Eur J Med Chem (1997) 32, 493–504

Interest of cluster significance analysis in structure–affinity relationships for non-xanthine heterocyclic antagonists of adenosine

M Adenot[1]*, V Benezech[2], J Bompart[2], PA Bonnet[2], JP Chapat[2], G Grassy[1]

[1]*Centre de Biochimie Structurale, UMR CNRS 9955, INSERM U414, Faculté de Pharmacie, 15 Ave Charles-Flahault, 34060 Montpellier Cedex;*
[2]*Laboratoire de Chimie Organique Pharmaceutique, Faculté de Pharmacie, 15 Ave Charles-Flahault, 34060 Montpellier Cedex, France*

(Received 9 October 1996; accepted 16 December 1996)

Summary — In order to define some predictive rules for the discrimination of adenosine antagonists by their A_1-receptor affinity, we performed a systematic QSAR analysis. As no significant descriptors of affinity were found, we then proposed to introduce a calculated enthalpy or entropy change for the interaction as a first approximation of the affinity descriptors. Since the structural details of the common receptor binding site remain to be determined, we utilized an indirect strategy involving the simulation of amino acid residues that are thought to interact with the ligand. Estimating enthalpic and entropic components by means of a semi-empirical quantum mechanical AM1 force calculation, we found a significant clustering of enthalpy change values. This method provides a good descriptor of interaction and also a simple tool for testing hypotheses on the nature of putative binding sites.

adenosine antagonist / QSAR analysis / cluster significance analysis / thermodynamics / AM1

Introduction

Adenosine receptors are currently known as ubiquitous G protein-coupled receptors with a large diversity of effectors [1]. This results in a great variety of physiological effects regarding adenosine and its derivatives. Recent discovery of an increasing number of physiopathological processes involving adenosine has led to an increased interest in purinergic transmission pharmacology. However, a number of questions remain concerning the topography of adenosine transmission pathways as well as the structure and function of adenosine receptors. The structure of adenosine receptors remains to be elucidated by current experimental techniques such as X-ray crystallography, neutron diffraction or NMR.

Adenosine receptors may be divided into two major classes based on the functional coupling to adenylyl cyclase activity. A_1 receptors are generally linked to the inhibition of adenosine 3',5'-cyclic monophosphate (cyclic AMP) generation, while A_2 receptors stimulate the production of cyclic AMP. In recent years, however, the subclassification of adenosine receptors has become more complex, with the acceptance of A_3 receptors and subtypes of A_2 receptors. The A_2 receptors are divided into two classes, based upon the observation that some A_2 receptors have EC_{50} values for adenosine in the high nanomolar range (1–20 nM) for the A_{2a} class rather than the micromolar range (5–20 μM) for the A_{2b} class [2]; A_{2a} and A_{2b} are both distinguished by their molecular weights and tissue localization. The A_3 receptor subtypes cloned from rat brain and from the hypophysial *par tuberalis* of sheep have recently been described [3, 4].

Both A_1 and A_2 receptors are ubiquitous and coexist in numerous tissues such as brain, kidneys or lungs where they generally have opposite effects. This results in widespread physiological effects concerning adenosine: A_1 receptor activation causes negative cardiac chronotropy, dromotropy and inotropy [5], vasoconstriction, bronchoconstriction [6], maintains inhibitory tonus in hippocampal neurons [7] and decreases the secretion of insulin [8], acid [9], renin [10] and erythropoietin [11]. A_2 receptor activation is responsible for renal artery dilatation [10], bronchodilatation, stimulation of gluconeogenesis and platelet inhibition [6].

In the present work, we proposed a simple method to test hypotheses about the nature of residues directly

*Correspondence and reprints

implicated in ligand binding. Such a theoretical tool based on thermodynamic considerations is complementary to site-directed mutagenesis studies: receptor sites are characterized by their capacity to discriminate ligands via structural or physicochemical features. We used a descriptor which characterizes the interaction process between ligand and receptor, rather than one of these two partners: enthalpy change during interaction. If a receptor site is constituted of one, two, three or more residues and is thought to be the site of the interaction, the enthalpy change during the recognition event should discriminate between high- and low-affinity ligands. In fact, all ligands that bind in a similar way to the receptor site should have a similar enthalpy change for the interaction. Simulation of the interaction of ligands with some theoretical receptor sites leads to corresponding ΔH values; a simple cluster significance analysis should predict if these values are grouped in a significantly narrow range of ΔH. This approach allows the selection of valid hypotheses of receptor–ligand interaction, in good agreement with site-directed mutagenesis studies.

In order to define some simple predictive rules for the discrimination of adenosine antagonists by their A_1 affinity, we used this technique with a set of 23 non-xanthine heterocyclic adenosine antagonists. First of all, classical electronic, steric, lipophilic and topological descriptors were calculated. We added the proton affinity as a specific interaction descriptor.

The modelling of the entire receptor site was not envisaged, as such an approach supposes a detailed experimental knowledge of the receptor structure. Since the mode of interaction of heterocyclic adenosine antagonists is still unknown, we had to make suppositions about which residues might interact with the ligand. A straightforward model consists of the reduction of the receptor site to only one amino-acid residue in interaction with the ligand [13]. For this reason, only residues supposed to interact were modelled. We calculated enthalpy and entropy changes during ligand–residue complexation equilibrium. The most likely type of interaction between the studied heterocycles and their putative receptor site is the creation of hydrogen bonds between the protein and some of the ligand heteroatoms. We supposed a priori that residues implicated in the recognition process were hydrogen bond donors or acceptors. The modelization of the receptor site was based on site-directed mutagenesis information [12]. Chemical modification experiments on the A_1 receptor have shown that 251-histidine in the sixth transmembrane helix is important for ligand affinity [12]. Ijzerman [12] docked cyclopentyladenosine in a cavity involving this 251-histidine. Peet [14] assumed that 246-serine and 277-threonine were implicated in the interaction and proposed a docking-model for (*R*)-phenylisopropyladenosine. We modelled receptor–ligand complexes based on both the Ijzerman and Peet hypotheses in order to determine if one of them would better discriminate among compounds by their A_1 receptor affinity. The histidine (His) model corresponds to the Ijzerman hypothesis; the serine (Ser) and threonine (Thr) models correspond to the Peet hypothesis in the two types of interaction. Water was explicitly taken into account in a hydrated ligand model; a single molecule of water was added in the vicinity of the hypothetical region of recognition of the ligand.

Statistical analysis of the results provides information on the ability of a receptor site model to discriminate among weak and strong affinity ligands. This approach might be a helpful tool in the formulation and discrimination of hypotheses for the synthesis of new antagonists.

Materials and methods

Materials

Model building and preliminary minimizations in the MM_2 force field were performed using MAD software [15] on a Silicon Graphics Indy workstation. Calculations of lowest energy conformers, proton affinity and thermodynamic functions were carried out using MOPAC version 6.0 [16] on a Hewlett–Packard Apollo 9000 series 700 workstation.

Similarity indices were calculated with ASP software [17]; topological descriptors and data analyses were performed with TSAR [18] on a Silicon Graphics Indy workstation.

Choice of structures and modelling

The structure–affinity relationships study was performed on a set of 67 A_1 receptor antagonists selected from the literature [19–23]. Although antagonists are thought to bind to different binding sites, we assume that the structural variability in our set of compounds is small enough to make all those molecules interact in the same way with the receptor. All these molecules possess a heterocyclic structure and their adenosine–receptor affinity (K_i or IC_{50}) was determined by homogenous binding tests. Three main classes of non-xanthine antagonists were retained: 4-amino[1,2,4]triazolo[4,3-*a*]quinoxalines, [1,2,4]triazolo[1,5-*c*]quinazolines and 4-amino-1*H*-imidazo[4,5-*c*]quinolines. Some pyrrolo[2,3-*d*]pyrimidine and pyrimido[4,5-*b*]tetrahydroindole derivatives were added.

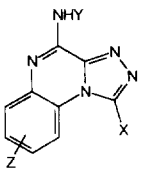
The 67 selected molecules are presented in tables I to IV. All were modelled in their neutral and all possible protonated forms. Our aim was to look for a priviliged site of protonation which might participate in a putative hydrogen bond. The geometry of these models was optimized by a quantum semi-empirical method using the AM_1 Hamiltonian (PRECISE Option) [24].

Molecular property descriptors

Electronic parameters

Electronic parameters are characterized by atomic charges. Electronic density on the frontier orbitals was evaluated with

Table I. Structures and properties of compounds **1**–**37**. 4-Amino[1,2,4]triazolo[4,3-*a*]quinoxaline



Compound	*X*	*Y*	*Z*	IC_{50} *(nM)*[a]	S_E1[b]	S_E9[b]
1	H	H	H	4600		
2	H	H	8-Cl	3400		
3	H	H	7-F	6100		
4	H	H	8-F	2900		
5	H	Et	H	29 000		
6	H	Et	7-F	11 000		
7	H	Et	8-F	3200		
8	H	*i*-Pr	H	4100		
9	H	*i*-Pr	7-F	4500	0	1503
10	H	*i*-Pr	8-F	680	0	1544
11	Et	H	7-Cl	75	61	826
12	Et	H	8-Cl	110	61	582
13	Et	H	7-F	580	59	552
14	Et	Et	8-OMe	710	0	2088
15	Et	*i*-Pr	6-Cl	6761	0	2075
16	Et	*i*-Pr	8-Cl	60	32	1829
17	Et	*i*-Pr	8-F	180	0	2331
18	Et	*c*-C_5H_9	8-Cl	20	0	4294
19	Et	*c*-C_6H_{11}	8-Cl	44	32	4725
20	CF_3	H	8-Cl	65	57	498
21	CF_3	*i*-Pr	H	170	0	2292
22	CF_3	*i*-Pr	8-Cl	24	0	2609
23	CF_3	*i*-Pr	8-F	57	0	2747
24	CF_3	*c*-C_5H_9	8-Cl	5.5	0	4311
25	CF_3	*c*-C_6H_{11}	8-Cl	28	0	4566
26	CF_3	*c*-C_6H_{11}	8-F	32	0	4288
27	C_2F_5	*i*-Pr	8-Cl	24	0	4289
28	Ph	*i*-Pr	8-Cl	3200	0	3952
29	OH	*i*-Pr	8-F	730	29	1473
30	Me	*c*-C_5H_9	H	34	0	2796
31	Et	*c*-C_5H_9	H	28	0	3590
32	Pr	*c*-C_5H_9	H	39	0	4577
33	Bu	*c*-C_5H_9	H	55	0	4696
34	CF_3	*c*-C_5H_9	H	7.3	0	3893
35	Et	*c*-Bu	H	58	0	2638
36	Et	*c*-C_6H_{11}	H	80	0	3698
37	Et	Ph	H	630	30	3356

[a]IC_{50} (nM) for A_1 receptor vs CHA (cyclohexyladenosine) binding to rat brain membranes; [b]S_E1 and S_E9 are the 1st and 9th components of the 3-D autocorrelogram weighted by electrophilic superdelocalizability (SE).

parameters such as electrophilic superdelocalizability (S_E) and nucleophilic superdelocalizability (S_N), as well as the frontier ratio (*FR*) [25]. Designating occupied and unoccupied molecular orbitals with indices from 1 to *i* and from *i* + 1 to *n* respectively, S_E and S_N can be written as:

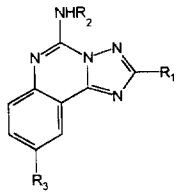
$$S_E = 2\ \Sigma_{(j=1,i)}\ C_{rj}^2/E_j$$
$$S_N = 2\ \Sigma_{(j=i+1,n)}\ C_{rj}^2/E_j$$

where C_{rj} is the *r*th atomic orbital coefficient in the *j*th molecular orbital, and E_j is this molecular orbital energy.

FR is the sum of the squares of C_{rj} coefficients over the two highest energy occupied orbitals (NHOMO and HOMO) weighted by these orbitals' energy:

$$FR = (\Sigma\ C^2{}_{r\mathrm{NHOMO}}/E_{\mathrm{NHOMO}}) + (\Sigma\ C^2{}_{r\mathrm{HOMO}}/E_{\mathrm{HOMO}})$$

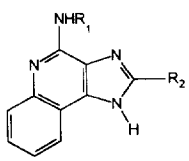
Table II. Structures and properties of compounds **38–48**. [1,2,4]Triazolo[1,5-*c*]quinazoline



Compound	R_1	R_2	R_3	IC_{50} *(nM)*[a]	S_E1[b]	S_E9[b]
38	2-Furyl	H	Cl	21	59	2928
39	2-Furyl	H	H	38	61	2075
40	2-Furyl	Me	Cl	89	31	3428
41	2-Furyl	C_2H_4OH	Cl	21	38	4836
42	2-Furyl	*i*-Pr	Cl	22	31	4960
43	2-Furyl	*i*-Bu	Cl	61	0	6440
44	2-Furyl	*t*-Bu	Cl	> 10 000	0	5642
45	2-Furyl	H	OMe	31		
46	2-Furyl	H	OH	9.5		
47	Pyrrolyl	H	Cl	610	88	3228
48	THF	H	Cl	827	60	2473

[a]IC_{50} (nM) for A_1 receptor versus CHA (cyclohexyladenosine) binding to rat brain membranes; [b]S_E1 and S_E9 are the 1st and 9th components of the 3-D autocorrelogram weighted by electrophilic superdelocalizability (SE).

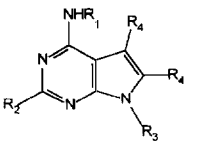
Table III. Structures and properties of compounds **49–53**. 4-Amino-1*H*-imidazo[4,5-*c*]quinolines



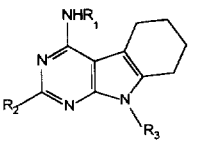
Compound	R_1	R_2	K_i *(nM)*[a]	S_E1[b]	S_E9[b]
77	*c*-C_5H_9	H	43		
79	R_1	H	1500	26	3568
82	*c*-C_5H_9	*c*-C_5H_9	39	27	6184
83	H	Ph	34	95	3123
85	*c*-C_5H_9	Ph	10	27	7242

[a]K_i (nM) for A_1 receptor versus CPX (8-cyclopentyl-1,3-dipropylxanthine) binding to calf brain cortical membrane; [b]S_E1 and S_E9 are the 1st and 9th components of the 3-D autocorrelogram weighted by electrophilic superdelocalizability (SE); R_1 = (*S*)-1-methyl-2-phenylethyl.

Table IV. Structures and properties of compounds **54–67**.



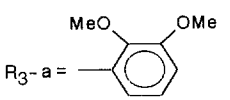
Type I
Pyrrolo[2,3-*d*]pyrimidines



Type II
Pyrimido[4,5-*b*]
tetrahydroindoles

Compound	R_1	R_2	R_3	R_4 *or type*	K_i *(nM)*[a]	S_E1[b]	S_E9[b]
54	H	H	H	Me	55000		
55	H	H	Allyl	Me	49000		
56	H	H	Ph	Me	18000	67	1143
57	H	H	R_3-a	Me	54300	67	2468
58	H	H	*p*-Br-Benzyl	Me	> 100000	66	2656
59	H	H	Ph	H	3100	67	2942
60	H	H	Ph	CHO	50100		
61	H	H	Ph	Type II	3240	67	1330
62	H	Me	H	Me	88000		
63	H	*o*-Cl-Ph	H	Me	5170	96	1662
64	H	Me	Ph	Me	30400	68	1378
65	H	Ph	Ph	Me	36	68	4486
66	H	*o*-Cl-Ph	Ph	Me	990	68	5970
67	H	Ph	R_3-b	Me	4.7	69	4931

[a] K_i (nM) for A_1 receptor versus *R*-PIA (*R*-phenylisopropyladenosine) binding rat brain membranes; [b] S_E1 and S_E9 are the 1st and 9th components of the 3-D autocorrelogram weighted by electrophilic superdelocalizability (SE).



R_3-b = (*R*)-1-phenylethyl

Lipophilicity parameter
Lipophilicity was evaluated by the determination of log *P* using the atomic incremental method of Ghose et al [26].

Steric parameters
Molar refractivity [27] and van der Waals (VdW) volume were selected to characterize the steric effect.

Topological parameters
We used molecular graph descriptors (Wiener [28], Randic [29], Balaban [30] and Kier indices [31]) as well as 2-D and 3-D autocorrelograms [32] weighted by electrophilic and nucleophilic superdelocalizabilities and frontier ratio [25].

Molecular similarity indices
In order to compare the selected molecules to the five leader compounds of each chemical series, we used Carbo and Hodgkin molecular similarity indices [34–37] relative to the electrostatic potential of each compound [25]. The leader compounds were chosen for their high A_1-receptor affinity.

The Carbo molecular similarity index [34] allows the comparison of the electronic density of two superimposed molecules and is defined by:

$$C_{AB} = \int \rho_A \rho_B dv / (\int \rho_A^2 dv)^{1/2} (\int \rho_B^2 dv)^{1/2}$$

where ρ_A et ρ_B are the electronic densities on molecules A and B integrated over the whole VdW volume. This formula was used to compare electrostatic potentials and fields.

Molecular similarity indices can be calculated by numerical integration over a 3-D grid surrounding both molecules to be compared. In this case, the previous equation can be written:

$$C_{AB} = \Sigma_{(i=1,N)} (\rho_A \rho_B) / (\Sigma_{(i=1,N)} \rho_A^2)^{1/2} (\Sigma_{(i=1,N)} \rho_B^2)^{1/2}$$

where N is the total number of grid points and ρ_A is the matched properties (electrostatic field or potential) for compound A at the *i*th grid point. This index takes a value of + 1 when the two compounds have similar properties and a value of –1 when they have totally dissimilar properties.

If one compound property value is the multiple of another compound property value ($\rho_A = n\rho_B$), the Carbo index takes a value of + 1. Hodgkin proposed an alternative index to overcome this inconvenience [35–37]:

$$H_{AB} = 2 \Sigma_{(i=1,N)} (\rho_A \rho_B) / (\Sigma_{(i=1,N)} \rho_A^2 + \Sigma_{(i=1,N)} \rho_B^2)$$

Proton affinity
In order to estimate an eventual contribution of the ligand nitrogen atoms in the binding process, we calculated the proton affinity *PA* [33]:

$$PA = \Delta H_f^\circ(B) + \Delta H_f^\circ(H^+) - \Delta H_f^\circ(BH^+)$$

where B is the neutral form of the ligand and BH^+ is its protonated form. Standard enthalpies of formation ΔH_f° are calculated using AM_1 [24] where parameters are optimized so as to reproduce the experimental heat of formation (ie, the enthalpy change to form 1 mol of compound at 25 °C from its elements in their standard state). The AM_1 Hamiltonian was preferred to PM_3 to reproduce proton affinity values in heterocyclic series [33]. Proton affinity allows the estimation of the role of proton binding on some of the ligand nitrogen atoms by exchange with an acidic function or creation of a hydrogen bond during the recognition process.

Thermodynamic descriptors of interaction

Free energy changes (ΔG) during the recognition process are dependent on the affinity constant according to the Van't Hoff law. We estimated the advantages of the ΔG value for discriminating among ligands by their receptor affinities. ΔG is usually written as an additive quantity:

$$\Delta G = \Delta G_{solvation} + \Delta G_{conformation} + \Delta G_{interaction} + \Delta G_{motion}$$

In this work, we investigated the motion component of ΔG for its ease of calculation and applicability to large series of compounds. ΔG_{motion} can be divided into three terms: ΔG_{vib}, ΔG_{rot} and ΔG_{tra}. The rotational terms of energy and entropy are expressed from the moment of inertia, and translational terms are expressed from the molecular weights. They are both relatively simple to calculate for small molecules. ΔG_{vib} values are computed by normal mode analysis [38–39]. This method extracts normal modes of vibrations from a spatial coordinates set. Normal mode analysis gives vibrational energy:

$$E_{vib} = N_A \Sigma_i \{(1/2) h\nu_i + [h\nu_i \cdot \exp(-h\nu_i/kT)/(1 - \exp(-h\nu_i/kT))]\}$$

and the vibrational part of entropy:

$$S_{vib} = R \Sigma_i \{(h\nu_i/kT) \cdot \exp(-h\nu_i/kT)/(1 - \exp(-h\nu_i/kT)) - \ln(1 - \exp(-h\nu_i/kT))\}$$

then the enthalpy of the system:

$$H_{motion} = U + pV = E_{vib} + E_{rot} + E_{tra} + pV$$

and the conformational part of the entropy:

$$S_{motion} = S_{vib} + S_{rot} + S_{tra}$$

There are two limitations to the normal mode analysis: a force calculation only makes sense if the structure is at a stationary point, ie, a minimum or a transition state, and the harmonic approximation should be justified. In the case of a ligand–protein complex, the surface potential is rather anharmonic. To make the calculation possible, we defined a model assembly of the complex. This model includes the ligand and the few amino-acidic residues involved in the recognition process. According to the putative pharmacophoric sites in the compounds, as well as experimental site-directed mutagenesis [12], the binding of heterocyclic antagonists of adenosine involves at least one hydrogen bond.

The interaction process between the ligand and its receptor has been defined by the equation:

$$L \cdots H_2O + R \rightleftharpoons L \cdots R + H_2O$$

[1] [2]

Hydrophilic ligands are always solvated before complexation and lose their solvation shell on complex formation. Water will modify the electrostatic component of the hydrogen bond in comparison with what is observed in the gas phase calculation. The solvation contribution in quantum mechanical calculations can be estimated by a continuum model [42] or simulation of the solvent effect using the numerical self-consistent reaction field method [43]. Alternately, a solvent molecule can be explicitly taken into account in the model [25]. This method is less realistic but gives an interpretable meaning for the calculated ΔG: it should be interpreted as a measure of the stability of the complexes with respect to the hydrates, ie, the spontaneity of complex formation in solution.

One water molecule was located close to the putative investigated hydrogen bond site which could be either the exocyclic NH_2 or one of the N sp^2 of the ligand. Each molecule presents several pharmacophoric sites suitable for creating intermolecular hydrogen bonds. Since we obtain no information about which sites are involved in the recognition process, we should screen each of them with different docking hypotheses. Among the different tested models, we present those which are consistent with the Ijzerman [12] or Peet [14] hypothesis.

H_{motion} and S_{motion} values can be calculated over the two states of the ligand–receptor system: before (state [1]) and after (state [2]) interaction. This method requires the previous modelling of the ligand in its hydrated (state [1]) and receptor-complexed (state [2]) forms. ΔG_{motion} is then expressed by:

$$\Delta G_m = (H_m[2] - H_m[1]) - T(S_m[2] - S_m[1])$$

Molecular motion contribution of enthalpy H_m and entropy S_m were calculated for different models after energy minimization of the hydrated and complexed forms over a limited set of 23 molecules selected from the initial set. Compounds were separated into high-affinity (K_i in the nanomolar range) and low-affinity groups (K_i in the micromolar range) (table V). Molecules with intermediate K_i values were excluded from this study. We did not envisage building a model of the whole receptor; only residues thought to be involved in the recognition process were included in the model. Energy calculations were performed using a semi-empirical method with an AM_1 Hamiltonian [24].

Analysis of the local geometry at the recognition site (interatomic distances, angles between partners of the hydrogen bond) allowed verification that the model was geometrically and energetically compatible with the existence of hydrogen bonds.

Data analysis

Discriminant analysis

In order to find more significant descriptors describing affinity to A_1 receptors, we used discriminant analysis for two classes: ligands with low affinity (IC_{50} > 580 nM) were distributed in class 1 and ligands with high affinity (IC_{50} < 180 nM) in class 2. Class 1 and 2 included 32 and 35 molecules, respectively. Variables used in the study were: log$P$, molar refractivity, VdW volume, atomic charges, proton affinity, molecular similarity indices, Wiener, Randic, Balaban and Kier indices, 2-D and 3-D autocorrelograms weighted by electrophilic and nucleophilic superdelocalizabilities (S_E and S_N) and by molecular frontier ratio *(FR)*.

Cluster significance analysis [40]

In the absence of linearity between affinity and ΔG_m, ΔH_m or ΔS_m descriptors, we found a clustering tendency for molecules

Table V. Selected molecules for cluster significance analysis.

Compound	*Log IC_{50}*	*Affinity class*	*ΔH (His) (cal/mol)*	*ΔH (Ser) (cal/mol)*	*ΔH (Thr) (cal/mol)*
9	3.65	1	2765	977	698
10	2.83	1	2751	976	761
12	2.04	2	2890	963	1006
13	2.76	1	2775	988	802
14	2.85	1	2664	997	790
15	3.83	1	2666	962	915
17	2.26	2	2661	1010	946
18	1.30	2	2701	920	966
21	2.23	2	2737	1023	928
23	1.76	2	2730	977	916
24	0.78	2	2643	1031	957
27	1.38	2	2774	1061	871
28	3.51	1	2682	–[a]	917
29	2.86	1	2761	1020	730
30	1.53	2	2428	724	930
33	1.74	2	2518	903	926
35	1.76	2	3467	1028	907
37	2.80	1	2448	668	1435
38	1.32	2	2453	901	923
40	1.95	2	2613	1050	995
44	4.00	1	2618	984	703
47	2.79	1	2615	964	815
48	2.92	1	2700	987	1059

[a]Missing data.

of the same class. Lack of linearity between affinity and physico-chemical descriptors implied the use of alternative non-linear data analysis methods. Cluster significance analysis is a technique used to group a set of points that consist of similar members based on their distances with chosen parameters through hyperspace. Mean square distance *(MSD)* is defined by:

$$MSD = \Sigma_{(j=1,N)} d^2(j, j')/N$$

for N compounds j separated by an euclidian distance d in the n-space of properties. The CSA algorithm compares the validity of the a priori classification by testing all possible combinations of individual points and comparing the *MSD* of clusters. The cluster significance is the probability that the proposed classification is more valid than those obtained by chance only.

This technique was applied to analyze log IC_{50} vs ΔH_m, log IC_{50} vs ΔS_m and log IC_{50} vs ΔG_m clouds. In this analysis, class 1 and 2 included 11 and 12 molecules, respectively.

Results and discussion

Classical analysis

A classical discriminant analysis indicates that only 81% of the molecules are well classified with two parameters in decreasing order of importance:

– the 9th component of the 3-D autocorrelogram weighted by S_E (S_E9): this component includes all pairs of atoms with an interatomic distance of 8 to 9 Å, each atom being weighted by its S_E value;

– the 1st component of the 3-D autocorrelogram weighted by S_E (S_E1): this component includes all pairs of atoms with an interatomic distance of 0 to 1 Å, each atom being weighed by its S_E value.

This analysis indicates that molecules with the greatest number of pairs of atoms separated by 8 to 9 Å and an important electrophilic superdelocalizability value have higher affinity for adenosine A_1 receptors. These include bulky mainly tricyclic compounds, like 8-chloro-4-cyclopentylamino-3-trifluoromethyl-(1,2,4)triazolo[4,3-*a*]quinoxaline. The introduction of bulky groups such as the phenyl or furyl moieties increases S_E and consequently their affinity.

These results give a rational basis to a structural optimization of the design of competitive antagonists of adenosine A_1 receptors. However, the more representative descriptors that we found to discriminate ligands according to their affinity are not easy to use and not really convincing. Classical analysis demonstrates the failure of linear QSAR despite the use of a wide range of descriptors. This is unfortunately a current situation in QSAR today. The need for alterna-

tive routes, either in data analysis or in the choice of relevant descriptors, leads us to explore the field of interaction descriptors.

Thermodynamic model analysis

Enthalpy values for the hydrated ligand, H_m [1], or the complex, H_m [2], were found to be not predictive at all. Then we used ΔH_m (or ΔS_m) as a descriptor of the interaction between the two partners rather than the static description of only one of them.

No linear relationship was shown between ΔG_m, ΔH_m or ΔS_m and log IC_{50}. However, in the Ser and Thr models, we observed a tendency of the enthalpy change ΔH_m to gather compounds of a same class into a cluster. This tendency has been analyzed using cluster significance analysis [40]. For each model, the geometry was checked and compared with geometric criteria for hydrogen bonds [44, 45] (table VI). These criteria are a maximum distance of 2.5 Å between the H atom and the acceptor and a minimum value of 90° for the (D–H–A) angle.

The His model

In this model, the 23 selected molecules were positioned close to a 5-methylimidazole molecule representing a histidine side chain so as to favour the formation of a complex by creation of a hydrogen bond either between the ligand exocyclic amino group and the sp^3 nitrogen atom of the histidine residue, or between the histidine NH group and a ligand nitrogen atom (fig 1).

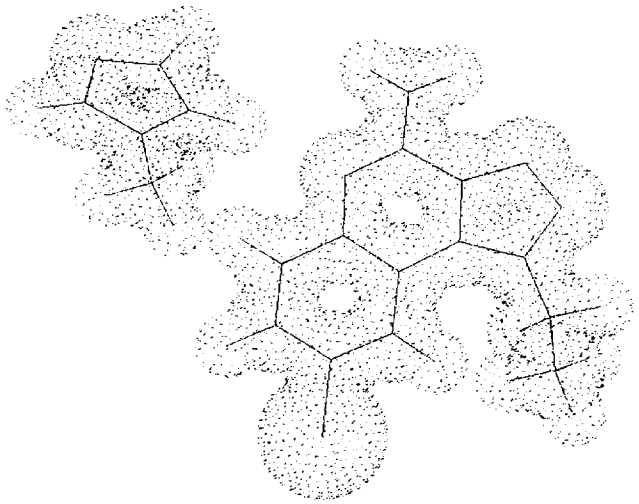
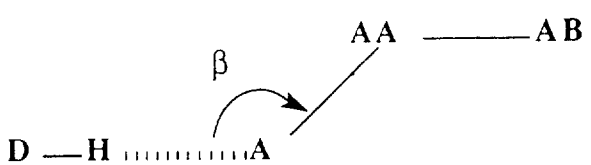


Fig 1. His-model represents interaction between compound **16** and 5-methylimidazole standing for the 251-histidine residue.

No significant clustering of compounds was observed in the cloud (fig 2). Therefore, this model does not appear to provide any predictive value. Furthermore, such a complex entity is not consistent with hydogen bonding since the distance between the H atom and the acceptor does not fit the distance criteria. The hypothesis that such a binding mode would occur with A_1 receptors seems irrelevant.

Table VI. Geometry analysis of the 4-amino[1,2,4]triazolo[4,3-*a*]quinoxaline complex for the three models. D is the donor heavy atom, H the hydrogen, A the acceptor and AA the acceptor antecedent.



		Models			*H-bond criteria*
		His	*Ser*	*Thr*	
d(H ··· A) (Å)	1[a]	*	2.12 (± 0.02)	2.11 (± 0.08)	< 2.5 Å
	2[b]	2.59 (± 0.06)	2.69 (± 0.03)	2.72 (± 0.06)	< 2.5 Å
d(H ··· AA) (Å)	1	*	3.07 (± 0.02)	3.09 (± 0.06)	< 3.9 Å
	2	3.48 (± 0.08)	3.25 (± 0.02)	3.10 (± 0.08)	< 3.9 Å
β	1	*	157.1 (± 4.0)°	168.9 (± 3.9)°	> 90°
	2	151.7 (± 8.0)°	117.7 (± 1.4)°	113.3 (± 2.1)°	> 90°

[a]H-bond between the amino-acid and the exocyclic NH_2 group of the ligand; [b]H-bond between the amino-acid and the N sp^2 of the ligand; *no data.

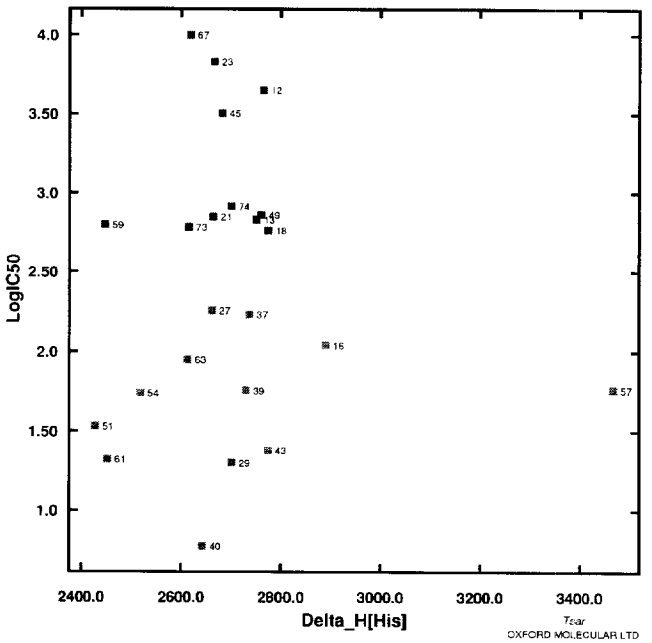


Fig 2. Log IC_{50} vs enthalpy change for the His-model.

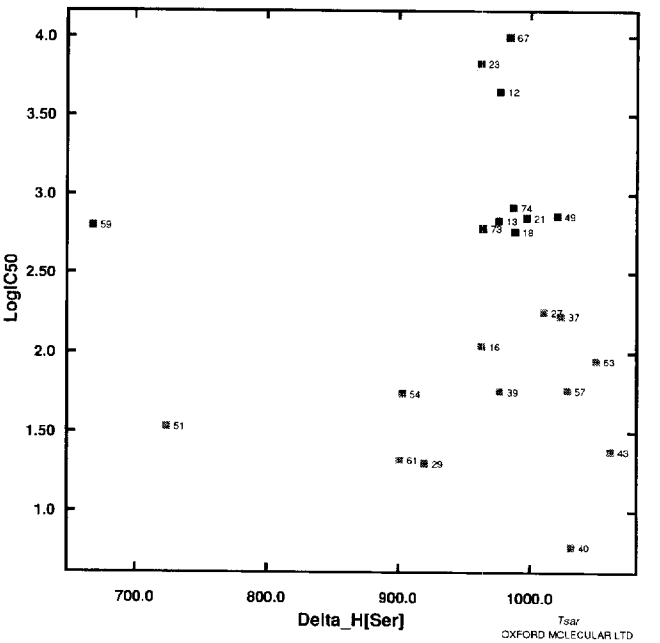


Fig 4. Log IC_{50} vs enthalpy change for the Ser-model.

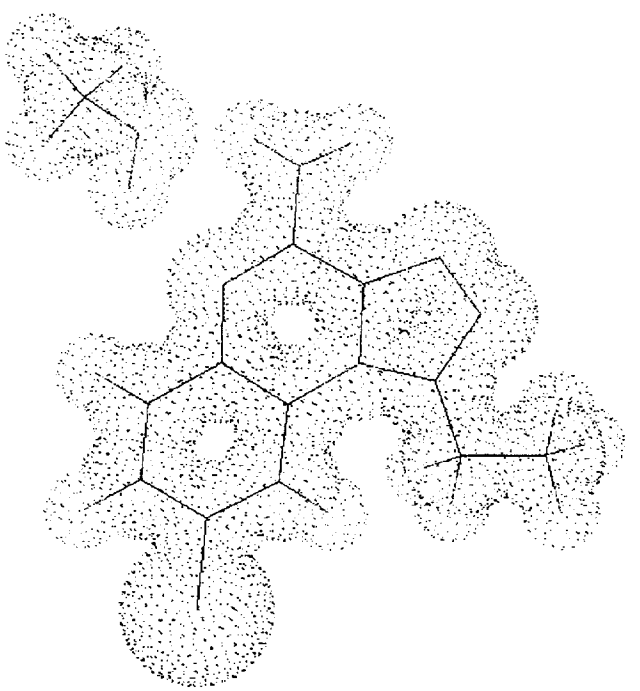


Fig 3. Ser-model represents interaction between compound **16** and methanol standing for the 246-serine residue.

The Ser model

Ligands were positioned close to a methanol molecule representing a serine side chain so as to favour creation of a complex by hydrogen bond formation between the methanol OH group and the exocyclic nitrogen atom of the ligand or one of the N sp^2 of the ligand (fig 3).

No significant clustering was observed (fig 4) due to compound **59**. However, examination of the log IC_{50} vs ΔH_m plot suggests a strong tendency to cluster the low-affinity compounds in a smaller free energy interval than it does for high-affinity compounds. If compound **59** is excluded from the analysis, we find a high probability that this clustering was not due to chance (P = 0.997). On the other hand, the enthalpy change ΔH_m seems to constitute a more favourable criterion (P = 1.0) to clustering than the entropy change ΔS_m (P = 0.997). However, none of the ΔH_m, ΔS_m and ΔG_m descriptors exhibited any discrimination between high- and low-affinity compounds. Thus, this model is not very relevant for structure–affinity relationships studies, and the hypothesis that this binding mode would occur with A_1 receptors is not retained.

Geometry analysis shows that the hydrogen bond preferentially occurs between methanol and the exocyclic nitrogen atom.

The Thr model

Ligands were positioned close to a methanol molecule representing a threonine side chain so as to favour creation of a complex by hydrogen bond formation between the methanol OH group and the exocyclic nitrogen atom or one of the sp^2 nitrogen atoms of the ligand (fig 5).

In this model, the high-affinity compounds cluster in a smaller free energy change interval than the low-affinity compounds. The probability that this clustering was not due to chance is high ($P = 0.952$). Again, the enthalpy change ($P = 1.0$) (fig 6) seems to be a more favourable clustering criterion than entropy change ($P = 0.972$). High-affinity compounds exhibit an enthalpy range from 860 to 1000 cal/mol, an entropy range from 10 to 14.5 cal/K/mol and a free energy range from –2000 to –3500 cal/mol. When ΔH_m, ΔS_m and ΔG_m values are within these ranges, it is reasonable to assume that we are in presence of high-affinity compounds. With ΔH_m, a discrimination appears: high-affinity compounds are contained in the previously defined interval of 800 to 1000 cal/mol, whereas low-affinity compounds remain outside this interval.

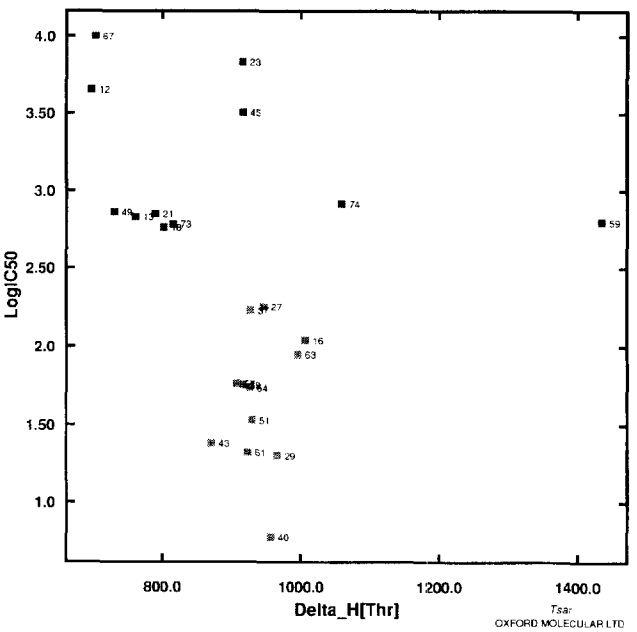


Fig 6. Log IC_{50} vs enthalpy change for the Thr-model.

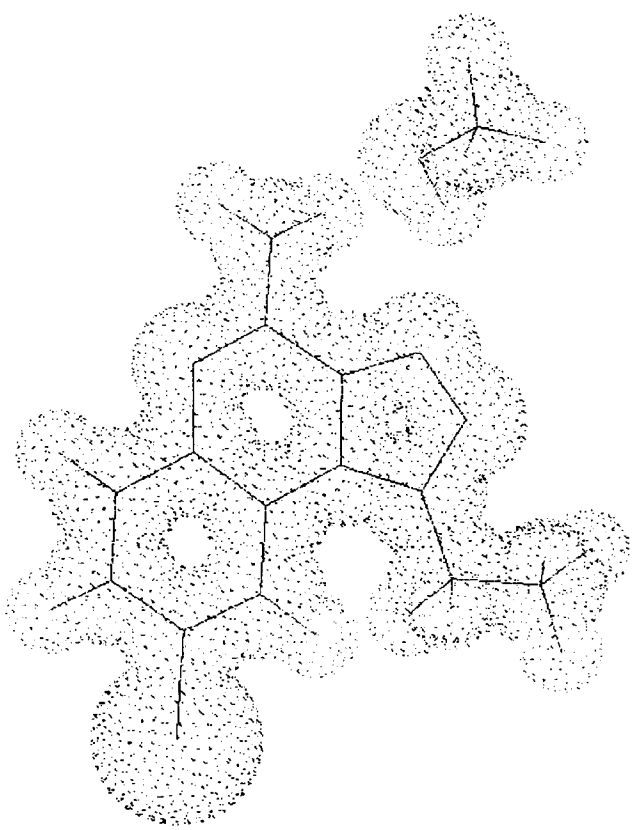


Fig 5. Thr-model represents interaction between compound **16** and methanol standing for the 277-threonine residue.

As previously, geometry analysis shows that the hydrogen bond preferably occurs between methanol and the exocyclic nitrogen atom.

The importance of the interaction in clustering is emphasized when we consider the sparse distribution of log IC_{50} vs H_m [1] (fig 7) and log IC_{50} vs H_m [2] (fig 8). It is obvious that H_m [1] or H_m [2] values considered individually provide no relevant information, but their difference leads to a very typical clustering.

This model is the most relevant for structure–affinity relationship studies and might be relevant for the selection of high-affinity compounds. The hypothesis that such a binding mode occurs between antagonist ligands and A_1 receptors can be retained.

It should be pointed out that the less positive mean value of ΔH_m for the class 1 molecules with respect to class 2 cannot be interpreted in terms of affinity scale for these two classes. $\Delta(\Delta H_m)$ relative values are in no case a reflection of the real free energy change, which depends on a variety of other favourable or unfavourable energetic components. In particular, when we considered the entropic loss term ΔS_m, we obtain negative ΔG_m values:

$$\Delta G_m \text{ (class 1)} < \Delta G_m \text{ (class 2)}$$

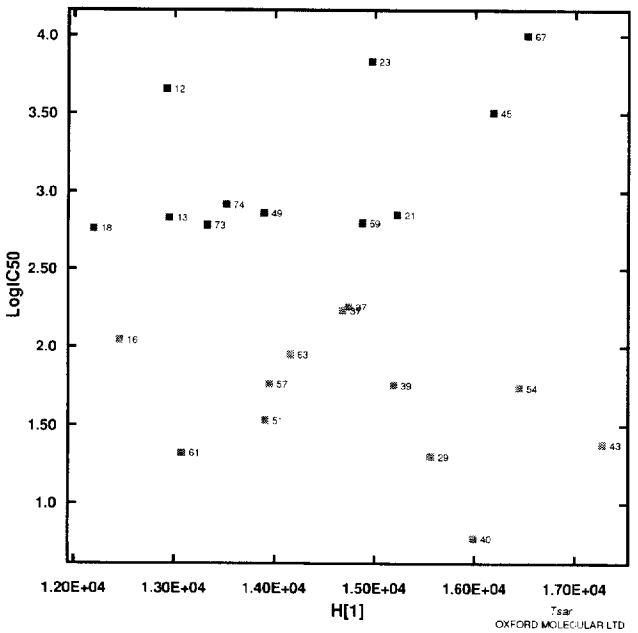


Fig 7. Log IC_{50} vs H [1] for the Thr-model.

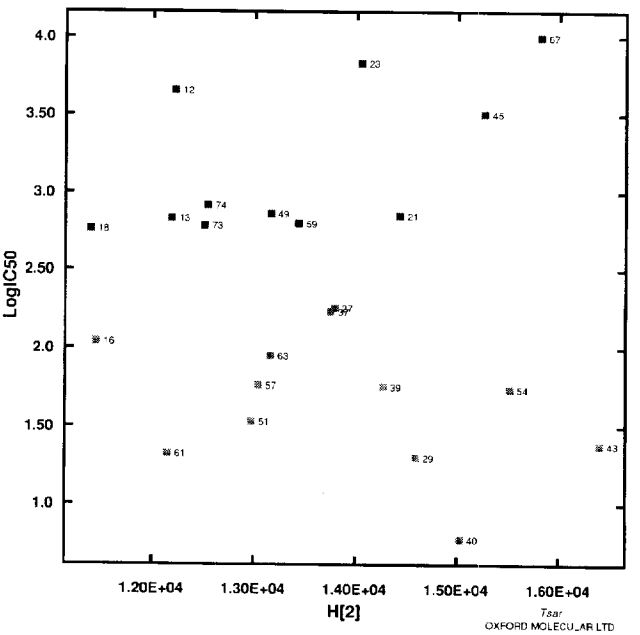


Fig 8. Log IC_{50} vs H [2] for the Thr-model.

which suggests that the exchange reaction is favoured and that the active compounds bind the receptor more strongly than the active ones.

The present results allow the discrimination among the different hypotheses concerning the nature of the residues involved in the recognition process. There is no experimental evidence that Ser binds to the exocyclic nitrogen and that Thr binds to N sp^2. This is why we tested these two docking modes with a methanol molecule as probe. It appears that the Thr model is superior to the other models for clustering compounds into a narrow range of ΔH. These results suggest that the Thr model corresponds to the more likely interaction model between adenosine antagonists and A_1 receptors. These calculations indicate the preponderance of the enthalpic component in ligand–receptor binding. In such a soft non-specific recognition process, most of the water molecules are conserved. The problem of water flush was not introduced into the calculations since normal mode analysis only leads to a conformational component of entropy. Such a method simply reflects the decrease of ligand and receptor site translational and rotational degrees of freedom during the complexation process. It results that the ΔG value does not correspond to the actual free energy change of thermodynamic equilibrium. Furthermore, the energy of the electrostatic interactions between the ligand and protein matrix charge distributions were not included in the ligand binding energy. These simplifications partly explain the lack of linearity between calculated free energy change and affinity [41]. The ideal calculation should take into account all components of ΔG but it would dramatically increase the cost of calculation, especially for the large series of compounds that are investigated in structure–activity studies. Moreover, calculation of the entropic solvent contribution leads to theoretical difficulty in distinguishing between strongly and weakly bound water molecules. It is well known that a water molecule can act as a real partner in the interaction, and not simply as a solvent.

Since our goal was not to accurately reproduce the ΔG values we used the molecular motion contribution to ΔG for discriminating among high- and low-affinity H-bonded ligands. This provides a relatively fast screening test for affinity.

Conclusion

Discriminant analysis based on descriptors usually used in QSAR did not reveal any particular relation between the structures of the nitrogen heterocyclic ligands and their affinities for A_1 receptors. Furthermore, the more discriminant descriptors are not representative of well defined molecular proper-

ties and are not easy to use routinely. Therefore we searched for other types of descriptors which might discriminate among high and low affinity ligands for A_1 receptors. A property that directly reflects ligand–receptor interaction was chosen in order to find a descriptor that was well correlated with the equilibrium affinity constant: the change in the thermodynamic parameters upon complexation equilibrium between ligand and receptor. The monoparametric analysis of log IC_{50} vs ΔH_m applied to the Thr model shows a significative clustering for high affinity A_1 antagonists. Therefore this model constitutes a predictive tool for the affinity of nitrogen heterocyclic A_1 ligands and appears to provide a rational approach for directing the pharmacological profile of compounds towards A_1-antagonistic properties. When the ligand–receptor interaction mode is unknown, it is necessary to put forward an a priori hypothesis concerning the type of residue implied in the recognition process. Otherwise, consideration of the existence of a highly significant discrimination may be envisaged between different hypothetical models as an argument to favour one type of ligand–receptor interaction. Thus the technique described here can be applied to the search and validation of hypotheses on ligand–receptor interaction sites in relation with site-directed mutagenesis studies.

Acknowledgments

This study was supported by a grant from the CNRS (ACC SV13). Thanks are due to C Royer from the University of Wisconsin-Madison for helpful comments. The authors acknowledge a grant of computer time from the CNUSC (F).

References

1 Van Galen PJM, Stiles GL, Michaels GS, Jacobson KA (1992) *Med Res Rev* 12, 423–471
2 Daly JW, Butts-Lamb P, Padgett W (1983) *Cell Mol Neurobiol* 3, 69–80
3 Linden J, Taylor HE, Robeva AS et al (1993) *Mol Pharmacol* 44, 524–532
4 Zhou QY, Li C, Olah ME, Johnson RA, Stiles GL, Civelli O (1992) *Proc Natl Acad Sci USA* 89, 7432–7436
5 Drury A, Szent-Györgi A (1929) *J Physiol (Lond)* 68, 213–237
6 Collis MG, Hourani MOS (1993) *Trends Pharmacol Sci* 14, 360–365
7 Alzheimer C, Sutor B, Ten Bruggencate G (1993) *Neurosciences* 57, 565–575
8 Bertrand G, Nenquin M, Henquin JC (1989) *Biochem J* 259, 223–228
9 Gerber JG, Payne NA (1988) *J Pharmacol Exp Ther* 244, 190–194
10 Speilman WS, Arend LJ, Klotz KN, Schwabe U (1990) *Purines in Cellular Signalling: Targets for New Drugs* (Jacobson KA, Daly JW, Manganiello V, eds) Springer Verlag, New York
11 Jacobson KA, Trivedi BK, Churchill PC, Williams M (1991) *Biochem Pharmacol* 41, 1399–1410
12 Ijzerman ADP, Van Galen PJM, Jacobson KA (1992) *Drug Des Discov* 9, 49–67
13 Wang S, Milne GWA, Nicklaus MC, Marquez VE, Lee J, Blumberg PM (1994) *J Med Chem* 37, 1326–1338
14 Dudley MW, Peet NP, Demeter DA et al (1993) *Drug Dev Res* 28, 237–243
15 MAD 2.20. Oxford Molecular Ltd, Magdalen Centre, Oxford Science Park, Sandford-on-Thames, Oxford OX4 4GA, UK
16 MOPAC, QCPE No 455. Dept of Chemistry, Indiana Univ, Bloomington, IA, USA
17 ASP 3.02. Oxford Molecular Ltd, Magdalen Centre, Oxford Science Park, Sandford-on-Thames, Oxford OX4 4GA, UK
18 TSAR V2.31. Oxford Molecular Ltd, Magdalen Centre, Oxford Science Park, Sandford-on-Thames, Oxford OX4 4GA, UK
19 Sarges R, Howard HR, Browne RG, Lebel LA, Seymour PA, Kenneth Koe B (1990) *J Med Chem* 33, 2240–2254
20 Trivedi BK, Bruns RF (1988) *J Med Chem* 31, 1011–1014
21 Francis JE, Cash WE, Psychoyos S et al (1988) *J Med Chem* 31, 1014–1020
22 Van Galen PJM, Nissen P, van Wijngaarden I, Ijzerman AP, Soudijn W (1991) *J Med Chem* 34, 1202–1206
23 Müller CE, Hide I, Daly JW, Rothenhaüsler K, Eger K (1990) *J Med Chem* 33, 2822–2828
24 Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) *J Am Chem Soc* 107, 3902–3909
25 Richards WG (1983) *Quantum Pharmacology. 2nd edn* Butterworths, 153–154
26 Viswanadhan VN, Ghose AK, Revankar GR, Robins RK (1989) *J Chem Inf Comput Sci* 29, 163–166
27 Hansch C, Leo A (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology*. John Wiley & Sons, NY
28 Wiener H (1947) *J Am Chem Soc* 69, 2636–2638
29 Randic MJ (1975) *J Am Chem Soc* 97, 6609–6615
30 Balaban AT (1982) *Chem Phys Lett* 89, 399–402
31 Hall LH, Kier LB (1992) *Reviews in Computational Chemistry*. Lipkowitz and Boyd, DB
32 Broto P, Moreau G, Vandycke C (1984) *Eur J Med Chem–Chim Ther* 19, 66–70
33 Szafran M, Koput J (1991) *J Comp Chem* 12, 675–680
34 Carbo R, Leyda L, Arnau M (1980) *Int J Quant Chem* 17, 1185–1189
35 Hodgkin EE, Richards WG (1978) *Int J Quant Chem* 14, 105–110
36 Richards WG, Hodgkin EE (1988) *Chem Br* 24, 1141–1144
37 Good AC, Hodgkin EE, Richards WG (1992) *J Chem Inf Comput Sci* 32, 188–191
38 Goldstein H (1950) *Classical Mechanics*. Addison-Wesley, Reading, MA
39 Wilson EB, Delcius JC, Cross PC (1955) *Molecular Vibrations*. Dover Publications Inc, NY
40 MacFarland JW, Gans DJ (1986) *J Med Chem* 29, 505–514
41 Perakyla M, Pakkanen TA (1995) *Proteins: Struct, Funct, Genet* 21, 22–29
42 Honig B, Sharp K, Yang AS (1993) *J Phys Chem* 97, 1101–1109
43 Rauhut G, Clark T (1993) *J Am Chem Soc* 115, 9174
44 Backer EN, Hubbard RE (1984) *Prog Biophys Molec Biol* 44, 97–197
45 Thornton JM, Mac Arthur MW, McDonald IK et al (1993) *Phil Trans Roy Soc Ser A* 345,113–129