J Mol Model (2001) 7:334–342
DOI 10.1007/s008940100048

ORIGINAL PAPER

Vernon G. S. Box · Fleumingue Jean-Mary

The roles of aromaticity in the structure of DNA and its intercalation complexes with quinones

Received: 8 May 2001 / Accepted: 19 July 2001 / Published online: 8 September 2001

Abstract The X-ray crystallographic coordinate data of a 56 DNA double helical oligomers were examined, using the molecular modeling program STR3DI32.EXE, in order to ascertain the aromatic statuses of the Watson–Crick hydrogen bonded base pairs. Several oligomers that were intercalated with anthraquinoid molecules (like the daunomycin and nogalamycin aglycones) were also included in the study in order to evaluate the aromatic statuses of the intercalated entities. This study revealed that the base pairs were aromatic in their Watson–Crick hydrogen bonded double helices, whereas they are known to be non-aromatic in situations in which they are not involved in Watson–Crick hydrogen bonding. The resonance energy gained by the aromatization of these bases, while engaged in Watson–Crick hydrogen bonding, must contribute to the stability of these DNA double helices. The anthraquinoid intercalates were revealed to be in their radical anion form, having received an electron from one of the bases between which these intercalates were sited. These anthraquinoid intercalates are therefore "held" in position by ionic – charge transfer – interactions, as well as hydrogen bonding due to their glycosidic entities. These observations are also relevant to investigations of the electrical conductivity of DNA double helices that are similarly intercalated.

Keywords DNA · Stability · Heteroaromaticity · Intercalation · Anthraquinone

Introduction

The DNA double helices

The DNA double helical polymers and RNA are undoubtedly unparalleled in their biological importance and have been the subjects of many studies. Biochemists and molecular biologists have struggled to decipher the information encoded in these polymers, their enzymatic activities, and their involvement in the construction of important biochemicals. Physical and organic chemists have been puzzled by other aspects of these polymers, especially the factors that contribute to the formation of these stable double helices and their ability to intercalate some molecules.

We have tried to ascertain, by thorough and detailed examinations of their X-ray crystallographic data, whether the DNA bases, as their simple derivatives and in non-Watson–Crick hydrogen bonded nucleosides, were aromatic, or not. [1] That study clearly suggested that only adenine was aromatic in its ground state. The study also clearly showed that there are marginally aromatic molecules, like imidazole, that could be "induced" into displaying significant aromaticity if they were simply protonated. [1] We also encountered non-aromatic molecules that do become aromatic when allowed to experience multiple hydrogen bonding, like the 2-pyridone dimer. [1]

Since the Watson–Crick pairing of bases relies on the establishment of two, or three, hydrogen bonds between the participants, we wished to find out if the DNA bases do become aromatic when they are present in their Watson–Crick hydrogen bonded, double helical oligomers/polymers. We have therefore examined the high resolution X-ray crystallographic atomic coordinate data of 56 DNA double helical oligomers, a few of which were intercalated with anthraquinoid entities (like those found in the daunomycin [2] and nogalamycin [3]).

Delocalization, resonance and aromaticity

Delocalization and resonance are among the most powerful and widely used concepts in organic chemistry. For many years organic chemists have assumed, often without the support of experimental data, that any planar conjugated π-system that can be represented by a delocalized structure, and hence might be capable of reso-

V.G.S. Box (✉) · F. Jean-Mary
Department of Chemistry,
City College of the City University of New York,
Convent Avenue @ 138th Street, New York, NY 10031, USA
e-mail: box@scisun.sci.ccny.cuny.edu
Tel.: 212-650-8266

nance, must indeed become delocalized and hence stabilized (or destabilized) by resonance. Thus, organic chemists have assumed that all molecules that possess the characteristics that should enable them to obey Hückel's Rules for aromaticity, must indeed, in their ground states, be truly delocalized, be resonance stabilized, and hence be aromatic. This assumption has become a "rule" that can only be tempered by the existence of structural or stereochemical factors that would prevent delocalization, or if the accompanying energetic consequences of delocalization would obviously and undoubtedly be severely unfavorable.

Molecular modeling and theoretical organic chemistry have also become powerful forces in modern organic chemistry. They are widely used and highly respected, but, as will be discussed below, they have also contributed to some of the erroneous ideas and applications of the three concepts highlighted above. This is especially true in the area of computational theoretical organic chemistry.

Molecular modeling is currently hampered by the fact that, until recently, no molecular modeling program had been developed with the ability to construct its own atom connectivity and bond type/order list, by itself, only from existing coordinate data. In practice, the user of the molecular modeling program must construct both lists, often without reference to the coordinate data, and then the program uses this connectivity data to "show" the user the molecular structure that is embedded in this connectivity data. This is especially true for bond type/order assignments that eventually make a profound impact on the user's understanding of π-electron delocalization patterns in the molecule being examined.

Consequently, there are many organic molecules, whose structures were determined by X-ray crystallography, but whose intriguing chemical and physical properties have not been recognized, because their X-ray crystallographically determined structures are "masked" by erroneous connectivity and bond type/order data. Quite often, the X-ray crystallographers that use molecular modeling programs during their structure refinement processes do not recognize errors in their bond length–bond type data, because these molecular modeling program shows these X-ray crystallographers what they have embedded in the connectivity list.

The theoretical (computational) treatments of conjugated π-systems have also allowed us to ignore instances in which molecules disobey the hallowed rules of delocalization, resonance and aromaticity. The popular molecular orbital theoretical methods allow bonding interactions over very large distances, much greater than those bonds of the same types whose parameters have been determined from the diffraction studies. For example, while there are no instances in which an isolated C=C (carbon–carbon double) bond has ever been shown by experimental diffraction methods to exceed 1.4 Å in length, we often see π-like bonding interactions being invoked in theoretical simulations over distances that are often considerably longer than 1.53 Å, the length of a simple isolated C–C (carbon–carbon) single bond.

To make matters worse, most theoreticians (including those that use the molecular orbital methods) also use molecular modeling programs that require user-generated connectivity and bond type/order data, and hence, regrettably, these theoreticians construct and simulate only the structures that they wish to.

We have tried to draw attention to some of the more glaring instances in which molecules that seem to be capable of delocalization do not, in fact, experience delocalization, or aromaticity. [1, 4] This area of organic chemistry, with its several logical contradictions, often sees discussions of simple molecules (like the free rotation about the C–C bonds of 1,3-butadienes, and the carbonyl groups of the pyridones and the anthrones) turn into feasts of speculation and confrontation. Indeed, given the current trend in which some theoreticians ascribe all of the properties of molecules to molecular orbital interactions, we have recently suggested that theoreticians first perform a detailed study of the dipolar and stereoelectronic effects within a molecule in order to determine which of these effects can fully, or partially, rationalize the molecule's chemical and physical data. Then, if the dipolar molecular simulation still falls short, the theoreticians should ascertain the natures of the molecular orbital interactions involved. [5]

STR3DI32.EXE and QVBMM

During the development of the molecular mechanics force field – QVBMM – and the molecular modeling program – STR3DI32.EXE – begun in 1986, the bond length data, available from diffraction (mainly X-ray crystallographic structural) studies, of many thousands of organic molecules were carefully analyzed. It was shown that the bond lengths of a conjugated π-system can unequivocally signal the presence, or absence, of delocalization in that π-system. The resulting new algorithms that correlate bond lengths to bond types were incorporated into STR3DI32.EXE and the QVBMM molecular mechanics force field. [6]

To date, these bond length–bond type/order algorithms have never incorrectly identified the type of any bond found in a molecule whose structure was determined by any of the diffraction methods. This remarkably successful atomic coordinate data analytical method has therefore enabled STR3DI32.EXE and the QVBMM molecular mechanics force field to become the first molecular modeling programs to reliably, automatically and independently, generate their own connectivity and bond type/order data solely from the atom type and atomic coordinate data supplied to them.

The STR3DI32.EXE molecular modeling program enables users to either corroborate their views of the bonding patterns in a molecule whose atomic coordinate data were reliably (experimentally by a diffraction method) determined, or to recognize the existence of unforeseen patterns of bonding or delocalization in that molecule. Thus, STR3DI32.EXE is an extremely power-

ful tool for the analysis of the structural features of organic molecules whose coordinate data were obtained from diffraction studies. The use of experimentally obtained data in these studies should help to remove most reservations about the integrity of the conclusions arrived at, whereas the use of theoretically generated structural data will always raise questions and controversy.

Isolated, localized, carbon–carbon double bonds, C=C, normally have lengths very close to 1.33 Å, the "perfectly" delocalized C–C bonds in simple benzenoids normally have lengths close to 1.390 Å, and the "perfectly" delocalized C–N bonds in simple pyridinoids normally have lengths close to 1.336 Å. The localization of a C=C bond in a conjugated π-system will cause that bond's length to be close to that of an isolated C=C bond, 1.33 Å. This localized C=C bond will experience a reduced efficiency of orbital overlap with the rest of the flanking π-bonds, and so reduce the efficiency of delocalization in that π-system. Similarly, C=C bonds in a conjugated π-system that are longer than the delocalized C=C bonds of benzene will also reduce the efficiency of orbital overlap and so reduce the efficiency of delocalization in that π-system.

The experimentally determined atomic coordinate data of any molecule that contains a non-aromatic, but potentially delocalizable, π-system can be examined by STR3DI32.EXE and the existence, or absence, of a truly delocalized π-system in that molecule demonstrated. Since the atomic coordinate data for most simple π-systems can be obtained from existing diffraction studies of close molecular analogues, these data can be used as the basis of an independent (from NMR etc.) method for the establishment of the existence of true delocalization and aromaticity in the common π-systems.

It is quite remarkable that the lengths of truly delocalized bonds in non-aromatic π-systems are almost identical with those in the simple classically-aromatic systems like benzene, thiophene and pyridine. [1, 4, 7] Truly delocalized amidic C–N bonds are almost identical in length with the C–N bonds of pyridine, and the bond lengths in truly delocalized carbocations and carbanions are almost identical with those of benzene. This is very much what one would expect since the multiple bonds are much more rigid than single bonds, and are not easily distorted by steric/electronic factors that do not directly involve their bonding.

The detection of aromaticity by molecular modeling

If indeed there is true global delocalization in an aromatic entity, then the bond lengths and the overall geometry of that moiety will unequivocally reflect this situation. The molecular modeling program STR3DI32.EXE uses a sophisticated atomic coordinate data analytical algorithm, based on the lengths of the bonds in the conjugated π-systems, to indicate whether that π-system is delocalized, or not. If the π-system is cyclic, then STR3DI32.EXE will determine whether the detailed geometry of that π-system allows it to be aromatic, or not. STR3DI32.EXE also uses the experimentally determined mean bond length deviation (the RMSD that is estimated and reported in the X-ray crystallographic data) in its analysis, in order to accommodate bonds whose lengths are greater than the lengths of truly delocalized bonds, but are within the limits of this error. Since the X-ray crystallographic bond length error can be included in the bond length analysis, even diffraction data obtained at lower resolution can be used in a meaningful fashion. Thus, STR3DI32.EXE has facilitated the detailed structural analyses of molecules that have been the subjects of high resolution diffraction atomic coordinate data.

STR3DI32.EXE assesses the degree of aromaticity of an aromatic π-system based on the extent to which the lengths of the bonds in that π-system vary from those in the classical aromatic molecules pyridine and benzene. Thus the Delocalization Index (DI) of an aromatic molecule, defined below, is an expression of the extent to which the bond lengths in that molecule are distorted from their "ideal lengths". This index provides evidence for, and a method of comparison of, the reduced delocalization in that π-system, and hence the reduced aromaticity.

$$\text{Delocalization Index (DI)} = \frac{\sum_1^{Nb} \text{Abs}\,(d - d0) \times 100}{N_b}$$

As usual, d is the length of a given bond in the π-system, d_0 is the length of the same type of bond in the classically-aromatic π-system, and N_b is the total number of delocalized bonds in that π-system. Thus, the DI measures the mean deviation of the bond lengths of that π-system from those of a classically delocalized aromatic molecule. The multiplier (100) simply converts the bond length data, normally given in ångstrøms in STR3DI32.EXE, to picometers (pm).

The DI of simple benzenoids, or simple azabenzenoids, should be close to zero, as too should be that of any "perfectly" delocalized aromatic entity. However, as the bond lengths in the delocalized π-system depart from their ideal values, so too will the efficiency of electron delocalization decrease, and so too will the DI of that aromatic entity increase. Aromatic systems that have large DIs will obviously be weakly aromatic, while those with small DIs will be strongly aromatic. The maximum observed DI for a very weakly aromatic aza-heterocyclic is approximately 5.3.

The minimum DI observed for benzenoids from high resolution X-ray crystallographic atomic coordinate data seems to be about 0.8±0.2 depending on the quality of the study (RMSD value) and the effective resolution achieved. For example, we occasionally encounter X-ray crystallographic atomic coordinate data of simple benzenoids (like unhindered benzoate esters) that yield DIs as high as 3.0, and this high DI indicates that the effective resolution (and analysis of the data) achieved in that diffraction study was quite low.

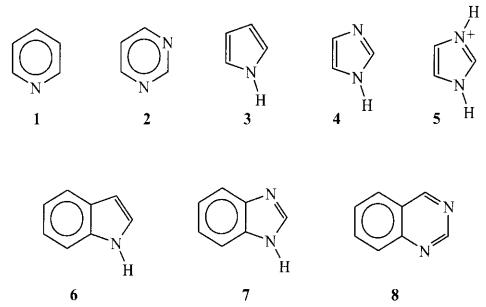


Fig. 1 Chart of monocyclic heterocycles and benzo-heterocycles examined in [1]

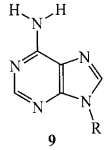


Fig. 2 Chart for adenine

The aromaticity of the common aza-aromatics

The atomic coordinate data for the simple monocyclic heterocycles have been examined in order to compare their aromaticities with those of the archetypal aromatic molecules benzene and pyridine. [1] It was shown that pyridine **1** and pyrimidine **2** were "strongly" aromatic molecules, that pyrrole **3** was "weakly" aromatic and that imidazole **4** (see chart in Fig. 1) was borderline on being non-aromatic. [1] The imidazolium ion **5** was, however, as aromatic as pyrrole **3**, and it was suggested that the increased aromaticity of the imidazolium ion **5** is a contributing factor to the significant basicity of imidazole **4**. [1]

When these monocyles were "fused" with benzene to provide the benzo-heterocyles **6–8** (shown in the chart in Fig. 1), the bond length data showed a marked reduction of the aromaticities of the heterocyclic moieties. This clearly refutes the notion that the "fusion" of a strongly aromatic entity to a weakly aromatic, or anti-aromatic, entity will enhance the stability/aromaticity of that moiety, and instead points towards the opposite trend. For example, the central bonds of dibenzocyclobutadiene are just as long as simple C–C single bonds, so showing that the molecule has resolved (energetically) that anti-aromaticity issue by become non-aromatic. [1]

Adenine **9** (R=H) (chart in Fig. 2) can be regarded as a fused heterocycle composed of the strongly aromatic pyrimidine and the borderline aromatic imidazole, and the bond length data for adenine show it to be weakly aromatic, at best, in the imidazole ring. The X-ray crystallographic data for the other DNA bases and their simple nucleosides clearly show that, with the sole exception of adenine, these bases are non-aromatic. [1]

The other DNA bases are quite different from adenine in that they all possess at least one amidic carbonyl group. The presence of the carbonyl group has a considerable effect on the molecular structure, the possibility of these bases being aromatic, and the types of circumstances in which they could become aromatic. The carbonyl group is much stronger than a carbon–carbon double bond, which in turn is stronger than a carbon–nitrogen double bond. Some of these DNA bases will hence show the tendency to be in the "carbonyl" form rather than the "ene" or "imine" form". Indeed this trend is well established since:

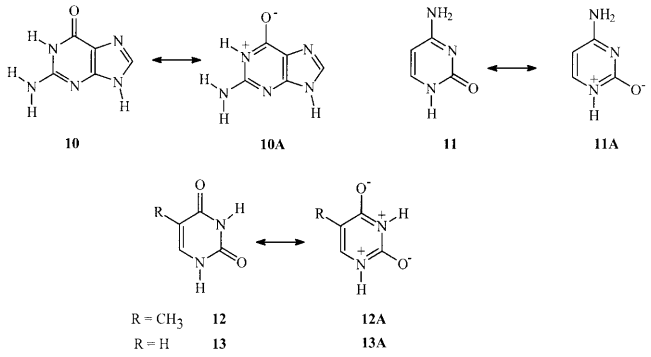


Fig. 3 Chart of delocalized structures of guanine **10**, cytosine **11**, thymine **12** and uracil **13**

1. Ketones and aldehydes are much more stable than their derived enols.
2. 9-Hydroxyanthracene does not exist, rather that molecule exists as anthrone.
3. In solution, 1,3,5-trihydroxybenzene is in equilibrium with cyclohexan-1,3,5-trione, and shows some typical "ketone" chemistry.
4. In solution, the 2- and 4-hydroxypyridines are in equilibrium with the corresponding pyridones and exist predominantly in the "carbonyl" form.

Thus, even when we consider their globally delocalized structures, guanine **10**, cytosine **11**, thymine **12**, uracil **13** (shown in the chart in Fig. 3) and their derivatives should tend to be less aromatic than adenine. Further, since thymine and uracil possess two amidic carbonyl groups, these molecules should be the least aromatic of the lot. The zwitterionic natures of these globally delocalized structures, below, also suggest that these highly energized resonance forms should contribute very little to the structures of their resonance hybrids, which will then resemble the non-aromatic "carbonyl" forms.

The synergistic role of Watson–Crick hydrogen bonding

Watson–Crick hydrogen bonding in these molecules, above, can change their electron density distribution quite dramatically. The resonance forms of guanine **10** and cytosine **11** are instructive. The amidic resonance form of guanine **10** clearly does not possess the features of an aromatic entity, while the non-amidic zwitterion **10A** would be expected to be only weakly aromatic in the pyrimidine ring because of the separated charges.

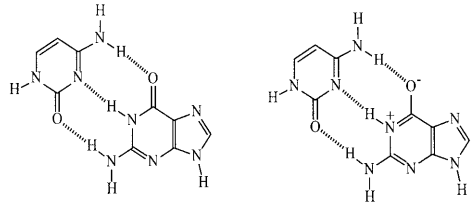


Fig. 4 Chart for guanine **10** in a Watson–Crick type of hydrogen bonding with cytosine **11**

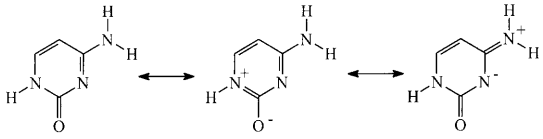


Fig. 5 Chart of resonance forms of cytosine **11**

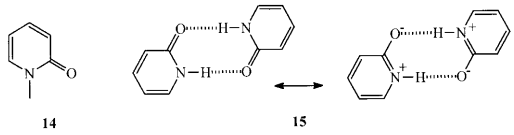


Fig. 6 Chart of *N*-methyl-2-pyridone **14** and 2-pyridone **15** as a hydrogen bonded dimer

However, if we were to engage guanine **10** in a Watson–Crick type of hydrogen bonding with cytosine **11** (see chart in Fig. 4), then the N–H bond of **10** should be weakened, so decreasing the size of the positive charge on the nitrogen, and the electron density on the oxygen should be decreased, so decreasing the size of the negative charge on that oxygen. Both effects would lead to an enhancement of the stability of the charged resonance form of guanine and favor the establishment of true delocalization, and hence aromaticity, in the guanine.

Simultaneously, the resonance forms of the cytosine molecule mirror the charge redistribution due to the hydrogen bonding and this enhances the stability of the zwitterionic resonance form of the cytosine, and hence the aromaticity of the cytosine (see chart in Fig. 5). The other normally encountered Watson–Crick base pair, adenine and thymine, also shows a similar, desirable structural and electronic complimentarity to the guanine–cytosine pair examined above.

This type of analysis is dramatically strengthened by the available X-ray crystallographic data for *N*-methyl-2-pyridone **14** and 2-pyridone **15** (chart in Fig. 6). The bond length data for the *N*-methyl compound **14** clearly shows that it is not aromatic in the solid phase of its ground state. The compound **15** crystallizes as a hydrogen bonded dimer, and the bond length data show that the units in the dimer are clearly aromatic. [1] Hydrogen bonding clearly stabilizes the charge separation that is required for the compound **15** to become aromatic, and one must therefore conclude that the aromaticity of compound **15**, in its solid phase dimer, is largely the result of this strategic hydrogen bonding, which the non-aromatic compound **14** cannot experience.

Discussion

The X-ray crystallographic data and its analysis

We selected only high resolution X-ray crystallographic atomic coordinate data (resolution≤2.5 Å) of 56 DNA oligomers that were present in the Brookhaven Protein Databank. This group of molecules was divided into two sets, one whose structures had been determined at resolution≤1.5 Å, and the rest. Every base in each X-ray study was individually assessed for its aromaticity (DI) using STR3DI32.EXE. The data were analyzed for all of these molecules, the higher resolution group, and the lower resolution group. This allowed us to assess the confidence with which we could use the data from the lower resolution group.

One of the problems that exists in many X-ray crystallographic studies is the presence of small random errors in the atomic coordinate data that will cause some atoms to be slightly mispositioned. These coordinate data errors seem to increase with the size of the molecule being studied. These errors usually show themselves when, for example, five of the six bonds of a benzenoid molecule have lengths close to ideal, but one bond has a length slightly outside of the range of acceptable lengths. This situation must obviously be due to one of the atoms in the benzenoid being slightly mispositioned. We have tried to assess the severity of these errors in the atomic coordinate data of each molecule we studied. The data for a potentially aromatic ring that obviously suffered from this error were ignored, since we did not wish to adjust any of the experimentally determined data. Further, any molecule's data that showed several of these errors were excluded from the study since a large number of these errors clearly point to poorly acquired, or processed, data.

We downloaded the atomic coordinate data of several molecules from the Brookhaven Protein Databank (Appendix A). [8] Of these, PDB1DAO, PDB1D11, PDB1IMS, PDB1D35, PDB224D, PDB245D, PDB427D, PDB482D and PDB1D35 were intercalated with anthraquinoid molecules.

After examination, we decided to omit PDB292D, PDB284D, PDB293D, PDB2DCG, PDB308D, PDB315D, PDB332D, PDB336D and PDB362D from the study because the quality of the data was poor (approximately 50% of the bases in each structure showed instances of the "mispositioned atom" problem). The data for PDB1D97 were omitted because we could not confidently assess the effects of the thiophosphate units on the structure.

We first measured and tabulated the DIs of each type of heterocyclic base in each molecule studied. Bases that were mismatched, in non-Watson–Crick pairs, were omitted, as too were bases that were not paired. All of

Table 1 Average DIs for all Watson–Crick molecules studied

Base	Pyrimidine ring	Imidazole ring	Purine average
Adenine	1.34	3.14	2.29
Guanine	2.45	3.44	2.95
Cytosine	3.27		
Thymine	4.15		
Uracil	3.57		

Table 2 Average DIs for Watson–Crick molecules studied (2.5 Å <res.> 1.5 Å)

Base	Pyrimidine ring	Imidazole ring	Purine average
Adenine	1.27	3.14	2.21
Guanine	2.37	3.66	3.01
Cytosine	3.26		
Thymine	4.20		
Uracil	none		

the data for a given base were then compiled and the average DI for that kind of base, over all of the molecules studied, was calculated, Table 1. A separate analysis of the data obtained at high resolution (≤1.5 Å) and at lower resolution (>1.5 Å and <2.5 Å) was performed to ensure that the lower resolution data did not distort the entire analysis. The lower resolution data are shown in Table 2. Tables 1 and 2 show that the quality of the data obtained from the lower resolution studies was as acceptable as that from the higher resolution studies, after STR3DI32.EXE had compensated for the experimental bond length error.

The data in Tables 1 and 2 supported our earlier conclusion that adenine was the most aromatic of the heterocyclic bases, and that uracil and thymine were the least aromatic. We must remember that the bases guanine, cytosine, thymine and uracil were shown to be non-aromatic from the X-ray crystallographic coordinate data for their simple derivatives (like their *N*-methyl derivatives), [1] but now Table 1 clearly shows that these heterocyclic bases were aromatic when in their Watson–Crick pairs.

Table 1 also showed that the "pyrimidine" rings of adenine and guanine were more aromatic than their "imidazole" rings. However, these "imidazole" rings were comparably aromatic to the single "pyrimidine" rings of cytosine and thymine/uracil.

The aromaticity of unpaired bases and bases in mismatched pairs

We had intentionally included in the group of molecules studied some with mismatched base pairs (PDB3DNB, PDB1D91, PDB113D, PDB112D, PDB111D, PDB1D99, PDB1D92, PDB1D8X and PDB1D9R) and three others (PDB463D, PDB476D and PDB475D) that had nucleotide sequences that forced a base at the end of each helix to remain unpaired. Our intention was to ascertain whether, or not, the unpaired bases and the mismatched (non-Watson–Crick paired) bases would be aromatic. After all, we had suggested that the aromaticities of guanine, cytosine, thymine and uracil should depend on their hydrogen bonding milieu.

All of the structures that had unpaired bases (all guanine) showed these bases to be aromatic. We were intrigued to see that each unpaired base had set up a series of hydrogen bonds with water molecules that were held between these bases and the phosphate unit of the adjacent nucleoside. Thus, these unpaired guanine units certainly had sufficient hydrogen bonding at their important sites to trigger them into aromaticity. Another example of guanine becoming aromatic in a non-Watson–Crick situation is in the structure GMPTRGL (7-methylguanosine-5'-phosphate tryptophanylglutamic acid complex). [9] Here the dipeptide unit supplies suitable hydrogen bonding.

The non-Watson–Crick paired bases encountered in this study (A:G, A:C, T:G pairs) all had two pairs of mutual hydrogen bonds linking the bases. Since adenine is aromatic in any environment (even without hydrogen bonding) then we did expect to see these mismatched adenines retain their aromaticities, and they all did. All of the mismatched guanine bases were also aromatic, and this suggests that guanine only needs two hydrogen bonds to initiate its aromaticity.

The cytidines in the A:C pairs found in PDB1D99 seemed to be non-aromatic. However, the data for all of the cytidines in this structure, including those in G:C pairs, seemed to be flawed by the above-mentioned "mispositioned atom" problem, since all of their ring bonds, except the N-1–C-2 bonds were well within the ranges of lengths suitable for aromatic/delocalized bonds. The cytidines in G:C pairs in most of the other structures were indeed aromatic and so we are inclined to suggest that these mismatched (A:C) cytidines should also be regarded as aromatic.

The thymines in non-Watson–Crick T:G pairs also seemed to be aromatic, but with significantly attenuated aromaticities. The average DI for these thymines was 4.833, while that for the thymines in the study as a whole was 4.149. Thus, mismatching thymines seemed to have significant effect on their aromaticities.

The aromaticity of synthetic bases in Watson–Crick pairs

The study embraced molecules with non-traditional bases. PDB114D had deaminoguanine; PDB1D9R had 5-bromouracil; PDB1D40 and PDB1D41 had 5-methylcytidine; PDB1D76 had 2-aminoadenine; and PDB1D61 had inosine. The Watson–Crick paired 2-aminoadenines (A*:T) were aromatic. The non-Watson–Crick paired deaminoguanines (A:G*) in PDB114D were aromatic since they all had two hydrogen bonds to their adenine partners. PDB1D61 showed a single strand and so we could not assess the hydrogen bonding milieu of the inosine unit. However, that inosine unit was aromatic.

One of the two 5-bromouracils in A:U* pairs in PDB1D9R was aromatic (DI=4.6) and the other seemed

to be non-aromatic. However a close examination of the non-aromatic 5-bromouracil showed that it too suffered from the above-mentioned "mispositioned atom" problem, since all of their ring bonds, except the C-4–C-5 bonds were well within the ranges of lengths suitable for aromatic/delocalized bonds. We suggest that this base should also be regarded as being aromatic.

The 5-methylcytidines in G:C* pairs in PDB1D40 seemed to be non-aromatic, while two of the four of those in PDB1D41 (same structure) were aromatic. It might well be that we should not have include PDB1D41 and PDB1D41 in this study because the data for these molecules were low in quality.

The role of aromaticity in the structure and properties of DNA

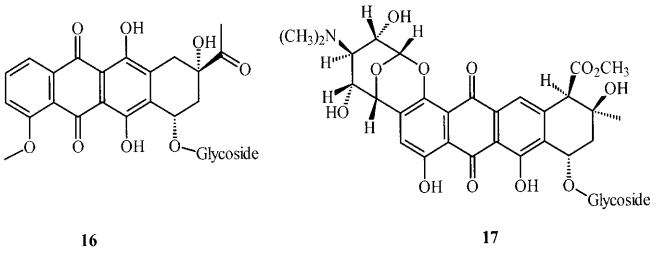
The synergistic hydrogen bonding effects described above had suggested that the DNA bases would all be aromatic in the very favorable environment of Watson–Crick hydrogen bonded pairs. The X-ray crystallographic data presented have substantiated this notion. The data also confirmed that the bases in mismatched pairs were aromatic when these pairs shared at least two hydrogen bonds.

Since the DNA bases do indeed become aromatic in the Watson–Crick milieu, then the stability of the double helix must be due, at least in part, to the resonance energies of the "newly" aromatized entities, guanine, cytosine, thymine and uracil. The "unwinding" of the double helix must also be accompanied by a substantial energy increase, unless the aromaticities of the bases are preserved by engaging them in Watson–Crick type hydrogen bonds.

The intercalation of anthraquinoid entities

The intercalation of anthraquinoid molecules, like daunomycin and nogalamycin, into DNA double helices has intrigued us. Hydrogen bonding between the glycosidic moieties of these molecules and the DNA backbone obviously plays a role in stabilizing the complex, but we have wondered if stereo-electronic factors that had not been considered also contributed. We therefore decided to examine the X-ray crystallographic data for some of these intercalated systems – PDB1D11, PDB2DES, PDB1DA0, PDB1IMS, PDB482D, PDB1D35, PDB427D, PDB224D and PDB245D – in the hope of uncovering any new information.

PDB1D35 and PDB427D were unique and intriguing structures that had their glycosidic entity linked to a neighboring non-adjacent adenine by a methylene bridge (from formaldehyde). This linkage must restrict the positioning of the anthraquinoid molecule in the complex, and so must perturb the entire structures of these materials to some extent. We therefore decided not to include these data in this study, even though we shall discuss these molecules in the near future.



16 17

Fig. 7 Chart of the two anthraquinoid aglycone groups found in the daunomycin group **16** or in the nogalamycin group **17**

Table 3 Average DIs for intercalated molecules studied

Base	Pyrimidine ring	Imidazole ring	Purine average
Adenine	1.62	3.25	2.44
Guanine	2.45	3.98	3.22
Cytosine	3.33		
Thymine	4.18		
Uracil	None		
Intercalate	Ring A	Ring B	Ring C
Anthraquinone **16**	1.35	1.66	1.98

The other molecules fell into two categories – those that had the anthraquinoid aglycone **16** found in the daunomycin group of molecules, and those (PDB224D and PDB245D) that had the anthraquinoid aglycone **17** (see chart in Fig. 7) found in the nogalamycin group. The nogalamycin aglycone **17** is not only highly substituted on each benzenoid ring, but also the dioxabicyclo [3,3,1] nonene unit should restrict access to one face of the molecule while the other face is completely exposed and the other quite unhindered. On the other hand, the daunomycin aglycone has both faces of the anthraquinoid entity available for interactions.

The data for molecules PDB224D and PDB245D were extensively flawed by the "mispositioned atom" problem. Indeed, we have shown that adenine is aromatic in each ring, regardless of the molecule's environment. However the adenines in these two studies were only aromatic in their six-membered rings and the 5-membered rings showed distinct evidence of the "mispositioned atom" problem. Indeed, this "mispositioned atom" problem was seen in the data for almost every base in these studies, and we could not be confident that the data for the nogalamycin entities did not suffer from this problem.

We therefore, reluctantly, decided to omit PDB224D and PDB245D from the study. The data for the daunomycin group of molecules PDB1D11, PDB2DES, PDB1DA0, PDB1IMS and PDB482D are shown in Table 3.

The most dramatic discovery made in this study is the realization that the anthraquinoid aglycone **16** had been transformed into a truly aromatic anthracene. This could only occur if the quinone had been completely reduced to the hydroquinone, or to a radical anion in a charge-transfer process. Indeed, the average DIs for the agly-

cone **16** clearly show that the central ring B is just as aromatic as either the ring A or ring C, even when we assume a minimal error range of ±0.2 units in these DI measurements. Remember that the DI assessments are based on the experimentally measured bond lengths and that the errors in the measurements of these bond lengths are reflected in the reported, experimentally measured, RMSD values.

The bond length data clearly suggest that reduction to the radical anion is much more likely to be correct since the C–O bonds of the previously "quinoid" (the central) ring of the anthraquinone still show significantly more double bond character than the C–O bonds of any phenol or hydroquinone we have ever examined. Thus, we must conclude that the anthraquinoid aglycone has indeed participated in a charge-transfer process with one of the bases, and the intercalate is stabilized both by hydrogen bonding between the glycosidic moiety and the oligomer backbone and by an ionic charge-transfer interaction.

It is very significant that in every instance the anthraquinoid entity was in a "pocket" made up of two G:C pairs, and so it seemed very likely that the guanine should be the source of the single electron. We have tried to locate the base most likely to be the radical cation partner, but the data do not allow us to do so. However, while it could logically be argued that the radical cation partner should be the guanine, of the four bases that form the "pocket" for the intercalation process, the "hole" could conceivably migrate to a neighboring electron rich base like an adenine. Indeed, the only base whose DI seemed to be significantly affected by this intercalation process is adenine, as is shown in Table 3.

We have reported a study that confirmed the extensive delocalization of the lone pairs of anilinic amine over their aromatic rings, [7] and the present study also showed that the lone pairs of the amino-groups in adenine and guanine were extensively delocalized. Hence one would expect that electron transfer from an anilinic entity would result in few geometric changes to that system, and certainly less than those shown in the transformation of a quinone to its radical anion. Thus, we were not surprised that the DIs for the guanines that encapsulated the intercalated entity were still quite similar to those of the guanines in non-intercalated molecules, nor was it surprising that the DIs for the adenines in the intercalated molecules show only a small increase (a small decrease in their aromaticities) in comparison to those in non-intercalated molecules.

Thus, the process of intercalating entities into DNA double helices can be facilitated either by hydrogen bonding alone, or with some electron deficient molecules, by charge-transfer formation. Indeed, after we started this work (F.J.-M.'s undergraduate research project 1994–1996) single electron (charge) transfer from DNA to anthraquinones was observed and characterized spectroscopically. [10] Those charge-transfer stabilized intercalated systems will obviously have very different physical properties to the others, and to the simple DNA double helices, especially in areas like their electrical conductivities. It is now easy to understand why some researchers have not found evidence of electrical conductivity in non-intercalated DNA sequences, [11] while others have found some evidence of electrical conductivity in intercalated DNA sequences. [12]

Conclusion

In a previous study we have shown, by careful analyses of the X-ray crystallographic coordinate data of DNA bases, [1] that only adenine is aromatic in its native state and its simple derivatives, while the other DNA bases were non-aromatic. In this study we have confirmed, by careful analyses of the X-ray crystallographic coordinate data of DNA double helices, that the DNA bases are indeed aromatic in the milieu of Watson–Crick hydrogen bonded pairs. We also confirmed that mismatched (non-Watson–Crick) base pairs that had two hydrogen bonds between them also provided enough stabilization for these bases to allow them to become aromatic. Indeed, the aromatization of 2-pyridone in its dimer should have foretold this. We also show that guanine can respond to hydrogen bonding from a variety of sources and be induced into aromaticity.

We do not have any definitive structural data regarding the possible aromatization of the DNA bases in aqueous solution, or when these bases are solvated by (hydrogen bonded to) water alone. Thus, we cannot rule out this possibility. It is striking that the X-ray crystallographic data do not show water molecules within the double helices, rather they show extensive solvation of the external surfaces of these structures. We also do not have structural data on water solvated denatured DNA and so cannot comment on the possible aromatization of the bases in these entities.

However, we have shown that the X-ray crystallographic data clearly suggest that the process of forming the double helix also induces guanine, cytosine, thymine and uracil to become aromatic. We therefore wish to suggest that this additional, newly developed, aromatic stabilization must now be listed among the other factors that stabilize the DNA double helix.

We have also shown that the process of intercalating the daunomycin anthraquinoid glycosides into the DNA double helix is assisted and stabilized by hydrogen bonding and charge-transfer interactions. These surprising discoveries should allow us to design more efficient intercalators, especially some that could dramatically affect the electrical conductivity of the DNA double helix.

We had previously suggested that nature might manipulate the aromaticities of molecules for the purposes of adjusting the intensities of intermolecular interactions, in much the same way that it uses hydrogen bonding, dipole–dipole, ion–dipole and ion–ion interactions. [1] This would require that there exist molecules that are either extremely weakly stabilized by aromaticity, or be borderline on being non-aromatic, which

could be induced into full aromaticity by some simple process. The induction of aromaticity in these molecules by protonations or hydrogen bonding is obviously a very suitable "trigger" in a biochemistry that is acid–base controlled.

Appendix A

Resolution≤1.5 Å – PDB1D61, PDB1D76, PDB3DNB, PDB1D11, PDB1D35, PDB1D48, PDB2DCG, PDB1D39, PDB1D41, PDB1D40, PDB1D8G, PDB1D8X, PDB1D9R, PDB1DA0, PDB1DC0, PDB1DJ6, PDB1DN0, PDB1DPL, PDB1IMS, PDB2DES, PDB224D, PDB245D, PDB284D, PDB292D, PDB293D, PDB295D, PDB308D, PDB315D, PDB475D, PDB336D, PDB355D, PDB362D, PDB427D, PDB428D, PDB431D, PDB440D, PDB441D, PDB455D, PDB463D, PDB476D, PDB482D and PDB485D.

Resolution >1.5 and <2.5 Å – PDB118D, PDB2D95, PDB1D96, PDB1D93, PDB1D91, PDB1D98, PDB113D, PDB114D, PDB1D99, PDB111D, PDB112D, PDB332D, PDB1D97 and PDB1D92.

References

1. Box VGS (1992) Heterocycles 34:1631–1659
2. a) Vedaldi D, Dall'acqua F, Caffieri S, Rodighiero G (1982) Farmaco Educ Sci 37:571–581; b) Veselkov AN, Eaton RJ, Baranovskii SF, Osetrov SG, Bolotin PA, Djimant LN, Pahomov VI, Davies DB (1999) J Struct Chem (Engl Transl) 40:230–238
3. a) Arora SK (1983) J Am Chem Soc 105:1328–1332; b) Williams HE, Searle MS (1998) J Chem Soc, Perkin Trans 1 1:3–6
4. Box VGS (1991) Heterocycles 32:2023–2041
5. Box VGS (2001) J Mol Model 7:193–200. DOI 10.1007/s008940100030
6. Box VGS (1997) J Mol Model 3:124–141
7. Box VGS, Yu HW (1997) J Chem Educ 74:1293–1296
8. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank, Nucleic Acids Res 28:235–242 http://www.rcsb.org/pdb/
9. Ishida T, Iyo H, Ueda H, Doi M, Inoue M, Nishimura S, Kitamura K (1991) J Chem Soc, Perkin Trans 1 1847–1853
10. Breslin DT, Schuster GB (1996) J Am Chem Soc 118:2311–2319
11. a) Lewis FD, Liu X, Liu J, Hayes RT, Wasielewski MR (2000) J Am Chem Soc 122:12037-12038; b) Zewail AH, Wan C, Fiebig T, Schiemann O, Barton JK (2000) Proc Natl Acad Sci USA 97:14052–14055
12. a) Barton JK, Holmlin RE, Dandliker PJ (1997) Angew Chem, Int Ed Engl 36:2715–2730; b) Giese B, Meggers E, Michel-Beyerle ME (1998) J Am Chem Soc 120:12950–12955