

Exploratory analysis of structural properties of DNA-intercalator complexes in crystalline state

M. M. C. Ferreira, R. Kiralj

Instituto de Química, Universidade Estadual de Campinas, Campinas SP, 13083-970, BRAZIL

Keywords: Principal Component Analysis, Hierarchical Cluster Analysis, X-ray structure analysis.

1 Introduction

Intercalator units of DNA intercalators are aromatic, heteroaromatic, conjugated or other π -electron systems that possess significant charge delocalization, elevated polarizability and good electron affinity, and frequently hydrogen bonding groups [1,2]. Such planar systems consist of 2-5 fused rings [3] or a ring coupled with double bonds and some heteroatoms (nitrogen, oxygen [2]). Non-intercalating units of the intercalators, flexible chains and macrocycles, serve to establish covalent and non-covalent interactions with the DNA helix and link two or more intercalating units. It has been shown recently that the stacking interactions between an intercalator and the neighboring base pairs are considerably stronger than the stacking between the two base pairs [4]. Stacking is the major force for intercalation [5], due to which structural changes of the DNA helix occur: additional base-pair separation of 3.4 Å [3,4], helix unwinding and extension [4,6], alterations in the helix geometry and its backbone conformation. The intercalator acts as a rigid body, with little conformational change upon binding to the DNA [7]. This has as a consequence the fact that the binding geometry strongly depends on the properties of the free intercalator, particularly of the intercalating unit, such as shape [3] and planarity [1]. However, the binding geometry seems to be a complex phenomenon, because it is determined also by the nature of base pairs and the intercalator multiple binding modes [1].

This work deals with structural properties of X-ray crystal structures of DNA-intercalator complexes, deposited in the Nucleic Acid Database (NDB) [8]. It is based on previous observations [3] that the DNA helix in such structures forms infinite straight or zig-zag base-paired stacks along crystallographic directions and that the average π -system planes separation distance (including base pairs and intercalating units) can be rationalized in terms of intercalator molecular properties. Structural descriptors of intercalators, DNA oligomers, and DNA-intercalator complexes in crystalline state were generated and analyzed by means of Principal component analysis (PCA) and Hierarchical cluster analysis (HCA) [9]. These analyses were carried out with the purpose to classify intercalated DNA helices with respect to intercalation type (*mono*-intercalators, *bis*-intercalators or *mono*-intercalators covalently bound to DNA), intercalator classes, and other structural features of helical structure and conformation.

2 Methods

2.1 Analysis of crystal structures

All the crystal structures of DNA-intercalator complexes were retrieved from the NDB and analyzed by molecular graphics methods. The crystallographic direction **p** along which the DNA double helices stack to each other frequently intermediated by intercalating units, and along which infinite layers of base pairs and intercalators are formed in a straight or zig-zag manner, was identified. The length *p* of the axis **p** was determined inside the unit cell. The number *I* of intercalating units and the number *A* of base pairs were determined along **p** in the unit cell. The average π -system planes separation distance, the π number, was calculated as $\pi = p / N$, where $N = I + A$ is equal to the number of all π -system planes stacked to each other along the **p** and also to the number of π -system planes separations along the same axis.

2.2 Generation of structural descriptors and exploratory analysis

Two properties of the DNA-intercalator complexes were defined: *I/A* and *I/N* as the intercalating unit frequencies in infinite π -system stacks in crystalline state. Two descriptors for DNA helix were generated from the crystal structures: pairs - the number fraction of bases involved in base pairs or quadruplets; GC% - the number fraction of guanine and cytosine bases in the helix.

The following descriptors for intercalating units of the intercalators were generated from their hydrogen-depleted two-dimensional structures (assuming the planarity of these systems): *Zav* - average atomic number, *Rav* - average van der Waals thickness, and *Score* - a measure of intercalability based on chemical knowledge on intercalators (steric and electronic properties of such systems that cause DNA intercalation).

PCA and HCA (complete linkage) analyses were performed by Pirouette [10] for all helices described by the seven descriptors (*Zav*, *Rav*, *Score*, pairs, GC%, *I/N* and *I/A*). These analyses were done on autoscaled data with the main intention to see what would be the most appropriate classification criteria for intercalated DNA helices: intercalation type (*mono*-intercalators, *bis*-intercalators or *mono*-intercalators covalently bound to DNA), classes of intercalators, and other structural features of intercalated DNA helices.

3 Results and discussion

Total of 98 crystal structures of DNA-intercalator complexes containing 102 symmetrically independent helices were retrieved from the NDB. The DNA oligomer lengths varied from 2 to 22 bases, being complexed with 52 intercalators which belonged to 14 chemical-structural classes (Figures 1-3).

The mean π -systems interplanar separation distance, the π number (Figure 4), ranges from 3.150 to 3.600 Å, with mean 3.346 Å, median 3.335 Å and standard deviation 0.072 Å, being closely related to the value of 3.4 Å for generally accepted base-pairing separation distance. The frequency distribution for π (Fig. 4 left) shows that most helices are concentrated between 3.28 and 3.38 Å. Although the π number depends primarily on the helix geometry and less on crystal packing forces, it seems that the π number depends also on intercalator structures (Fig. 4 right). This is also illustrated by three groups of intercalators in Figures 1-3.

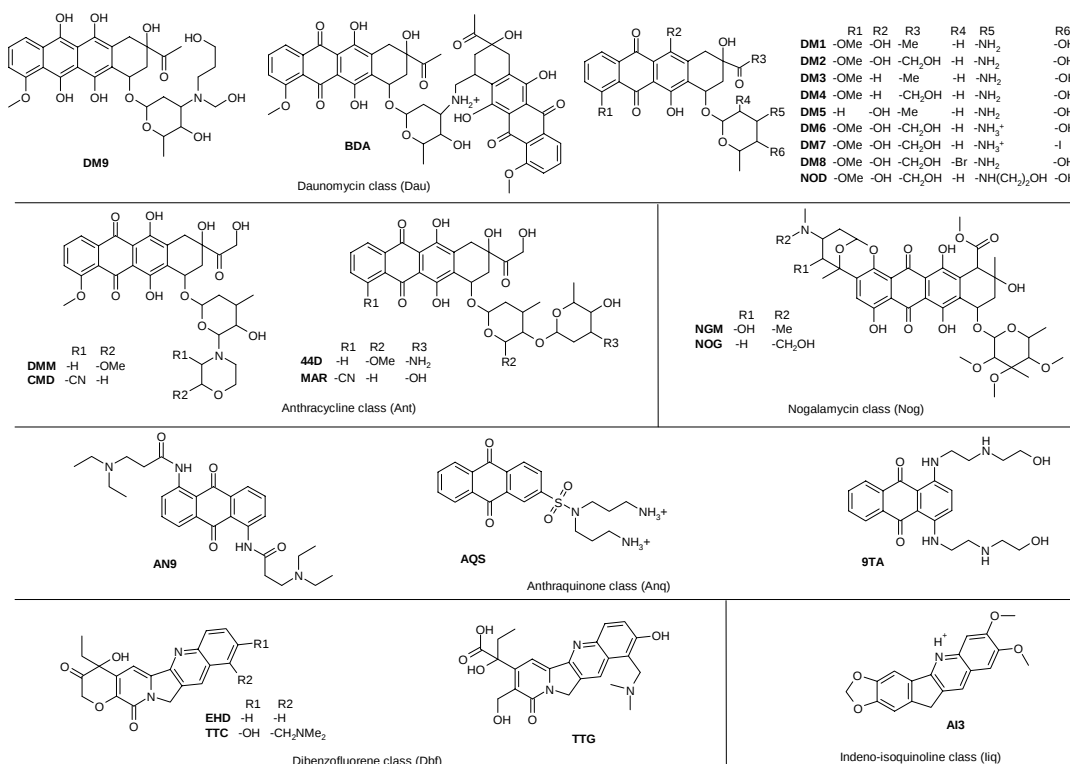
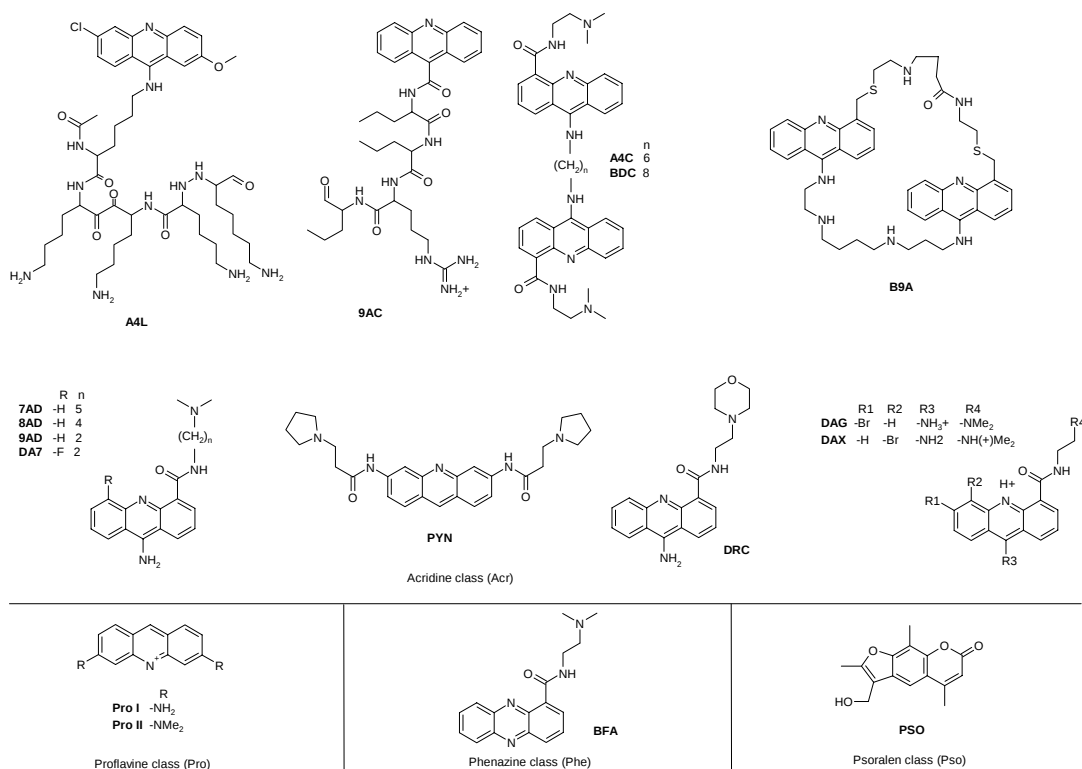
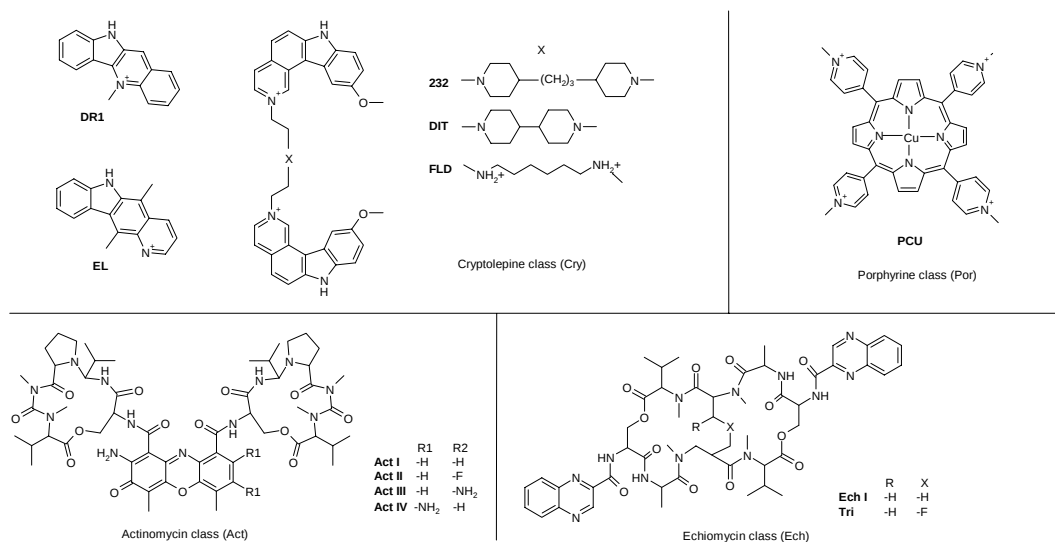


Figure 1 – Tetracycline classes (Dau, Ant, Nog) and other classes that have lower π number values

Figure 2 – Three fused ring intercalator classes with medium π number valuesFigure 3 – Intercalator classes with elevated π number values

The data matrix $X=(102 \times 7)$, consisting of 7 descriptors (Z_{av} , R_{av} , Score, pairs, GC%, I/N and I/A) for 102 helices, was explored by HCA and PCA, and the results are demonstrated in Figures 5 and 6, respectively. Data compression is visible from the PC contents of the total variance: PC1 - 36.2%, PC2 - 27.4%, PC3- 15.3%, PC4 - 9.3%, PC5 - 6.6%, PC6 - 4.9%, PC7 – 0.3%. Both PCA and HCA plots do not exhibit clear separation of *bis*-intercalators and covalently bound intercalators from *mono*-intercalators, although some systematic separation tendencies can be observed.

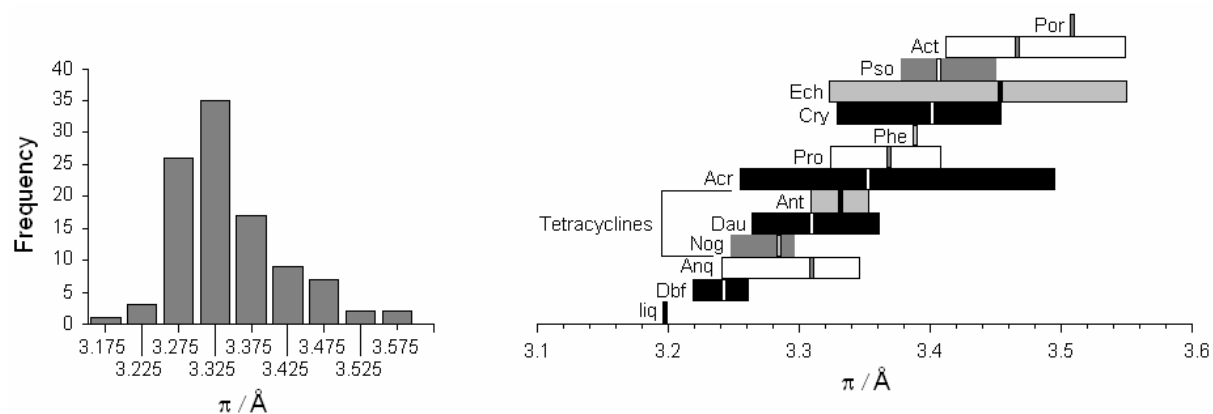


Figure 4 – Statistics on the π numbers. Left: frequency histogram. Right: ranges and means (narrow boxes) for all 14 intercalator classes. In case of one sample per class, only the narrow box is shown.

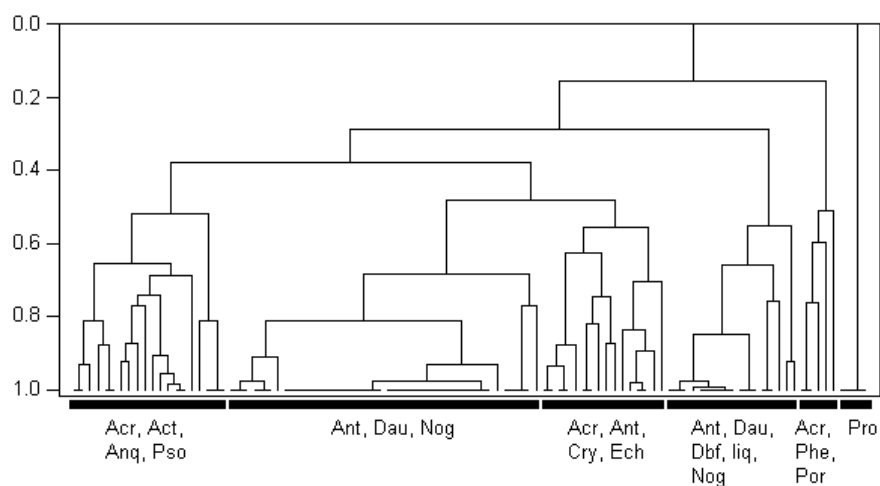


Figure 5 – HCA analysis showing 14 intercalation classes as defined in Figures 1-3

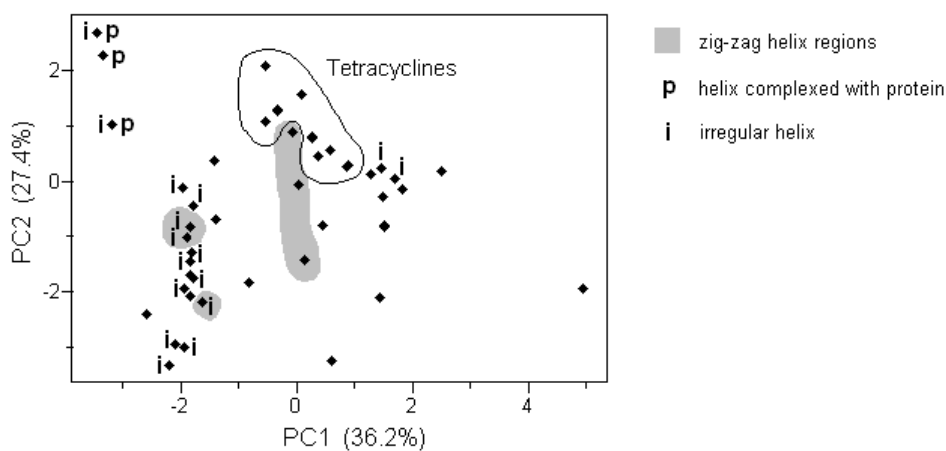


Figure 6 – PC1-PC2 scores plot with some patterns of DNA helix structure

When analyzing the clustering patterns of 14 intercalator classes, it can be observed that some classes are well discriminated from the others, whilst some classes are distributed over two or more clusters in the HCA plot (Fig. 5). Such a separation trend is similar to the π number distribution over classes (Fig. 4 right), where some classes are overlapped due to long ranges of the number π , but the general tendency of the π number dependence on classes is still visible. The scores plot (Fig. 6) shows that three tetracycline classes (Dau – daunomycines, Ant – anthracyclines and Nog – nogalamycines) are the best clustered classes of intercalators. Other interesting features of helical structure can be clearly observed in the plot: helices complexed with both intercalators and proteins are placed at negative PC1 and positive PC2, irregular helices are mainly concentrated at negative PC1, whilst zig-zag helices occupy distinct regions in the left part of the scores plot.

It is interesting to note that among seven selected descriptors, two describe helices (pairs and GC%), two are related to DNA-intercalator complexes (*I/N* and *I/A*), and three exclusively describe intercalating units of intercalators (*Zav*, *Rav*, and *Score*). Whilst the three later descriptors are based on molecular structures of intercalators and chemical knowledge on such compounds, and GC% is derived from chemical composition of DNA oligomers, the exact values of the other descriptors can be determined only from crystal structures of DNA-intercalator complexes. By other words, these descriptors (*I/N*, *I/A* and pairs) are determined mainly by intercalator structures and crystal packing, and less by helical composition.

4 Conclusion

This work has presented preliminary results of chemometric exploration of structural data describing DNA-intercalator complexes in crystalline state. The results suggest further improvement of the methodology by including more structural data and new variable selection. According to the presented results, there is a possibility to distinguish 14 intercalator classes and types of intercalators (*bis*-intercalators, covalently bound intercalators and *mono*-intercalators), and to obtain useful discrimination patterns on helical structure and conformation. Existing relationship between the π number and intercalator classes suggest a possibility for a SPR (Structure-Property Relationship) study.

5 Acknowledgements

The authors acknowledge The São Paulo State Funding Agency (FAPESP) for financial support.

6 References

- [1] B. Keeny, T. Time D. Řeha, M. Kabeláč, F. Ryjáček, J. Šponer, J. E. Šponer, M. Elstner, S. Suhai, P. Hobza. Intercalators 1. Nature of stacking interactions between intercalators (ethidium, daunomycin, ellipticine, an 4',6-diaminide-2-phenylindole) and DNA base pairs. *Ab Initio* quantum chemical, density functional theory, and empirical potential study. *J. Amer. Chem. Soc.*, 124: 3366-3376, 2002.
- [2] R. D. Snyder, J. McNulty, G. Zairov, D. E. Ewing, L. B. Hendry. The influence of *N*-dialkyl and other cationic substituents on DNA intercalation and genotoxicity. *Mutation Res.*, 578: 88-99, 2005.
- [3] R. Kiralj: PhD Thesis. University of Zagreb, Zagreb, 1999.
- [4] T. Kubār, M. Hanus, F. Ryjáček, P. Hobza. Binding of cationic and neutral phenanthridine intercalators to a DNA oligomers is controlled by dispersion energy: quantum calculations and molecular mechanics simulations. *Chem.-Eur. J.*, 12: 280-290, 2006.
- [5] S. A. Kelley, R. E. Holmlin, E. D. A. Stemp, J. K. Barton. Photoinduced electron transfer in ethidium-modified DNA duplexes: dependence on distance and base stacking. *J. Amer. Chem. Soc.*, 119: 9861-9870, 1997.
- [6] M. F. Braña, A. Gradillas, A. Gómez, N. Acero, F. Llinares, D. Muñoz-Mingarro, C. Abradelo, F. Rey-Stolle, M. Yuste, J. Campos, M. Á. Gallo, A. Espinosa. Synthesis, biological activity, and quantitative structure-activity relationship study of azanaphthalimide and aryl-naphthalimide derivatives. *J. Med. Chem.*, 47: (2004) 2236-2242, 2004.
- [7] J. B. Chaires. Energetics of drug-DNA interactions. *Biopoly.*, 44: 201-215, 1997.
- [8] H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, B. Schneider. The Nucleic Acid Database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.*, 63: 751-759, 1992.
- [9] K. R. Beebe, R. Pell, M. B. Sheasholtz: *Chemometrics: a Practical Guide*. Wiley, New York, NY, 1998.
- [10] *Pirouette 3.11*, Infometrix Inc., Woodinville, WA, 2003.