

3D-WHIM Pattern Recognition Study for Bisamidines. A Structure-Property Relationship Study.

Fabiano A. S. Menezes^a, Carlos A. Montanari^{a*} and Roy E. Bruns^b

^aDepartamento de Química, Universidade Federal de Minas Gerais, Campus da Pampulha, 31270-901, Belo Horizonte - MG, Brazil

^bInstituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970, Campinas - SP, Brazil

Um modelo para a interação de bisamidinas aromáticas com o B-DNA foi estabelecido através de estudos de relações estruturas-propriedades derivadas dos cálculos de descritores tridimensionais WHIM-3D. Uma análise de componentes principais, PCA, dos descritores revelou três componentes principais significativas e agrupou as bisamidinas em diferentes conformações: estendida, semi-estendida e semifechada com interações tipo π - π e também por ligações de hidrogênio. O método SIMCA classificou as conformações de acordo com essas características. A interação das 29 bisamidinas estudadas com o B-DNA dá-se através de suas propriedades de forma, distribuição e dimensão.

A proposed model for the interaction of bisamidine analogues with the B-DNA receptor is established by structure-property relationship studies derived from 3D-WHIM descriptor calculations. Three classes, each with relevant information about structural relationships, were determined by PCA and SIMCA analyses for molecular conformations described by 3D-WHIM descriptors for a set of 29 bisamidines with antileishmaniasis and anti-PCP activities. Shape, distribution and dimension properties mostly govern the interaction of bisamidines with B-DNA through the minor groove AT rich regions.

Keywords: 3D-QSAR, principal component analysis, pattern recognition

Introduction

AIDS is a fatal disorder for which no successful chemotherapy has yet been developed. Patients who suffer from this disease are also susceptible to *Pneumocystis carinii* pneumonia, PCP, which leads to a 100% death rate¹. Another *causa mortis* that is seriously widespread in tropical countries, is visceral leishmaniasis, caused by the protozoa *Leishmania amazonensis*². It has been estimated by WHO that over 17 million deaths were due to infectious or parasitic diseases³.

Various compounds have been reported to treat both diseases. Pentamidine, a highly flexible bisamidine analogue, has been found to be useful as anti-PCP⁴ and antileishmaniasis agents⁵⁻⁷. Although it prolongs the life of AIDS⁸⁻¹³ and Leishmania patients^{10,14}, it does exhibit some side effects. For this reason, several other drugs¹⁵ have been tried for treating these patients, but their usefulnesses have not been established, yet.

Since most of the drugs tried so far belong to the same structural class, one common mechanism of action has been accepted for these cationic drugs, that is, interaction with B-DNA through the minor groove AT rich sequences^{16,17,19a}.

This provides good reason to believe that any drug of this class which encompasses DNA isohelicity could have restricted side effects. Figure 1 shows the drug pentamidine: 1A shows the structure obtained from CCDC (Cambridge Crystallographic Data Centre), and 1B represents the preferred isohelical conformation adopted upon binding to B-DNA. Therefore, attempts have been made to find out the *pharmacophoric conformation*¹⁶ within this chemical family that may be more sequence specific and less toxic. As a result, 29 members of this family, Figure 2, were studied in order to disclose the structural features that would lead to the development of a rationale for synthesising new isohelical drugs to B-DNA. Thus, the present paper discusses a structure-property relationship (SPR) investigation undertaken and aimed at obtaining better understanding of the mode of action with the goal of rationalising substituent selection.

Methods

Initial structures for molecules were built and the conformational analysis of each pentamidine analogue, showed in Figure 2, was carried out using HyperChem[®] program. The molecular mechanics-molecular dynamics-molecular

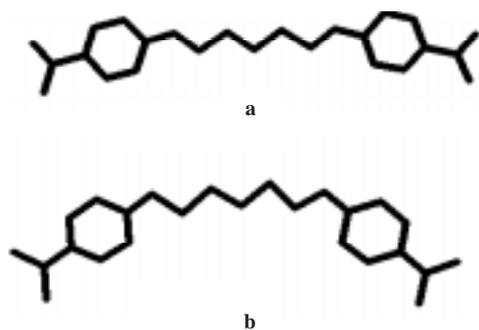
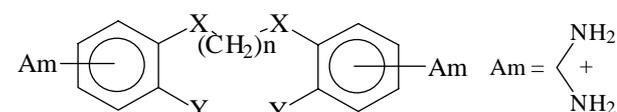


Figure 1. a: X-Ray structure for pentamidine; b: "Bounded" isohelical conformation to B-DNA.



1. $n = 2$, $X = O$, $Y = H$, Am = para position
2. $n = 3$, $X = O$, $Y = H$, Am = para position
3. $n = 4$, $X = O$, $Y = H$, Am = para position
4. $n = 5$, $X = O$, $Y = H$, Am = para position
5. $n = 6$, $X = O$, $Y = H$, Am = para position
6. $n = 3$, $X = O$, $Y = H$, Am = meta position
7. $n = 4$, $X = O$, $Y = H$, Am = meta position
8. $n = 5$, $X = O$, $Y = H$, Am = meta position
9. $n = 6$, $X = O$, $Y = H$, Am = meta position
10. $n = 2$, $X = O$, $Y = NO_2$, Am = para position
11. $n = 4$, $X = O$, $Y = NO_2$, Am = para position
12. $n = 5$, $X = O$, $Y = NO_2$, Am = para position
13. $n = 2$, $X = O$, $Y = NH_2$, Am = para position
14. $n = 3$, $X = O$, $Y = NH_2$, Am = para position
15. $n = 4$, $X = O$, $Y = NH_2$, Am = para position
16. $n = 3$, $X = O$, $Y = OCH_3$, Am = para position
17. $n = 4$, $X = O$, $Y = OCH_3$, Am = para position
18. $n = 5$, $X = O$, $Y = OCH_3$, Am = para position
19. $n = 4$, $X = O$, $Y = Cl$, Am = para position
20. $n = 5$, $X = O$, $Y = Cl$, Am = para position
21. $n = 5$, $X = O$, $Y = Br$, Am = para position
22. $n = 3$, $X = NH$, $Y = H$, Am = para position
23. $n = 4$, $X = NH$, $Y = H$, Am = para position
24. $n = 5$, $X = NH$, $Y = H$, Am = para position
25. $n = 6$, $X = NH$, $Y = H$, Am = para position
26. $n = 3$, $X = NH$, $Y = NO_2$, Am = para position
27. $n = 5$, $X = NH$, $Y = NO_2$, Am = para position
28. $n = 2$, $X = NH$, $Y = NH_2$, Am = para position
29. $n = 4$, $X = NH$, $Y = NH_2$, Am = para position

Figure 2. Structures for compounds investigated in this work.

mechanics, MM-MD-MM, routine in the AMBER¹⁸ force field was used, according to published procedure¹⁹. The simulation temperature was set to 900 K, but the final temperature was 300 K. The simulation temperature was adopted to allow molecules to span a better range of possible conformations. The solvent effect was simulated by the use of a distance-dependent dielectric constant of the form $\epsilon = 4r_{ij}^{-1}$ ¹⁹. All atomic charges are derived from AM1 calculations²⁰. Atom-centred charges for each molecule were calculated and fitted to the entire molecules.

Ninety-eight WHIM²¹ descriptors were calculated and subjected to PCA analysis using the TSAR^{22a}, SIMCA in the Sybyl-QSAR^{22b} and ARTHUR^{22c} packages. WHIM^{23,24}

provide 3D molecular descriptors that are invariant to rotational and translational transformations, thus avoiding molecule alignment problems. WHIM descriptors are able to distinguish different conformations of the same molecule, and thus it seems appropriate for carrying out this study.

Priorities to PCA analysis, the original 3D WHIM data were subjected to a scaling procedure according to average/standard deviation calculations. The averaged values were used for pre-classification using PCA and also for SIMCA calculations.

Results and Discussion

WHIM (Weighted Holistic Invariant Molecular) descriptors are 3D molecular indices that represent different sources of chemical information. They contain information on 3D molecular structure in terms of size, shape, symmetry and atom distribution. The indices are calculated from x , y , z -coordinates of a 3D structure of the molecule, i.e. from a spatial conformation of minimum energy^{23,24}.

The representative conformations for all studied compounds can be found in Figure 3.

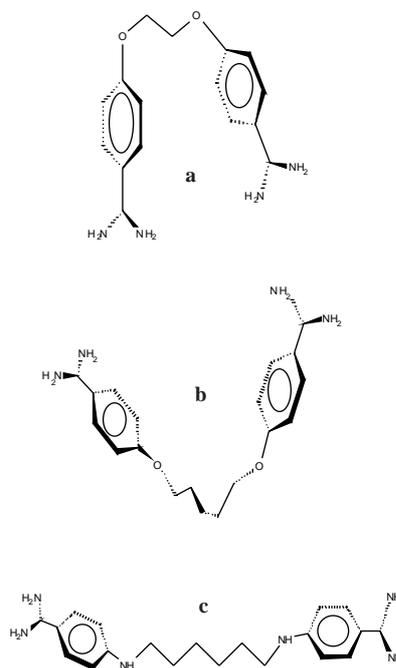


Figure 3. Conformational families identified by WHIM descriptors through PCA and SIMCA. a: π -stacking family 1: analogues 1, 10-12, 17, 26-28. b: semi-extended family 2: 2, 6, 13-16, 20-23. c: extended family 3: 3-5, 7-9, 18, 19, 24, 25 29.

Figures 4 and 5 show the PCA score and loading plots, respectively. Supplementary information on calculated descriptors and their magnitude values can be found either from the authors or the Journal.

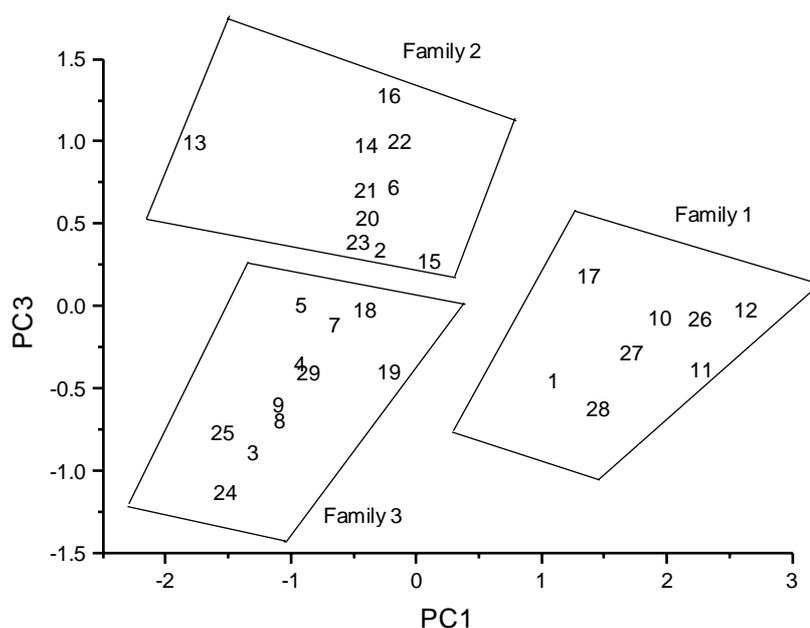


Figure 4. Score plot of PC1 versus PC3. Compounds are numbered according to Figure 2. The contoured lines were manoeuvred to show the so-called families, which were identified by PC1 and PC3 scores.

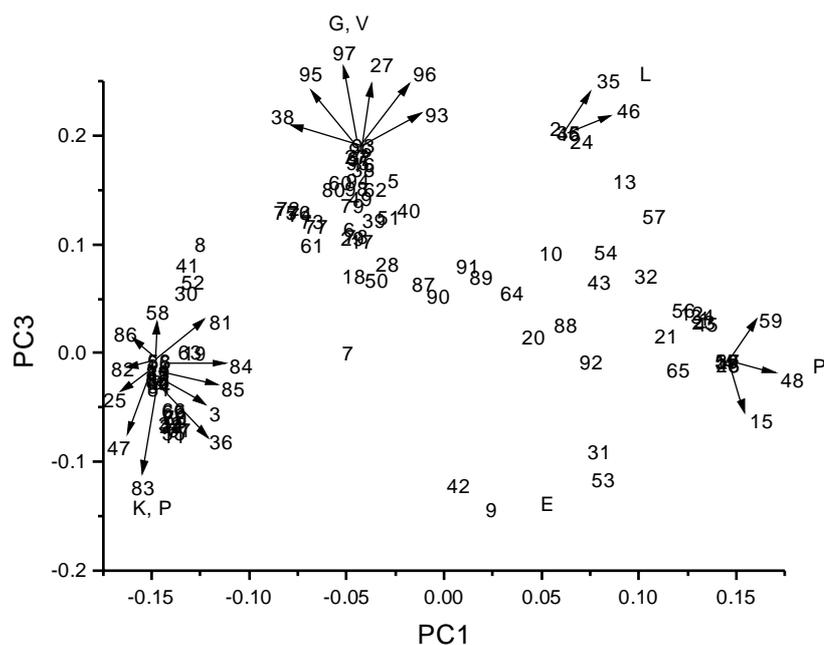


Figure 5. Loadings plot of PC1 versus PC3. This Figure includes all calculated variables in their x, y-coordinates. Note that many of them have similar values either in PC1 and PC3.

The first three principal components explain 46.0, 22.2 and 11.2% of the total data variance. However, the first and third components show the larger discriminating powers in defining three characteristic groups of compounds, which are detailed in Figure 3. The score and loading graphs in Figures 4 and 5, respectively, account for 57.2% of the total variance in the data. The three clusters of molecules in Figure 4 differ owing to the

differing number of carbon atoms in the bridge between the two aromatic amidines as well as to the type of substituent. Thus, they can adopt π -stacking intramolecular and also H-bonding (family 1 in Figure 3), semi-extended (family 2) and extended (family 3) conformations. The SIMCA analysis confirmed this clustering pattern, classifying all molecules in the three groups shown in the PCA score graph within 100%.

Accordingly, it has to be pointed out that the classification showed in Figure 4 is only dealt with 3D WHIM descriptors discriminating power and is, implicitly, in accordance with each conformation, though there may exist some apparent discrepancies in the classification. However, it has no meaning in terms of their conformational energies. Nevertheless, the energy ranges (kcal.mol⁻¹) for all three families are as following: family 1: 28.6-40.8, family 2: 21.4-35.5 and family 3, 27.6-49.9. Thus, it seems, of course, that these values are not capable of separating the three groups. Moreover, the apparent discrepancy, let us say, between compounds 27 and 28 might be better explained taking into consideration the conformations such compounds could adopt. In this case, there is a hydrogen bond between moieties (NH₂ and NO₂) of 1.72 Å, which explains their classifications. After all, molecules with different energies may assume different conformations. In the process of drug-receptor interactions, the receptor may recognise only one out of many. Hence, it is worthwhile mentioning that such conformations may play an important role in the nature of the binding processes. It could also be reasoned that families 1 and 2 are fairly similar. Compound 1, for instance, is classified into family 1 instead of family 2 due to the fact that its conformation is in closer π -contact (3-7Å), which resembles this family. A comparison to its nearest neighbour, compound 15 classified in family 2, where the π -contacts are in longer range (6-10Å), sheds some light into their classifications.

It should be mentioned that although the score graph of the first two principal components explains more variance (68.3%) than the one in Figure 4, it is not capable of discriminating between the three groups of compounds. Only two clusters are evident in this graph (not shown), family 1 and combined families 2 and 3. The separation owes to the discriminating power of the first principal component, PC1, given by Equation 1.

The terms in Equation 1 are the most important ones for PC₁, having loadings with the larger absolute magnitudes as can be seen upon inspection of the loading graph in Figure 5. This principal component is better understood in terms of molecular size and shape. The **P_{1m}** and **K** (that can assume one of the following weights: unit, mass, van der Waals volume, electronegativity, polarizability and electrotopological as holistic measurement) descriptors have the largest absolute magnitude of loadings in Equation 1 and Figure 5. **P_{1m}** and **P_{2m}** are related to molecular shape according to atomic masses^{22,23}. They are directional descriptors that search for the principal axes (spread along orthogonal axes) with respect to the atomic mass properties.

$$PC1 \sim -0.146(Ks) - 0.147(P1v) + 0.145(P2m) - 0.147(P1m) \quad (1)$$

Nevertheless, **K** represents shape within any axis

direction. Thus, these descriptors comprise the eigenvalue proportions for all studied conformations.

Other descriptors showed so far in Figure 5 play a similar trend. **G** describes the symmetry of molecules according to Van der Waals volumes and electronegativity, **V** accounts for all dimensions (unit, mass, Van der Waals, electronegativity, electrotopological and polarizability). **L** stands also for dimension, but as directional WHIM descriptors, as **G** - not **V**, that are non-directional descriptors. **E** is distribution embedded along axes, and it is also directional.

Hence, each "straight molecule" that in this study is represented by isohelicity to B-DNA, as pentamidine itself, Figure 1, should be the choice ones for best binding selectivity.

Equation 2 shows the major descriptors depicted in Figure 5.

$$PC3 \sim -0.146(E2u) + 0.205(L3u) + 0.189(Vu) \quad (2)$$

Finally, it has to be pointed out that there is a correlation between variables depicted in Figure 5. For instance, **K** and **P** are correlated in the range of 0.7-0.9 (r^2); **G**'s are in the range of 0.7. However, this might not be the case for individuals. This means that in the case of selecting variables, those highly correlated may be represented by just one member.

Based upon the above, a simple model can be proposed for the binding of bisamidine derivatives with the receptor, Figure 6. There are three distinct characteristics that appear to be of relevance: (i) the isohelicity to DNA through the "bridge" between the two bisamidine moieties. In this case, compounds that belong to families 2, and certainly 1, would not fit properly into the proposed B-DNA minor groove mode of action for such compounds^{17,19a}; (ii) hydrogen bonding via the amidines, due to the size of molecules. If molecules can adopt family 1 conformations, an intramolecular H-bonding and/or π -stacking interaction would prevent the capabilities of H-bonding formation with minor groove base pairs of B-DNA; and (iii) the alkyl linker lipophilicity. This might be due to the fact that linker's size would suffice for different shape molecules can adopt, and perhaps to any lipophilic interaction into the DNA wells.

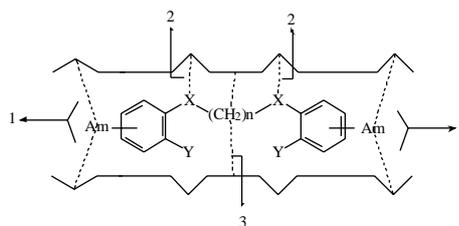


Figure 6. Proposed model for the interactions of bisamidine derivatives with the receptor. 1. Hydrogen bonding between amidine moiety and receptor; 2. Hydrogen bonding within B-DNA walls; 3. Hydrophobic interaction and molecular shape (isohelicity to B-DNA²⁵⁻²⁷).

Conclusions

It seems quite reasonable, therefore, that the above characteristics could induce selective binding by promoting rigidity of analogues. This rationale could lead to less side effects if selectivity is achieved. Nonetheless, the calculation of holistic 3D WHIM descriptors is capable of dealing with conformations that can be envisaged through pattern recognition using PCA and SIMCA. It is, nevertheless, noteworthy that the most prominent descriptors that account for classifying all conformations come from shape, distribution and dimension.

Overall, conformations were explored and classified based on physicochemical descriptors that encompass such 3D information content. This seems to be a very powerful way of dealing with molecules where the search for pharmacophoric conformations may play an important role in the drug design field.

Acknowledgements

We wish to thank CNPq for the grants and Oxford Molecular for providing us with TSAR. Thanks are also due to Prof. Roberto Todeschini, for 3D WHIM.

Electronic supplementary information

Includes definitions and values for the calculated descriptors. Available at: http://www.s bq.org.br/jbcs/2000/vol11_n4/indice.

References

- Gazzard, B. G. *J. Antimicrob. Chemother.* **1989**, *23*, 67.
- (a) Bell, C. A.; Hall, J. E.; Kyle, D. E.; Grogl, M.; Ohemeng, K. A. E.; Tidwell, R. R. *Antimicrob. Agents Chemother.* **1990**, *34*, 1381. (b) Jeronimo, S. M. B.; Teixeira, M. J.; Sousa, A. D.; Thielking, P.; Pearson, R. D.; Evans, T. G. *Clin. Infect. Dis.* **2000**, *30*, 608.
- World Health Report **1997**: <http://www.who.ch/whr/1997/factse.htm>.
- Graves, D. C.; CharyReddy, S.; BeckerHapak, M. *Mol. Cell. Probes* **1997**, *11*, 1.
- Basselin, M.; BadetDenisot, M. A.; Lawrence, F.; RobertGero, M. *Exp. Parasitol.* **1997**, *85*, 274.
- Basselin, M.; Lawrence, F.; RobertGero, M. *Biochem. J.* **1996**, *315*, 631.
- Kandpal, M.; Tekwani, B. L.; Chauhan, P. M. S.; Bhaduri, A. P. *Life Sci.* **1996**, *59*, 75.
- Montgomery, A. B.; Luce, J. M.; Turner, J.; Lin, E. T.; Debs, R. J.; Corkery, K. J.; Brunette, E. N.; Hopewell, P. C. *Lancet*, **1987**, *2*, 480.
- Bielawski, K.; Galicka, A.; Bielawska, A.; Sredzinska, K. *Acta Biochim. Pol.* **2000**, *47*, 113.
- Wispelwey, B.; Pearson, R. D. *Infect. Control. Hosp. Epidemiol.* **1991**, *12*, 375.
- Donkor, I. O.; Queener, S. F.; Dalton, J. T. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1967.
- Santamauro, J. T.; Storer, D. E. *Med. Clin. North Am.* **1997**, *81*, 299.
- Miller, R. F.; Lenoury, J.; Corbett, E. L.; Felton, J. M.; Decock, K. M. *J. Antimicrob. Chemother.* **1996**, *37*, 33.
- Sands, M.; Kron, M. A.; Brown, R. B. *Rev. Infect. Dis.* **1985**, *7*, 625.
- Lee, M. B.; Gilbert, H. M. *Infect. Med.* **1999**, *16*, 37.
- Greenidge, P. A.; Jenkins, T. C.; Neidle, S. *Mol. Pharmacol.* **1993**, *43*, 982.
- Montanari, C. A.; Tute, M. S.; Beezer, A. E.; Mitchell, J. C. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 67.
- Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagano, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
- (a) Montanari, C. A.; Trent, J. O. and Jenkins, T. C. *J. Braz. Chem. Soc.* **1998**, *9*, 175. (b) Orozco, M.; Laughton, C. A.; Herzyk, P.; Neidle, S. *J. Biomol. Struct. Dyn.* **1990**, *8*, 359.
- Guadagnini, P. H. E.; Bruns, R. E. *Quim. Nova* **1996**, *19*, 148.
- WHIM-3D/QSAR version 3.2, Talete srl, Milano (Italy), 1996.
- (a) TSAR, Version 3.21, Oxford Molecular, Ltd.. (b) Sybyl/QSAR, Version 6.3, Tripos, Inc... (c) Scarmenio, I. S.; Bruns, R. E. *Trends Anal. Chem.* **1989**, *8*, 326.
- Todeschini, R.; Gramatica, P. *Quant. Struct.-Act. Relat.* **1997**, *16*, 113.
- Todeschini, R.; Gramatica, P. *Quant. Struct.-Act. Relat.* **1997**, *16*, 120.
- Bostock-Smith, C. E.; Searle, M. S. *Nucleic Acids Res.* **1999**, *27*, 1619.
- Clark, G. R.; Boykin, D. W.; Czarny, A.; Neidle, S. *Nucleic Acids Res.* **1997**, *25*, 1510.
- Neidle, S.; Nunn, C. M. *Nat. Prod. Rep.* **1998**, *15*, 1.

Received: May 12, 1999

FAPESP helped in meeting the publication costs of this article.