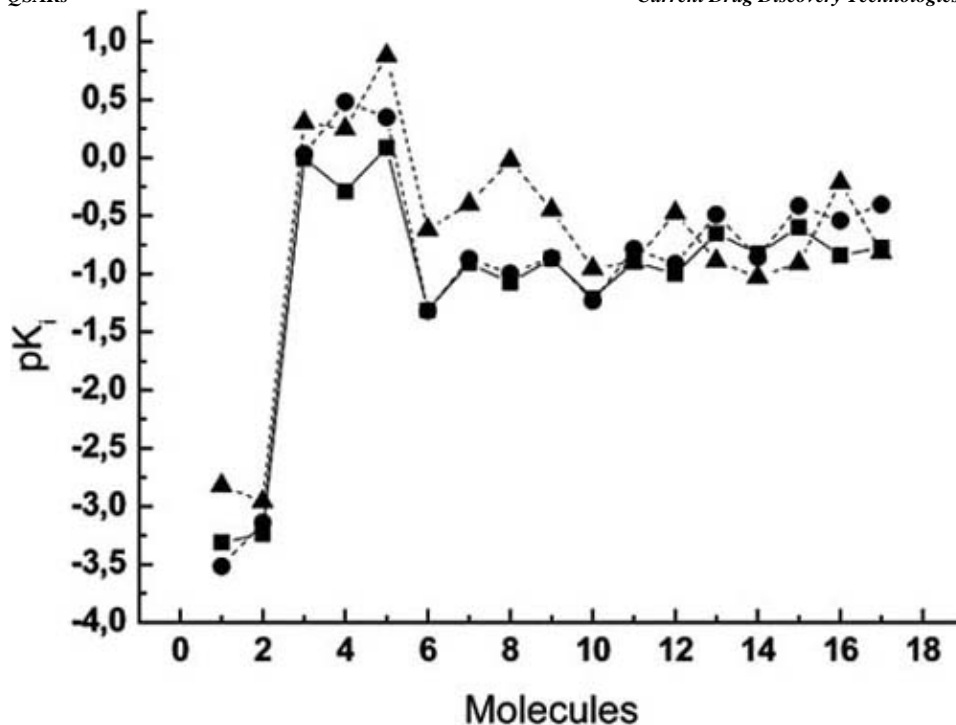


Fig. (3). Comparison of experimental pK_i values (■) with those predicted by Autocorrelation MEP/PLS (○) and CoMFA (●) Models.

performed according to Varani *et al.* [19].

Computational methodologies. All 3D-QSAR studies were carried out on a 6 CPU (PIV 2.0-3.0 GHZ) linux cluster running under openMosix architecture [20].

Autocorrelation MEP studies have been done using ADRIANA (version 1.0) suite [21]. CoMFA studies were performed on a Silicon Graphics Power Indigo2 R8000 workstation running SYBYL 6.5 [22].

Molecular structure building. 3D models of all 106 A_3 antagonists were obtained by using the 3D structure generator Corina [15]. Corina is an integral part of Adriana QSAR-suite [21]. Conformers generation and best conformer selection has been carried out using standard parameters of Corina.

Molecular Electrostatic Potential (MEP) calculation. In the present work MEPs derive from a classical point charge model: the electrostatic potential for each molecule is obtained by moving a unit positive point charge across the van der Waals surface and it is calculated at various points j on this surface by the following equation [5]:

$$V_j = \sum_i^{\text{atoms}} \frac{q_i}{r_{ji}}$$

where q_i represents the partial charge of each atom i and r_{ji} is the distance between points j and atom i . Starting from the 3D model of a molecule and its partial atomic charges, the electrostatic potential or another appropriate property is calculated for points on the molecular surface. Partial atomic charges were calculated by the PEOE method [16] and its extension to conjugated systems [17] implemented by Petra module of Adriana suite [21]. Connolly's solvent accessible surface with a solvent radius of 2.0 Å has been used to

project the corresponding MEP. For the pyrazolotriazolopyrimidine derivative (**1**) (shown in Table 1, about 3500 points are obtained which are characterized by their Cartesian coordinates and the value of the electrostatic potential. After applying the autocorrelation function, the autocorrelation vector is obtained. Connolly's solvent accessible surface and the corresponding MEP have been calculated by Surface module of Adriana [21].

Autocorrelation vector. The first application of these vectors as molecular descriptors has been published by Moreau and Broto [23,24], who applied the classical mathematical notion of an autocorrelation function to the topology of molecular structures. The autocorrelation vector is presented as an intrinsic descriptor of the distribution of an atomic property along the molecular graph. Each component of the autocorrelation vector is calculated as follows:

$$A(d) = \sum_{ij} p_i p_j$$

where A is the autocorrelation coefficient referring to atom pairs i, j , p_i is the atomic property and d the i, j topological distance.

Starting from this concept a new 3D descriptor has been introduced, which is based on the autocorrelation of properties at distinct points on the molecular surface [5]. The different components of the autocorrelation vector are derived in this way:

$$A(d_{\text{lower}}, d_{\text{upper}}) = 1/L \sum_{ij} p_i p_j (d_{\text{lower}} < d_{ij} < d_{\text{upper}})$$

where the i, j distance d belongs to the $d_{\text{lower}}, d_{\text{upper}}$ interval and L is number of distances in the same interval. The application of this concept made possible to compare different molecular properties, as this 3D descriptor represents a compressed expression of the distribution of the

property p on the molecular surface [5]. The parameters for the calculation of the autocorrelation coefficient are the following: $d_{lower} = 1 \text{ \AA}$; $d_{upper} = 13 \text{ \AA}$; $L = 12$ point density = 10 points/ \AA^2 ; vdW radius reduction factor = 1.000. All parameters have been changed in a various way to see if it was possible to improve the model capability, but non-significance results were derived. Considering distances from 1 to 13 \AA , with a step width of 1 \AA , twelve autocorrelation coefficients are calculated. This transformation produces a unique fingerprint of each molecule under consideration. Autocorrelation vector have been calculated by Surface module of Adriana [21].

Partial Least Square (PLS) analysis. All PLS analysis have been carried out using "The Unscramble" statistical software [25].

CoMFA - Molecular Superposition. The pyrazolo-triazolo-pyrimidine moiety is believed to be a determinant key for binding interactions of described derivatives. The energetically more stable conformation has been used for all compounds, as input for the CoMFA field calculations.

CoMFA - Field Calculations and Regression Techniques. The electrostatic and steric fields were sampled along a three-dimensional lattice encompassing all molecules in each receptor data set. The lattice consisted of 720 sample points based on a 2.0 \AA lattice spacing with boundaries extending 4.0 \AA beyond the largest structure in all directions. 0.75 and 1.5 \AA lattice spacing were also used without improvements of the CoMFA results. The lattice points within the union volume of the superimposed structures were dropped. The probe used to calculate the CoMFA fields consisted of a sp³ carbon atom with a +1 charge and a van der Waals radius of 1.52 \AA . The steric and electrostatic fields were calculated separately for each molecule using respectively a Lennard-Jones 6-12 potential and a Coulombic potential with a $1/r$ distance-dependent dielectric. The steric and electrostatic energies were truncated at 30 kcal/mol. The field values corresponding to the 720 sample points for each molecule, together with binding affinity data, were stored in a SYBYL Molecular Spreadsheet to facilitate statistical analysis.

Partial least squared regression analysis was performed on the human A₃ antagonist dataset using a subset of CoMFA field sample points falling with a standard deviation

1.0 kcal/mol. The steric and the electrostatic fields were scaled to equalize their weighting in the CoMFA models (SYBYL command "scaling CoMFA_std"). PLS was performed using cross-validation to evaluate the predictive ability of the CoMFA models [26]. The optimal number of latent variables came from cross-validation equation having the lowest standard error and a significance level $\geq 99.5\%$ was estimated using the stepwise F-test. Bootstrap analysis of the dataset was used to evaluate the statistical confidence limits of the results [26]. A q^2 value of 2.0 was adopted for both the cross-validated and non-cross-validated analysis. values of 1.0 or 0.5 did not significantly change the calculated r^2 .

Initial PLS analyses were performed in conjunction with the cross-validation option (leave-one-out method) to obtain the optimal number of components to be used in the subsequent analyses of the dataset. The PLS analysis was

repeated with the number of cross-validation groups set to zero. The optimal number of components was designated as that which yielded the highest cross-validated r^2 values in the non-cross-validated (conventional) analyses. The final PLS analysis with 10 bootstrap groups and the optimal number of components was performed on the complete dataset.

The corresponding calibration equation (resulting from the simultaneous contribution of all the observations) was derived after the optimal dimensionality of each receptor-model was established, by PLS analysis and cross validation. The calibration equation with latent variables was then converted to the original parametric space represented by probe-ligand interaction energies. A 3D-QSAR whose coefficients were associated with statistically significant lattice locations was therefore derived. CoMFA coefficient contour maps were generated by interpolation of the pairwise products between the 3D-QSAR coefficients and the standard deviations of the associated energy variables.

CoMFA- "Predictive" r^2 Values. The "predictive" r^2_{pred} was based only on molecules not included in the training set and is defined as explained by Marshall and co-workers [27].

Test Sets. The test sets consisted of 17 new synthesized molecules for the considered training set **107-123** (Table 2). These structures were chosen to maximize a uniform sampling of biological activity. All predicted activities for the test set molecules were calculated using both autocorrelation MEP/PLS and CoMFA models. The results of the non cross-validated model on the test sets are summarized in Table 2.

CONCLUSION

Concluding, our preliminary results seem to indicate that this new alternative approach could efficiently compete with other well consolidate 3D-QSAR methodologies such as CoMFA approach. We consider that the transformation of MEP surfaces into autocorrelation vectors is a very useful tool to generate a unique chemical fingerprint of each molecule (*autocorrelation vectors*), and we consider this strategy very promising for a large molecular database screening applications. In fact, using autocorrelation strategy it is not necessary to align all molecules before generating the corresponding QSAR model. Following this feeling, we are running an intense virtual screening program to severally validate the real capability of our MEP/PLS model to discover new chemically diverse A₃ antagonists from very large chemical libraries.

ACKNOWLEDGMENT

We thank Molecular Network GmbH (Erlangen, Germany) for the assistance in using ADRIANA modeling suite. We also thank Dr. Paolo Braiuca (Department of Pharmaceutical Sciences, University of Trieste, Italy) and Dr. Christian Montopoli (Department of Pharmaceutical Sciences, University of Padova, Italy) for their support in preparing CoMFA model. The molecular modeling work coordinated by S.Moro has been carried out with financial supports of Associazione Italiana per la Ricerca sul Cancro (AIRC), Milan, and the Italian Ministry for University and Research (MIUR), Rome, Italy.

REFERENCES

- [1] Kubinyi H., Folkers G., Martin Y.C.: *3D QSAR in Drug Design*; Kluwer Academic Publishers: Vol. 3, Recent Advances, (1998).
- [2] Wold S. *Chemometric Methods in Molecular Design*; Waterbeemd H. van der Ed. VCH: Weinheim, (1995).
- [3] Cramer R.D., Patterson D.E., Bunce J.D.: *Comparative Molecular Field Analysis (CoMFA). I. Effect of Shape on Binding of Steroids to Carrier Proteins*. J. Am. Chem. Soc. 110, 5959, (1988).
- [4] Gasteiger J., Li X., Rudolph C., Sadowsky J., Zupan J.: *Representation of Molecular Electrostatic potentials by Topological Feature Maps*. J. Am. Chem. Soc. 116, 4608, (1994).
- [5] Wagener M., Sadowsky J., Gasteiger J.: *Autocorrelation of Molecular Surface Properties for Modeling Corticosteroid Binding Globulin and Cytosolic Ah Receptor Activity by Neural Networks*. J. Am. Chem. Soc. 117, 7769, (1995).
- [6] Clementi S., Cruciani G., Riganelli D., Valigi R., Costantino G., Baroni M., Wold S.: *Pharm. Pharmacol. Lett.* 3, 5, (1993).
- [7] Gancia E., Bravi G., Mascagni P., Zaliani A.: *Global 3D-QSAR methods: MS-WHIM and autocorrelation*. J. Comput. Aided Mol. Des. 2000 14 293.
- [8] Pastor M., Cruciali G., McLay I., Pickett S., Clementi S.: *Grid Independent descriptors (GRIND). A Novel Class of Alignment-Independent Threedimensional Molecular Descriptors*. J. Med. Chem. 43 3233, (2000).
- [9] Klein C.T., Kaiser D., Ecker G.: *Topological Distance Based 3D Descriptors for Use in QSAR and Diversity Analysis*. J. Chem. Inf. Comput. Sci. 44, 200, (2004).
- [10] Muller C.E.: *Medicinal Chemistry of Adenosine A₃ Receptor Lignads*. *Curr. Top. Med. Chem.* 3, 445, (2003).
- [11] Moro S., Defflorian F., Spalluto G., Pastorin G., Cacciari B., Kim K., Jacobson K.A.: *Demystifying the three dimensional structure of G protein-coupled receptors (GPCRs) with the aid of molecular modeling*. *Chem Commun (Camb)* 21, 2949, (2003).
- [12] Fishman P. Bar-Yehuda S.: *Pharmacology and Therapeutic Applications of A₃ Receptor Subtype*. *Curr. Top. Med. Chem.* 3, 463, (2003).
- [13] Moro S., Spalluto G., Jacobson K.A.: *Techniques: Recent developments in computer-aided engineering of GPCR ligands using the human A₃ adenosine receptor as an example*. *Trends Pharmacol. Sci.* 26, 44, (2005).
- [14] Baraldi P.G., Cacciari B., Moro S., Spalluto G., Pastorin G., Da Ros T.N., Varani K., Gessi S., et al.: *Synthesis, Biological Activity, and Molecular Modeling Investigation of New Pyrazolo[4,3-e]-1,2,4-triazolo[1,5-c]pyrimidine Derivatives as Human A₃ Adenosine Receptor Antagonists*. J. Med. Chem. 45, 770, (2002).
- [15] CORINA; Molecular Networks GmbH: Erlangen - Germany, (2003).
- [16] Gasteiger J., Marsili M.: *Iterative Partial Equalization of Orbital Electronegativity - A Rapid Access to Atomic Charges*. *Tetrahedron* 36, 3219, (1980).
- [17] Gasteiger J., Saller H.: *Berechnung der Ladungsverteilung in konjugierten Systemen durch eine Quantifizierung des Mesomeriekonzeptes*. *Angew. Chem.* 97, 699, (1985).
- [18] Moro S., Braiuca P., Defflorian F., Pastorin G., Cacciari B., Baraldi P.G., Varani K., Borea P.A., et al.: *Combined target-based and ligand-based drug design approach as tool to define a novel 3D-pharmacophore model of human A₃ adenosine receptor antagonists: pyrazolo[4,3-e]1,2,4-triazolo[1,5-c]pyrimidine derivatives as a key study*. J. Med. Chem. 48, 152, (2005).
- [19] Varani K., Gessi S., Dionisotti S., Ongini E., Borea P.A.: *[³H]-SCH 58261 labelling of functional A_{2A} adenosine receptors in human neutrophil membranes*. *Br. J. Pharmacol.* 123, 1723, (1998).
- [20] OpenMosix: <http://www.openMosix.org> (2004).
- [21] ADRIANA; Molecular Networks GmbH: Erlangen - Germany, (2003).
- [22] SYBYL 6.5; TRIPOS Associates: St. Louis MO (1993).
- [23] Moreau G., Broto P.: *The Autocorrelation of a Topological Structure: a new Molecular Descriptor*. *Nouv. J. Chim.* 4, 359, (1980).
- [24] Moreau G., Broto P.: *Autocorrelation of Molecular Structures, Application to SAR Studies*. *Nouv. J. Chim.* 4, 757, (1980).
- [25] THE UNSCRUMBLER 9.0; CAMO Process AS: Oslo - Norway (2003).
- [26] Cramer RD III., Bunce JD., Patterson DE., Frank IE.: *Cross-Validation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies*. *Quant. Struct.-Act. Relat.* 7, 18, (1988).
- [27] Waller C.L.O., Giolitti A., Marshall G.R.: *Three-Dimensional QSAR of Human Immunodeficiency Virus (1) Protease Inhibitors. I. A CoMFA Study Employing Experimentally-Determined Alignment Rules*. J. Med. Chem. 36, 4152, (1993).