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

Rudolf Kiralj and Márcia Miguel Castro Ferreira. *marcia@iqm.unicamp.br, rudolf@iqm.unicamp.br, http://lqta.iqm.unicamp.br*
Laboratório de Quimiometria Teórica e Aplicada (LQTA), Instituto de Química, Universidade Estadual de Campinas - UNICAMP, Campinas SP, 13083-970, BRAZIL

THE OBJECTIVE OF THIS WORK

X-ray structures of DNA-intercalator complexes deposited in structural databases as Nucleic Acid Database (NDB), Protein Data Bank (PDB), or Cambridge Structural Database (CSD) reveal that DNA oligonucleotides and intercalators form stacks (straight or zig-zag) of base pairs and intercalating units along crystallographic directions, thus mimicking well an infinite DNA helix. This fact might be related to the mechanism of action of intercalator drugs in anti-cancer and anti-microbial chemotherapy. A question that may be posted is: Do intercalator's molecular properties, structure and chemical class to which it belongs determine the binding mode and strength of its intercalation to DNA? This work is a preliminary study dealing with these questions, by means of exploratory analysis of structural data (3 molecular properties of intercalators and 5 crystal properties of DNA-intercalator complexes). Presented results encourage further searching for structural descriptors which can improve the qualitative relationships and discrimination trends found, or even enable some quantitative relationships.

AN ILLUSTRATIVE EXAMPLE OF CRYSTAL STRUCTURE DATA: ACRIDINE BINDING TO A DNA HEXANUCLEOTIDE (PDB: 452D)

The A chain: (5'-D(*CP*GP*TP*AP*CP*G)-3'), a synthetic construct. Ligand: acridine derivative coded 9AD. Crystallographic data: *R* = 20.2%, resolution 1.60 Å, space group $P6_4$. Reference: A. K. Todd, A. Adams, J. H. Thorpe, W. A. Denny, L. P. G. Wakelin, C. J. Cardin, *J. Med. Chem.* 42 (1999) 536-540.

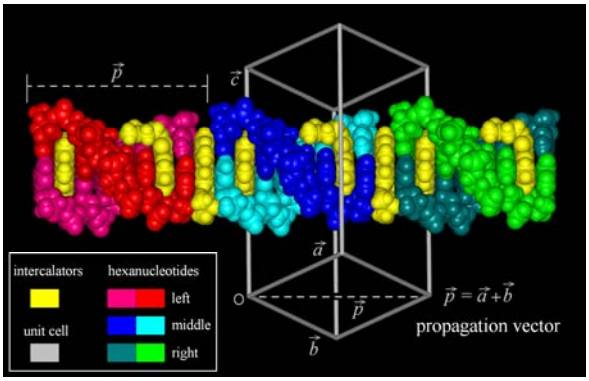

Figure 1. Space filling representation of three DNA hexanucleotides complexed with intercalator molecules in spatial relation to a unit cell. In this structure, the intercalator also serves as a glue between two neighboring hexanucleotides. The virtual "DNA helix" extends along the direction of the propagation vector $\vec{p}$, which is defined as linear combination of the unit cell vectors as $\vec{p} = \vec{a} + \vec{b}$. This vector is the period of the helix repetitive structure.

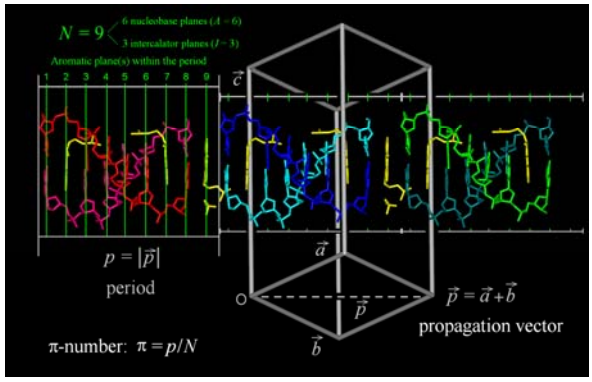

Figure 2. $|\vec{p}|$ is the period to which sets of aromatic planes can be attributed: *N* = 9 plane sets, from which *A* = 6 plane sets of nucleobases and *I* = 3 planes of interacalators. All the planes are nearly parallel and coplanar, at similar separation distance characteristic for aromatic systems. The number π defined as π = *p*/*N* can be considered as a measure of intercalation strength (stronger intercalation ⇒ smaller π) and related to intercalator structure, properties and binding modes. π is the mean separation distance for DNA-intercalator complexes.

14 INTERCALATOR CLASSES FOUND FOR 102 SYMMETRICALLY UNIQUE DNA OLIGONUCLEOTIDE HELICES IN 98 CRYSTAL STRUCTURES (NDB DATABASE)

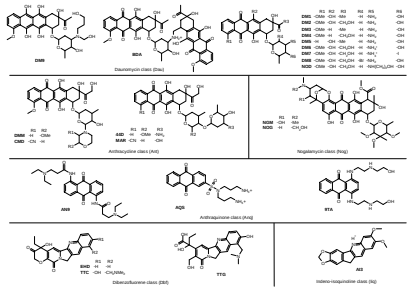

Figure 3. Tetracycline classes (Dau, Ant, Nog) and other classes that have lower π number values.

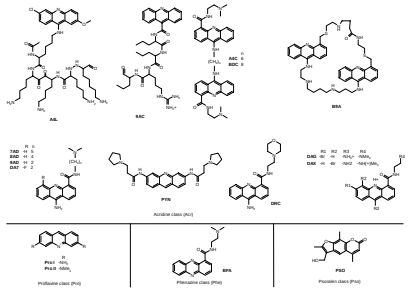

Figure 4. Three fused ring intercalator classes with medium π number values.

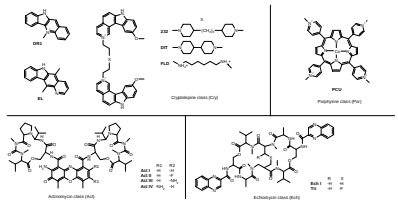

Figure 5. Intercalator classes with elevated π number values.

π NUMBER STATISTICS AND CORRELATION WITH STRUCTURAL DESCRIPTORS

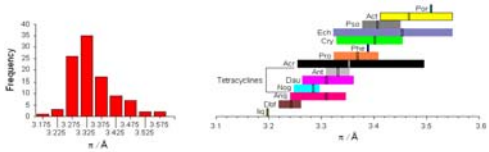

Figure 6. Left: The π number distribution showing significant variations: range 3.150-3.600 Å, mean 3.346 Å, median 3.335 Å, standard deviation 0.072 Å. Right: ranges and means (narrow boxes) for all 14 intercalator classes. In case of one sample per class, only the narrow box is shown. Conclusion: there is a certain dependence of the π number on intercalator structures.

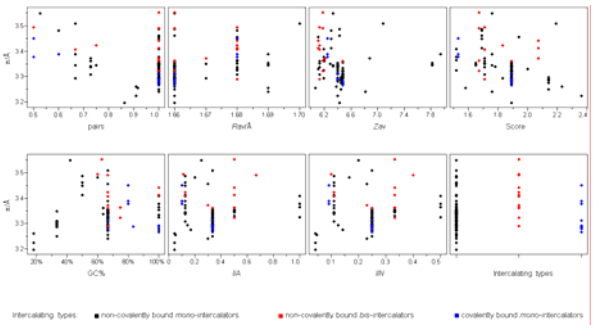

Figure 7. Correlation of the π number with other structural descriptors and three intercalating types. Four intercalated helix descriptors from crystal structures: pairs – number fraction of bases involved in base pairs or quadruplets; GC% - number fraction of guanine and cytosine bases; ratios *II*A and *II*N (see Figures 1 and 2 for definition). Three theoretical descriptors for intercalators: *Z*av – average atomic number of the intercalating unit; *R*av – average van der Waals thickness of the intercalating unit; Score – measure of the intercalator's intercalability based on chemical knowledge (steric and electronic properties of intercalators relative to cryptolepine). Conclusions: some descriptos show modest correlation with the π number: pairs, *R*av, score, GC%, and even the intercalating types.

ACKNOWLEDGEMENT:

FAPESP

EXPLORATORY ANALYSIS

DATA: autoscalled matrix X(102, 7) with 102 helices and 7 descriptors (*Z*av, *R*av, Score, pairs, GC%, *II*N and *II*A)

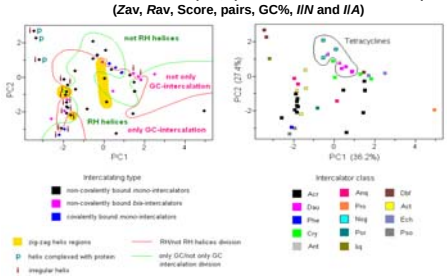

Figure 8. PC1-PC2 scores plot with some patterns of intercalated helix structures. A) helix types: RH (reverse Hoogsten double helix) *vs.* other types. B) presence *vs.* absence of proteins; C) zig-zag *vs.* straight helix propagation; D) regular *vs.* irregular helices; E) GC *vs.* other intercalation sites; E) tetracyclines *vs.* other classes. Intercalation types (*mono-* and *bis*-intercalations, covalent and non-covalent) are not distinguished well.

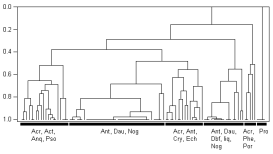

Figure 9. HCA with complete linkage. Some systematic tendencies in separation of the 14 intercalation classes can be noticed. The three intercalation types are not distinguished.