Journal of Computer-Aided Molecular Design, 10 (1996) 67–73
ESCOM

J-CAMD 326

Determination of receptor-bound drug conformations by QSAR using flexible fitting to derive a molecular similarity index

C.A. Montanari*, M.S. Tute**, A.E. Beezer and J.C. Mitchell

Chemical Laboratory, University of Kent, Canterbury, Kent CT2 7NH, U.K.

Received 26 August 1995
Accepted 26 November 1995

Keywords: QSAR; Conformation; Molecular similarity; Bisamidines; Pentamidine; Electrotopological state

Summary

Results are presented for a QSAR analysis of bisamidines, using a similarity index as descriptor. The method allows for differences in conformation of bisamidines at the receptor site to be taken into consideration. In particular, it has been suggested by others that pentamidine binds in the minor groove of DNA in a so-called isohelical conformation, and our QSAR supports this suggestion. The molecular similarity index for comparison of molecules can be used as a parameter for correlating and hence rationalising the activity as well as suggesting the design of bioactive molecules. The studied compounds had been evaluated for potency against *Leishmania mexicana amazonensis*, and this potency was used as a dependent variable in a series of QSAR analyses. For the calculation of similarity indexes, each analogue was in turn superimposed on a chosen lead compound in a reference conformation, either extended or isohelical, maximising overlap and hence similarity by flexible fitting.

Introduction

Aromatic diamidines display activity against a variety of viruses, bacteria, protozoa, fungi and even tumours. In particular, pentamidine shows a useful level of activity against *Pneumocystis carinii* and is currently used in treating such infections in AIDS patients. Its toxicity has stimulated a search for improved analogues [1].

Our interest in diamidines was aroused by the activity of pentamidine against leishmania, a disease prevalent in some regions of Brazil. Our starting point was the published activity [2] of a series of 37 analogues against *Leishmania mexicana amazonensis*, and the hope that by application of QSAR methods we might rationalise this activity and thus be able to suggest more fruitful directions for analogue synthesis.

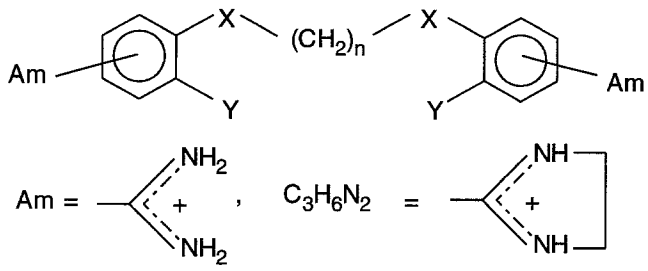
It has been found that pentamidine binds to AT-rich sequences of DNA in the minor groove [3,4]. Moreover, an 'isohelical' conformation has been proposed as a binding necessity [5]. Donkor et al. [6] have studied rigid analogues of pentamidine, and have found that an analogue resembling isohelical pentamidine in shape is preferred for binding to DNA.

Conformational flexibility is normally difficult to deal with by the classical methods of QSAR; indeed, usually a particular conformation (often the most stable, or 'all-trans', conformation) is assumed for the calculation of any property variables. However, by using molecular similarity as a variable, we realised that we may incorporate knowledge on conformation into the QSAR analysis, for similarity depends on comparison of shapes and electrical fields, which are conformation dependent. If we were to choose as the reference the receptor-bound conformation, perhaps similarity indexes would be more relevant, i.e., better correlated to biological activity.

Methods

Structures for pentamidine analogues were 'built' using the modelling package HyperChem [7]. Initially they were built with all-trans linking chains of methylene, ether or amine groups separating the aromatic rings. The amidine

*On leave from Núcleo de Estudos em Química Medicinal-NEQUIM, Departamento de Química, Universidade Federal de Minas Gerais, Brazil.
**To whom correspondence should be addressed.



Compound	n	X	Y	Am
1	2	O	H	*para*
2	3	O	H	*para*
3	4	O	H	*para*
4	5	O	H	*para*
5	6	O	H	*para*
6	3	O	H	*meta*
7	4	O	H	*meta*
8	5	O	H	*meta*
9	6	O	H	*meta*
10	2	O	NO_2	*para*
11	4	O	NO_2	*para*
12	5	O	NO_2	*para*
13	2	O	NH_2	*para*
14	3	O	NH_2	*para*
15	4	O	NH_2	*para*
16	3	O	OCH_3	*para*
17	4	O	OCH_3	*para*
18	5	O	OCH_3	*para*
19	4	O	Cl	*para*
20	5	O	Cl	*para*
21	5	O	Br	*para*
22	3	NH	H	*para*
23	4	NH	H	*para*
24	5	NH	H	*para*
25	6	NH	H	*para*
26	3	NH	NO_2	*para*
27	5	NH	NO_2	*para*
28	2	NH	NH_2	*para*
29	4	NH	NH_2	*para*
30	5	NH	NH_2	*para*
31	6	NH	NH_2	*para*
32	3	O	H	$C_3H_6N_2$
33	4	O	H	$C_3H_6N_2$
34	5	O	H	$C_3H_6N_2$
35	3	O	OCH_3	$C_3H_6N_2$
36	4	O	OCH_3	$C_3H_6N_2$
37	5	O	OCH_3	$C_3H_6N_2$

Fig. 1. Structures for the studied bisamidine compounds.

functions were built coplanar with the rings, and in the protonated form. Atom partial charges were added using the Charge-2 routine within the modelling package PIMMS [8]. Structure refinement was carried out by minimisation using the AMBER force field within HyperChem. For the isohelical conformation of pentamidine as bound to DNA, we used the relevant torsion angles taken from the X-ray structure [4].

Molecular similarity indexes were calculated using the flexible fitting option within the Automated Similarity Program (ASP) package [9,10]. This method permits each molecule to adopt that shape (conformation) which maximises the similarity to a chosen reference molecule, which has been preset to a reference conformation. Both extended (as built and minimised) and isohelical conformations of pentamidine were used as a reference molecule. The Carbo index for similarity was generated, based on the combination of shape and charge [9]. Also, the Carbo index was calculated for shape alone, and for charge alone, to investigate the individual contributions of these two descriptors. Molecular alignment was carried out using the default weighted molecular extent, and flexible optimisation was performed by including torsions of all single rotatable bonds, except those linking amidine groups to the aromatic rings.

For the QSAR analysis, physicochemical and structural descriptors, including similarity indexes, were correlated with biological activity through stepwise multiple regression using the Tools for Structure–Activity Relationships (TSAR) package [11].

Biological data were taken from Bell et al. [2] for the set of 37 bisamidines (Fig. 1) tested against *Leishmania mexicana amazonensis*, and used as log $1/IC_{50}$ (where IC_{50} is in mM) in the regression analyses. Structural variations include the amidine group being either para or ortho to the linking chain, the chain being -X-$(CH_2)_n$-X-, where -X- is either -O- or -NH-, and n = 2–6. Some ortho substituents were also included, and amidines were sometimes substituted by dihydroimidazoles.

Classical physicochemical parameters investigated included log P as calculated by the Hansch–Leo CLOGP routine [12] and reported by Bell et al. [3], and molar refraction and molar volume calculated within TSAR. Because of a likely role of the positively charged amidine functions in charge–charge interactions with the receptor, the electrostatic potential parameters V_{min} and V_{max} were calculated using a routine within PIMMS that gives a maximum potential (found adjacent to the amidinium cation) and a minimum potential (found adjacent to the linker chain of atoms) on the Connolly surface surrounding the extended molecule.

Structural variables investigated were the indicator variables for carbon chain length in the linker chain between aromatic rings, the indicator variables for presence or absence of ether and amine links, and the electrotopological state indexes S(i) for these ether and amine linker atoms, which were calculated using the Molconn-X program [13]. The electrotopological state index has not been much used in QSAR, but as an index that combines the relative polarity of an atom with a correction due to the influence of other atoms on this atom (its environment, or topology), it seemed deserving of consideration in the context of our study.

Results and Discussion

Table 1 shows the biological data and calculated parameters for the analysis. Compound **4** is pentamidine (n = 5, X = O). The most potent compound is **25** (n = 6, X = NH), and the least potent is **10**, which has a nitro group in the ortho position to the ether link and just two methylene groups. Although the number of data is insufficient, this loss of potency with n = 2 is probably due to an inability to span the complete binding site. This is reinforced by the most active compounds, which have n = 5 or 6, and these features appeared in the development of the early QSAR models. For the purpose of analysing these chain-length effects in quantitative terms, compound **10** (IC_{50} > 50 mM) was allowed to take part in the correlation by assigning potency as IC_{50} = 50 mM.

We first used a classical approach to generate Eq. 1, which includes indicator variables to describe the effect of presence or absence of groups or features (as in Free-Wilson analysis) and also includes the standard physicochemical parameters log P and V_{min} (calculated on the extended structures). A correlation matrix (Table 2) was

TABLE 1
PARAMETERS USED IN THE QSARs FOR ISOHELICAL PENTAMIDINE ANALOGUES

Compound	IC_{50} [a]	A [b]	B [c]	C [d]	D [e]	E [f]	log P [g]	V_{min} [h] (kcal/mol)	S(i) [i]	Carbo combined index [j]	Carbo shape index [k]
1	15.822	1.0	0.0	0.0	0.0	0.0	3.30	56.55	5.534	0.896	0.803
2	0.852	1.0	0.0	0.0	0.0	0.0	3.50	49.49	5.593	0.943	0.892
3	1.589	1.0	0.0	0.0	0.0	0.0	4.00	44.00	5.631	0.976	0.934
4	0.820	1.0	0.0	0.0	0.0	0.0	4.60	46.86	5.657	1.000	1.000
5	0.396	1.0	0.0	0.0	0.0	0.0	5.10	37.81	5.677	0.955	0.925
6	6.100	0.0	1.0	0.0	0.0	0.0	3.50	53.86	5.610	0.911	0.844
7	5.435	0.0	1.0	0.0	0.0	0.0	4.00	55.30	5.647	0.912	0.877
8	2.131	0.0	1.0	0.0	0.0	0.0	4.60	48.51	5.673	0.924	0.888
9	1.034	0.0	1.0	0.0	0.0	0.0	5.10	51.61	5.692	0.955	0.906
10	50.000	1.0	0.0	0.0	1.0	0.0	2.80	9.27	5.315	0.868	0.815
11	5.599	1.0	0.0	0.0	1.0	0.0	3.60	−14.81	5.435	0.923	0.859
12	1.997	1.0	0.0	0.0	1.0	0.0	4.10	−13.72	5.469	0.944	0.928
13	22.598	1.0	0.0	0.0	1.0	0.0	1.20	44.75	5.539	0.836	0.815
14	3.503	1.0	0.0	0.0	1.0	0.0	1.40	45.56	5.598	0.917	0.843
15	7.577	1.0	0.0	0.0	1.0	0.0	1.90	46.74	5.636	0.963	0.929
16	1.785	1.0	0.0	0.0	1.0	0.0	2.60	44.99	5.716	0.929	0.857
17	10.048	1.0	0.0	0.0	1.0	0.0	3.10	47.95	5.748	0.954	0.904
18	3.031	1.0	0.0	0.0	1.0	0.0	3.60	30.34	5.770	0.939	0.906
19	1.329	1.0	0.0	0.0	1.0	0.0	5.20	48.46	5.627	0.942	0.909
20	0.703	1.0	0.0	0.0	1.0	0.0	5.70	38.67	5.654	0.939	0.910
21	0.677	1.0	0.0	0.0	1.0	0.0	6.00	39.23	5.753	0.957	0.941
22	0.687	1.0	0.0	0.0	0.0	1.0	2.70	59.14	3.325	0.949	0.900
23	0.671	1.0	0.0	0.0	0.0	1.0	3.20	52.94	3.353	0.967	0.956
24	0.558	1.0	0.0	0.0	0.0	1.0	3.70	50.85	3.372	0.982	0.972
25	0.289	1.0	0.0	0.0	0.0	1.0	4.30	54.69	3.387	0.963	0.906
26	4.828	1.0	0.0	0.0	1.0	1.0	2.20	−10.10	2.937	0.922	0.845
27	1.129	1.0	0.0	0.0	1.0	1.0	3.30	−4.38	3.016	0.963	0.951
28	26.243	1.0	0.0	0.0	1.0	1.0	0.20	56.15	3.205	0.902	0.811
29	7.878	1.0	0.0	0.0	1.0	1.0	0.70	50.05	3.283	0.945	0.928
30	1.160	1.0	0.0	0.0	1.0	1.0	1.30	49.48	3.305	0.983	0.974
31	0.991	1.0	0.0	0.0	1.0	1.0	1.80	43.25	3.322	0.968	0.933
32	1.773	0.0	1.0	1.0	0.0	0.0	5.60	46.79	5.784	0.931	0.926
33	2.710	0.0	1.0	1.0	0.0	0.0	6.10	42.81	5.813	0.929	0.918
34	1.719	0.0	1.0	1.0	0.0	0.0	6.60	40.05	5.832	0.926	0.878
35	2.258	0.0	1.0	1.0	1.0	0.0	4.60	38.06	5.906	0.944	0.881
36	6.415	0.0	1.0	1.0	1.0	0.0	5.10	48.76	5.930	0.942	0.905
37	4.041	0.0	1.0	1.0	1.0	0.0	5.70	33.55	5.946	0.918	0.912

[a] IC_{50} values were taken from Bell et al. [2]. The values used in QSARs are log 1/C.
[b] A: *p*-Am = $(CH_4N_2)^+$.
[c] B: *m*-Am = $(CH_4N_2)^+$.
[d] C: Am = $(C_3H_6N_2)^+$.
[e] D: o-substitution.
[f] E: -NH-.
[g] Taken from Bell et al. [3].
[h] Calculated from PIMMS.
[i] Calculated from Molconn-X.
[j] Charge- and shape-combined Carbo similarity indexes calculated from ASP, 50/50.
[k] Shape index calculated from ASP.

TABLE 2
CORRELATION MATRIX (r) FOR MULTIPLE REGRESSION CALCULATION OF VARIABLES INCLUDED INTO EQ. 1

	X1	X2	X3	X4	X5	X6	X7	X8	Y
X1	1.0	0.329	0.370	0.212	−0.041	0.016	−0.540	0.022	0.075
X2	0.329	1.0	0.040	0.128	0.040	−0.223	−0.380	−0.198	−0.389
X3	0.378	0.040	1.0	−0.016	−0.096	0.180	−0.510	0.075	0.222
X4	0.212	0.128	−0.016	1.0	−0.212	−0.121	−0.393	0.199	−0.680
X5	−0.041	0.040	−0.096	−0.212	1.0	−0.212	0.005	0.055	−0.202
X6	0.016	−0.223	0.180	−0.121	−0.212	1.0	0.087	0.010	0.412
X7	−0.548	−0.380	−0.510	−0.393	0.005	0.087	1.0	−0.307	0.436
X8	0.022	−0.198	0.075	0.199	0.055	0.010	−0.307	1.0	−0.037
Y	0.075	−0.389	0.222	−0.680	−0.202	0.412	0.436	−0.037	1.0

X1 = *p*-amidine; X2 = ortho substitution; X3 = NH; X4: n = 2; X5: n = 4; X6: n = 6; X7 = log P; X8 = V_{min} (kcal/mol); Y = dependent variable = log $1/IC_{50}$.

examined to give an idea of intercorrelation between property and structural variables, and then stepwise regression was used to generate the equation, with the F-statistic for entering or leaving the correlation set to F = 4. Variables examined that never entered the equation were molar refraction (MR) and molar volume (MV), indicator variables for n = 3 and n = 5, and V_{max}. The numbers in parentheses are jackknife standard errors of the estimates of regression coefficients, which we used in preference to the usual t-statistic 95% limits [14].

$$\begin{aligned}\log 1/IC_{50} = {} & 0.528(\pm0.097)I_{p\text{-amidine}} - 0.182(\pm0.059)I_{ortho} \\ & + 0.330(\pm0.091)I_{NH} - 1.026(\pm0.275)I_{n=2} \\ & - 0.314(\pm0.057)I_{n=4} + 0.230(\pm0.117)I_{n=6} \\ & + 0.188(\pm0.056)\log P + 0.004(\pm0.002)V_{min} \\ & - 1.467(\pm0.380)\end{aligned} \quad (1)$$

$(n = 37,\ r = 0.96,\ r^2 = 0.92,\ s = 0.17,\ F = 41.69,\ r^2_{cv} = 0.63)$

Noticeable from this equation is the strong preference for the amidine group to be in the para position, and the deleterious effect of ortho substitution, implying some steric hindrance at the receptor site. The indicator variable for n = 2 is to enter the equation first, and alone will account for over 40% of variance in the biological data. This variable and the significance and sign of indicator variables for n = 4 and n = 6 show that increasing the length of the spacer increases the potency. It is also clear that replacing the -O- by an isosteric -NH- is favourable, and perhaps this indicates a role for the -NH- group as an H-bond donor to base pairs in the AT-rich DNA minor groove.

Although Eq. 1 appears to be a good predictive model, it is limited in three respects. Firstly, it cannot be used to predict any molecules having other than straight chain links between aromatic residues, so it is limited in application. Secondly, the required number of independent variables is more than would be desirable for the number of observations, and indeed three variables are at the limit of acceptable statistical significance, although dropping them from the equation decreases the explained variance. Thirdly, its diagnostic value is compromised, because with the inclusion of indicator variables it is somewhat qualitative. A better understanding of the meaning of the indicator variables was therefore sought.

First of all, we investigated the possibility of replacing the indicator variables by a broader descriptor of topology. Out of many topological indexes, the electrotopological state index S(i) [15–22] was considered for our purpose. The index was calculated for all atoms, but only considered for -O- and -N(H)- moieties of the linker. However, this index never entered the model and Eq. 1 was again produced as the best relationship.

We then calculated the Carbo similarity index using the fully extended conformation of pentamidine, **4**, as the reference structure, and calculated the index on flexible fitting and for a 50/50 combination of shape and charge features [9,23–29]. Figure 2A shows the extended conformation of pentamidine as obtained from minimisation with the AMBER force field in HyperChem. The correlation of potencies including just this index gave Eq. 2:

$$\begin{aligned}\log 1/IC_{50} = {} & 15.360(\pm3.140)Carbo_{extended} \\ & - 14.910(\pm2.980)\end{aligned} \quad (2)$$

$(n = 37,\ r = 0.69,\ r^2 = 0.48,\ s = 0.39,\ F = 32.29,\ r^2_{cv} = 0.43)$

The correlation of Eq. 2 explains less variance than Eq. 1 and it is clear that the sole use of a combined (shape and charge) Carbo index based on the extended pentamidine, though giving a simpler relationship, does not improve the model. This is in line with the reasoning of Benigni et al. [30], who have described similarity indexes as being non-complementary to classical descriptors. However, we believe that Eq. 2 shows the importance of the 3D QSAR descriptor, and as we go on to argue, it may play a very important role in differentiating between conformational features of the molecules.

When faced with conformational mobility, it is reason-

able to base the calculation of conformationally dependent parameters on the most stable conformation; this usually means the all-trans or extended one [31–33]. However, this should only be the starting point in an analysis, and other possible conformations should be examined to identify or explain the poor performance of the model described by Eq. 2. We therefore made another calculation of the Carbo index, but using the isohelical conformation (receptor conformation as judged by X-ray [5]), and thus derived Eq. 3:

$$\log 1/IC_{50} = 12.420(\pm 1.931)Carbo_{isohelical} - 12.040(\pm 1.843) \quad (3)$$

$$(n = 37,\ r = 0.74,\ r^2 = 0.54,\ s = 0.37,\ F = 41.43,\ r^2_{cv} = 0.49)$$

The variance explained by Eq. 3 is still not as good as that of Eq. 1, but it is quite clear that Eq. 3 is a better model than Eq. 2 and that the use of an alternative conformation of pentamidine is an improvement to the model that supports the flexible fitting method [34–38]. Moreover, it is in agreement with the earlier proposal of an isohelical or bowl-shaped conformation of rigid analogues [4,6,39–41]. Figure 2 shows the structures of extended (Fig. 2A) and isohelical (Fig. 2B) forms as well as a relatively rigid analogue (Fig. 2C). This implies that we search molecules with this shape.

We then tried as the reference the conformation of pentamidine as found in the crystal structure of the complex with the dodecamer $d(CGCGAATTCGCG)_2$ [4], which is shown in Fig. 2D. This procedure resulted in Eq. 4, which is somewhat inferior to Eq. 3 but slightly better than Eq. 2:

$$\log 1/IC_{50} = 19.370(\pm 2.650)Carbo_{isohelical(dodecamer)} - 18.810(\pm 2.514) \quad (4)$$

$$(n = 37,\ r = 0.71,\ r^2 = 0.50,\ s = 0.38,\ F = 34.58,\ r^2_{cv} = 0.47)$$

Begnini et al. [30] suggested that the similarity index would be highly dependent on the input conformation of the lead. It seems quite reasonable to assume that the isohelical conformation supported by Eq. 3 is most suitable to describe the receptor conformation for the bisamidine analogues at the receptor for action against *L. amazonensis*.

We next tried the analysis using not pentamidine, but the most active compound in the set, i.e., compound **25**, as the reference. We used both an extended (Fig. 2E) and an isohelical (Fig. 2F) conformation. The isohelical conformation of **25** had been generated by flexible fitting of **25** onto isohelical pentamidine during the derivation of Eq. 3. Equations 5 and 6 describe the QSAR models using the extended and isohelical conformations of **25**, respectively.

$$\log 1/IC_{50} = 13.50(\pm 0.867)Carbo_{extended,25} - 12.93(\pm 0.853) \quad (5)$$

$$(n = 37,\ r = 0.75,\ r^2 = 0.56,\ s = 0.36,\ F = 43.60,\ r^2_{cv} = 0.51)$$

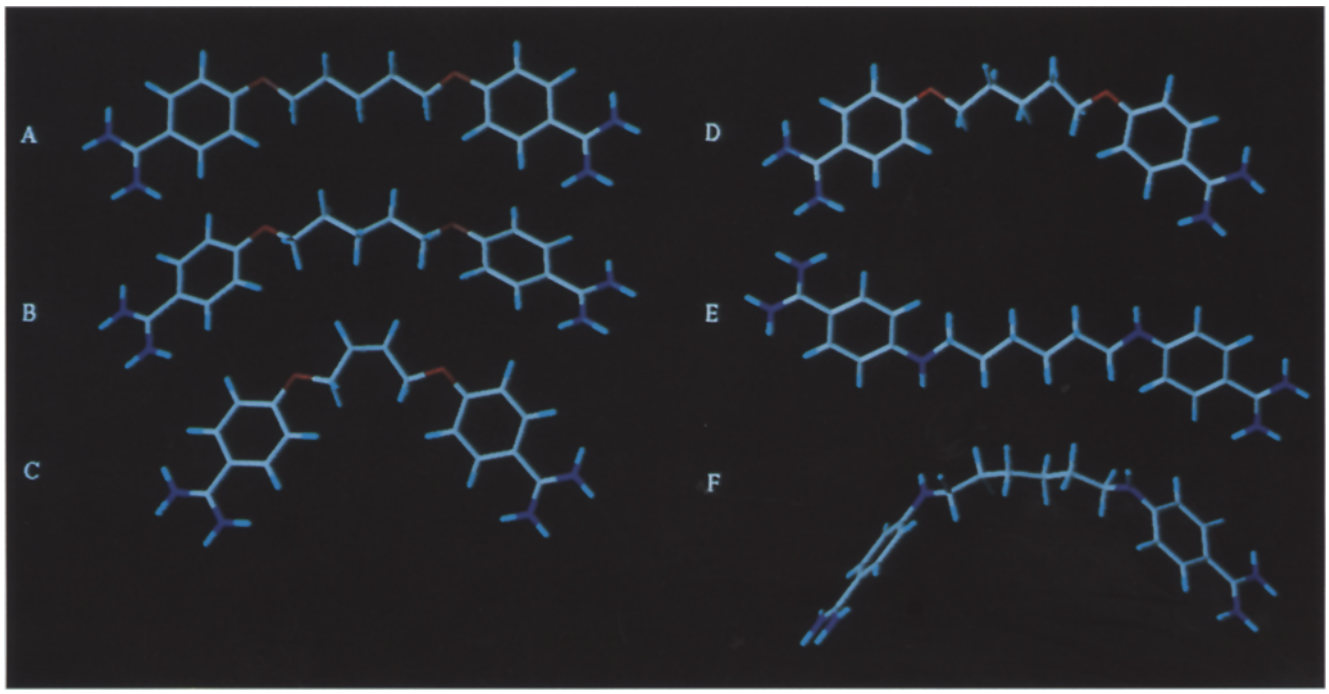


Fig. 2. Illustration of the conformations of some compounds described in the text. (A) Extended pentamidine; (B) isohelical pentamidine as used in Eq. 3; (C) rigid analogue as described in the text; (D) pentamidine structure from X-ray, used to generate Eq. 4; (E) extended compound **25**; (F) isohelical compound **25**; (E) and (F) generated from (A) and (B).

$$\log 1/IC_{50} = 13.60(\pm 0.511)Carbo_{isohelical,25} - 12.90(\pm 0.416) \quad (6)$$

$(n = 37,\ r = 0.81,\ r^2 = 0.66,\ s = 0.31,\ F = 68.97,\ r^2_{cv} = 0.59)$

We then allowed classical physicochemical parameters and our other structural parameters to be sampled by stepwise multiple regression, in addition to the Carbo index of Eq. 6. We sampled the calculated log P, MR, S(i), V_{max}, V_{min}, and indicator variables. However, none of these descriptors was able to satisfy our criterion of $F = 4$ to enter the equation.

We then returned to the Carbo index of Eq. 3, and tried to extend this equation by sampling the other variables. Thus, calculated log P and S(i) both entered at the required significance level to provide Eq. 7, isohelical pentamidine being the reference conformation.

$$\log 1/IC_{50} = 9.543(\pm 2.722)Carbo_{isohelical} + 0.165(\pm 0.031)\log P - 0.145(\pm 0.070)S(i) - 9.220(\pm 2.310) \quad (7)$$

$(n = 37,\ r = 0.83,\ r^2 = 0.69,\ s = 0.32,\ F = 23.94,\ r^2_{cv} = 0.51)$

The coefficient in the electrotopological index S(i) is negative. The value of the index is always larger for -O- than for -N(H)-, and the index thus reflects the fact that isosteric replacement of -O- for -N(H)- always results in an increase of potency, as can be seen by comparison of compounds **2–5** and **22–25** in Table 1. There is a small, linear dependency on lipophilicity. As in many QSAR studies the lipophilicity dependence is frequently parabolic, we allowed a squared term in log P to be sampled. This, however, did not enter the equation.

Note that similarity in lipophilicity is not included in the Carbo index used here, and so lipophilicity appears as a separate descriptor, not encompassed under similarity.

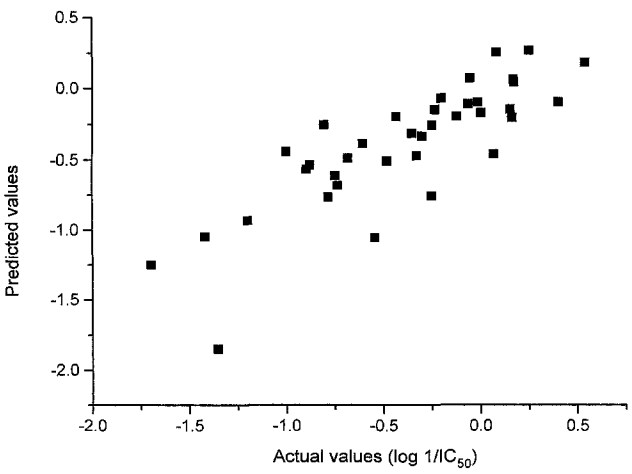


Fig. 3. Plot of predicted (Eq. 7) versus observed IC_{50} values.

Figure 3 is a plot of observed against predicted activities. There is clearly much scatter in the points, but the cross-correlated r^2 of Eq. 7 is reasonable.

The Carbo index, as we have stated, is by default a 50/50 weighting of similarity in shape and in electrostatics. As a final part of this investigation, we envisaged the possibility of finding which of these two factors is the most important by calculating them each separately. The electrostatic index by itself was found to be not significant. The index for shape alone, however, gave the quite reasonable regression expressed by Eq. 8, only marginally inferior to Eq. 7.

$$\log 1/IC_{50} = 5.303(\pm 1.102)Carbo_{isohelical,shape} + 0.167(\pm 0.062)\log P - 0.184(\pm 0.105)S(i) - 4.847(\pm 0.597) \quad (8)$$

$(n = 37,\ r = 0.79,\ r^2 = 0.62,\ s = 0.35,\ F = 17.63,\ r^2_{cv} = 0.35)$

The model expressed by Eq. 7 can be assumed as the 'best' QSAR for the studied compounds. Although giving only a marginal improvement in the explanation of variance (compare Eq. 3 with Eq. 2), a similarity index based on one particular conformation, thought to be the conformation adopted at the DNA receptor, is superior to one based on an extended conformation. We must point out that over the molecules of this set the variation within the Carbo index itself is but small, so one would not expect to find great differences between values for different reference conformations. We therefore present this result in the belief that it may stimulate others to look for more varied series of molecules, where the thesis may be proved that flexible fitting within the ASP program can either determine or support a particular receptor-binding conformation for a member of a set of flexible analogues.

Conclusions

The Carbo molecular similarity index has been evaluated as a QSAR descriptor using electrostatic charge and shape combined, and also as separate indices. The combined index and the shape index were derived by flexible fitting of diamidines to particular reference conformations, and the resulting QSAR equations thus permitted a suggestion for the pharmacophoric conformation. The combined index presents the best description, but shape is the more important component. This shape dependence has been recognised as the most relevant descriptor in biological applications and agrees with the fact that the interaction is primarily steric and to a lesser extent electrostatic. The ASP program is capable of distinguishing these possibilities.

The para-substituted bisamidines with six methylene units in the linker between aromatic rings, but with no ortho substitution (which might hinder receptor binding),

should be used as the basis for design of improved compounds against *L. mexicana amazonensis*. The bioisosteric change of -O- to -NH- gives more potent compounds, probably due to hydrogen bonding to base pairs in the minor helical groove of DNA. It is sensible that an isohelical conformation would fit better into this groove than an extended one. Already, a rigid-analogue approach has been adopted [6,42,43] to synthesize molecules with this preferred shape.

Our study provides (i) an insight into the possibility that binding to the receptor (possibly an AT-rich region of DNA in *L. mexicana amazonensis*) could play an important role in eliciting the potencies described; (ii) evidence that flexible fitting within the ASP program can be used to determine or support the receptor-binding conformation of a flexible molecule; and then (iii) a basis for the rational design of bisamidines with potential for antileishmaniasis activity.

Acknowledgements

C.A.M. wishes to thank the Brazilian Research Council (CNPq) for providing the grants to pursue this work. We thank the referees for their comments on the first draft of the paper, and Oxford Molecular Ltd. for software support and financial assistance.

References

1 Tidwell, R.R., Jones, S.K., Geratz, J.D., Ohemeng, K.A., Cory, M. and Hall, J.E., J. Med. Chem., 33 (1990) 1252.
2 Bell, C.A., Hall, J.E., Kyle, D.E., Grogl, M., Ohmeng, K.A., Allen, M.A. and Tidwell, R.R., Antimicrob. Agents Chemother., 34 (1990) 1381.
3 Bell, C.A., Cory, M., Farilay, T.A., Hall, J.E. and Tidwell, R.R., Antimicrob. Agents Chemother., 35 (1991) 1099 and references cited therein.
4 Edwards, K.J., Jenkens, T.C. and Neidle, S., Biochemistry, 31 (1992) 7104.
5 Lowe, P.R., Sansom, C.E., Schwalbe, C.H. and Stevens, M.F.G., J. Chem. Soc., Chem. Commun., (1989) 1164.
6 Donkor, I.O., Tidwell, R.R. and Jones, S.K., J. Med. Chem., 37 (1994) 4554.
7 HyperChem, Release 2 for Silicon Graphics, Autodesk, Inc., Sausalito, CA, 1992.
8 PIMMS v. 1.46, Oxford Molecular Ltd., Oxford, 1994.
9 Carbo, R., Leyda, L. and Arnam, M., Int. J. Quantum Chem., 17 (1980) 1185.
10 ASP v. 3.11, Oxford Molecular Ltd., Oxford, 1994.
11 TSAR v. 2.31, Oxford Molecular Ltd., Oxford, 1994.
12 Hansch, C. and Leo, A., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, NY, 1979.
13 Molconn-X, v. 1.0 for Macintosh, Hall Associates Consulting, Eastern Nazarene College, Quincy, MA, 1991.
14 Ståle, L. and Wold, S., Prog. Med. Chem., 25 (1988) 291.
15 Kier, L.B. and Hall, L.H., J. Pharm. Sci., 70 (1981) 583.
16 Hall, L.H. and Kier, L.B., Bull. Environ. Contam. Toxicol., 32 (1984) 354.
17 Kier, L.B., Quant. Struct.–Act. Relatsh., 4 (1985) 109.
18 Kier, L.B., Quant. Struct.–Act. Relatsh., 5 (1986) 1, 7.
19 Kier, L.B., Quant. Struct.–Act. Relatsh., 6 (1987) 8.
20 Hall, L.H. and Kier, L.B., Quant. Struct.–Act. Relatsh., 9 (1990) 115.
21 Kier, L.B., Hall, L.H. and Frazer, J.W., J. Math. Chem., 7 (1991) 229.
22 Hall, L.H., Mohney, B. and Kier, L.B., Quant. Struct.–Act. Relatsh., 10 (1991) 43.
23 Good, A.C., J. Mol. Graphics, 10 (1992) 144.
24 Good, A.C., Hodgkin, E.E. and Richards, W.G., J. Chem. Inf. Comput. Sci., 32 (1992) 188.
25 Good, A.C. and Richards, W.G., J. Chem. Inf. Comput. Sci., 33 (1993) 112.
26 So, S.S. and Richards, W.G., J. Med. Chem., 35 (1992) 3201.
27 Good, A.C., So, S.S. and Richards, W.G., J. Med. Chem., 36 (1993) 433.
28 Good, A.C., Hodgkin, E.E. and Richards, W.G., J. Comput.-Aided Mol. Design, 6 (1992) 513.
29 Good, A.C. and Richards, W.G., J. Chem. Inf. Comput. Sci., 33 (1993) 112.
30 Benigni, R., Cottaramusino, M., Giorgi, F. and Gallo, G., J. Med. Chem., 38 (1995) 629.
31 Martin, Y.C., Quantitative Drug Design: A Critical Introduction, Marcel Dekker, New York, NY, 1978.
32 Tute, M.S., Adv. Drug Res., 6 (1971) 1.
33 Tute, M.S., In Hansch, C. (Ed.) Comprehensive Medicinal Chemistry, Vol. 4, Quantitative Drug Design, Pergamon Press, New York, NY, 1990, pp. 1–31.
34 Koshland Jr., D.E., J. Cell. Comput. Physiol., 54 (1959) 245.
35 Koshland Jr., D.E., Adv. Enzymol., 22 (1960) 45.
36 Koshland Jr., D.E., Proc. Natl. Acad. Sci. USA, 44 (1958) 98.
37 Koshland Jr., D.E., Adv. Enzyme Regul., 6 (1968) 291.
38 Koshland Jr., D.E. and Kirtlwy, M.E., Natl. Cancer Inst. Monogr., 27 (1966) 129.
39 Lane, A.E., Jenkins, T.C., Brown, T. and Neidle, S., Biochemistry, 30 (1991) 1372.
40 Sansom, C.E., Laughton, C.A., Neidle, S., Schwalbe, C.H. and Stevens, M.F.G., Anti-Cancer Drug Des., 5 (1990) 143.
41 Jenkins, T.C., Lane, A.N., Neidle, S. and Brown, D.G., Eur. J. Biochem., 213 (1993) 1175.
42 Smissman, E.E., Nelson, W.L., LaPidus, J.B. and Day, J.L., J. Med. Chem., 9 (1966) 458.
43 Armstrong, P.D. and Cannon, J.G., J. Med. Chem., 13 (1970) 1037.