Order Number 9304680

Theoretical studies of anticancer drugs, lexitropsins

Kabir, Shahrzad, Ph.D.

City University of New York, 1992

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

THEORETICAL STUDIES OF ANTICANCER DRUGS, LEXITROPSINS

by

SHAHRZAD KABIR

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1992

This manuscript has been read and accepted for the Graduate Faculty in Chemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Sept 11, 1992
date

Anne-Marie Sapse
Chairman of Examining Committee

Sept. 14, 1992
date

Executive Officer

Supervisory Committee

The City University of New York

Abstract

THEORETICAL STUDIES OF ANTICANCER DRUGS, LEXITROPSINS

by

SHAHRZAD KABIR

Adviser: Dr. A. M. Sapse

The purpose of this research study was to gather information about the structure and activity of some anticancer drugs, leading eventually to better drug designs. The following studies were undertaken:

1) The investigation via geometry optimization of the structure of one small lexitropsin, amidinomycin, which is an oligopeptide that binds to the minor groove of B-DNA.

2) Proton affinities of some hydrogen acceptor rings that are present in some lexitropsin were studied in order to estimate their capacities to bind to GC sequences of DNA.

3) Binding power of one of the DNA bases , thymine, to either guanidinium ion as present in netropsin or aminopyrrolidinium ion moiety as is present in Anthelvencin was compared in order to determine how much these two groups contributed to the overall binding of netropsin and Anthelvencin to the base sequences of DNA. It was found that ab initio calculations on Amidinomycin agree well with the experimental results and the proton affinities of imidazole is much higher than the one of oxazole which in turn is much higher than the one of thiazole and a methyl group substituent increases the proton of imidazole, while a peptidic group decreases it. Also, it was found that the binding of guanidinium and aminopyrrolidinium ions to uracil as a model for thymine is very similar.

Acknowledgments

I would like to thank my mentor, Dr. A. M. Sapse for her years of encouragement and advice. I would also like to express my gratitude to my thesis committee, Dr. Charlotte Russell and Dr. Joel Gerstein for taking time to read this manuscript and Professor Richard Pizer for helping me to reach this goal. Also, I would like to thank my dear friend Kathy Roberts for typing this thesis.

Last, I am very grateful to my family for their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (continued)

LIST OF ILLUSTRATIONS (continued)

I) INTRODUCTION:

Cancer is one of the most serious illnesses that kills millions of people around the world each year. The major approaches to cancer treatment include surgery, radiation, chemotherapy, immunotherapy, or a combination of the four. Chemotherapy is one of the methods that is used often to treat cancer. However, the development of a logical plan for cancer chemotherapy requires that several assumptions be made.[1]

"1) New drugs with possible antitumor effectiveness should be selected rationally from the widest possible base among synthetic chemicals and natural products.

2) Drugs should be found with sufficient capacity to kill tumor cells and produce permanent, complete remission within the limits of acceptable morbidity to the patient.

3) Both tumor-cell kill capacity and toxic effects of drugs in patient can be predicted from the drugs effect in animal models.

4) Animal models based on tumor-cell kinetics can predict drug effects in patients with a higher probability for success than is possible with nonkinetic models.

5) Drug effects in vivo can be assessed adequately only when the exposure time of cells to specific amounts of drugs can be established.

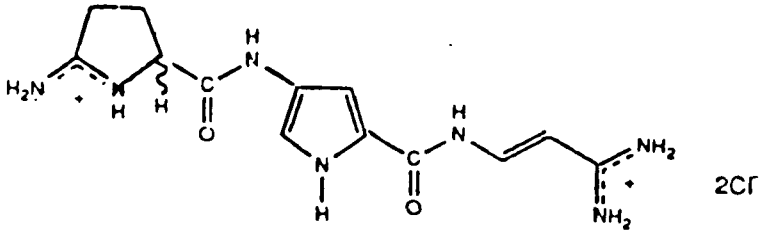
6) Animal models, having helped study known effective drugs, will help predict new effective drugs."

All highly effective and curative treatments with chemotherapy today are in the form of combination chemotherapy.[1-2] All effective combination chemotherapies are derived from biochemical and drug-resistance rational. These include primarily evidence for activity of the individual agent against the disease and evidence that toxicity is non-additive, so that the dosage of the individual agents are not seriously damaging when employed in combination.

One of the biggest problems with the use of chemotherapeutic agents has always been the side effects that these drugs may have. Many experiments have shown that the principle target of many anticancer drugs is DNA. Therefore, in order to reduce the side effects of these drugs which are due mostly to their lack of selectivity in killing cells, one must get

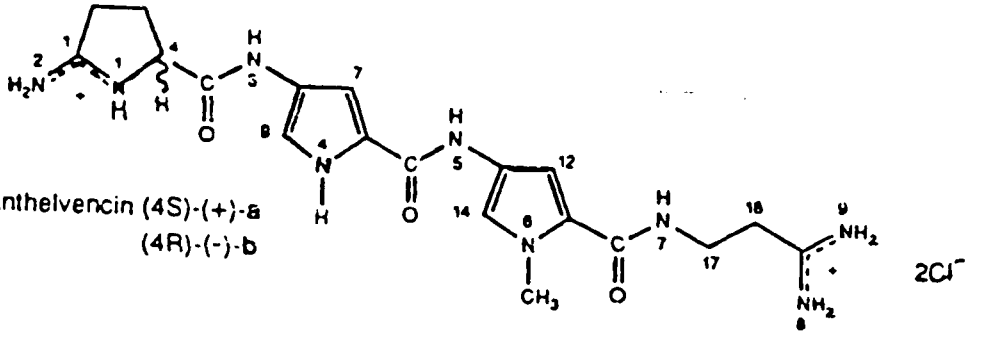
information about ways to increase the ability to distinguish between normal cells and cancer cells. A category of drugs which form the object of this study are oligopeptides termed "Lexitropsins". The lexitropsins' anticancer activity is based on their ability to bind to the DNA minor groove and this way prevent the DNA replication. The sequence selectivity of these drugs should play an important role in their ability to bind preferentially to cancer cells' DNA. It is not known as yet in what way the cancer cells' DNA differs as far as base sequences are concerned from the normal cells' DNA but it will probably be known in the future. Among the most important compounds of this type are: 1) kikumycin, 2) anthelvencin, 3) distamycin, 4) netropsin, 5) amidinomycin and 6) noformycin, fig(1-6).

Fig. (1)



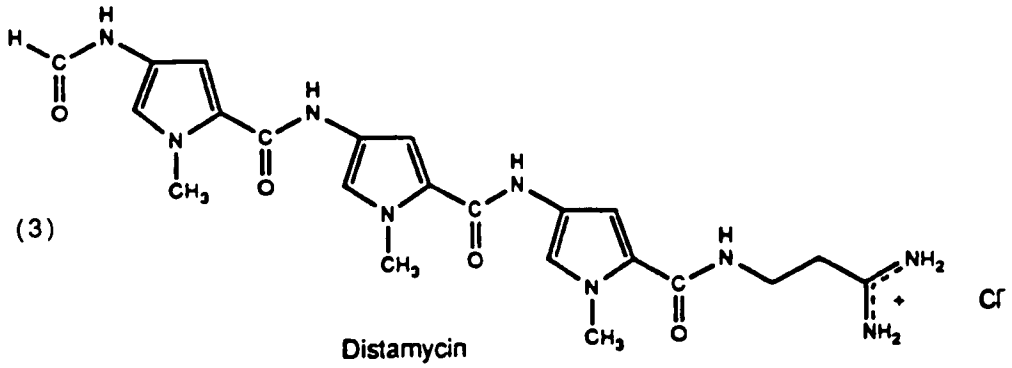
Kikumycin (4S)-(+)-a; (4R)-(-)-b

Fig. (2)



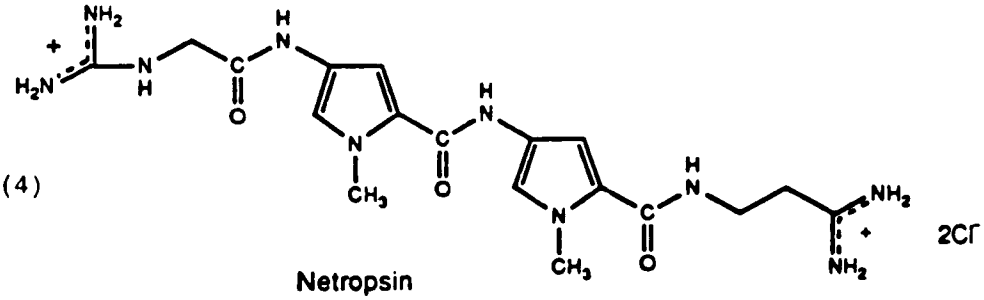
Anthelvencin (4S)-(+)-a
(4R)-(-)-b

Fig. (3)



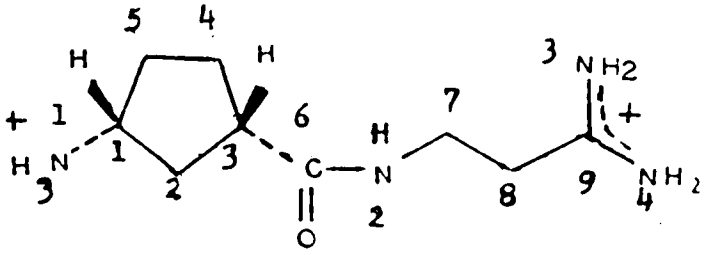
Distamycin

Fig. (4)



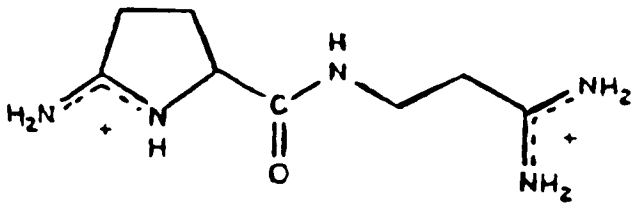
Netropsin

Fig (5)



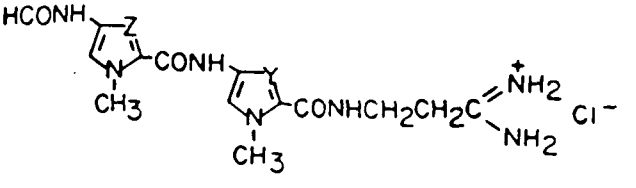
Amidinomycin

Fig. (6)



Noformycin

Evidence from biochemical pharmacology indicates that these oligopeptide antibiotics [3-7] act to block the template function of DNA by binding primarily to (AT)n sequences in the minor groove. While these natural products (1-6) show only moderate antiviral and anticancer activity, synthetic "lexitropsins" to which they gave rise are capable of recognizing alternative sequences and show higher antiviral and anticancer activity.[8] The chemical activity of these compounds has not yet been established. The research is concentrating on the improvement of the selectivity for DNA binding. For example, Lown and coworkers synthesized a series of monocationic lexitropsins that displayed high specificity for GC sequences. These compounds have an N-formyl group in place of guanidinium moeity normally present in netropsin. By replacing the N-methyl pyrrole groups of the dipeptide with N-methyl imidazole, a high degree of sequence specificity was seen in the sense that the preference for GC versus AT binding increases. The structure of these drugs are shown in fig.(7).



Z=CH, Y=CH
Z=N, Y=CH
Z=CH, Y=N
Z=N, Y=N

fig (7)

Heteroatomic Lexitropsins

Several factors effect molecular recognition and the binding of these drugs to DNA. 1) Size of the lexitropsins effect the binding of the drugs to B-DNA. Lexitropsins that have greater length show greater biological potency.[9] On the other hand , shorter oligopeptides are more sensitive in terms of sequence preference than longer ligands. For example, this effect can be seen in case of Dansyl-Distamycin-2 and Dansyl-Distamycin-3. Dansyl-Distamycin-2 is smaller with respect to size than Dansyl-Distamycin-3 in the sense that it has two N-methylpyrrole rings while Dansyl-Distamycin-3 has three N-methylpyrrole rings. It was shown that Dansyl-Distamycin-2 would bind to poly(dA-dT). Poly(dA-dT) of DNA is about ten times weaker than Distamycin-3.[10] The binding for Dansyl-Distamycin-3 and Dansyl Distamycin-2 to poly(dA-dT). Poly(dA-dT) are 1.7×10^7 M^{-1} and 2.0×10^6 M^{-1} respectively.

2) Chirality also plays an important role. Chiral ligands should be isohelical with the minor groove of DNA. For example, a great difference was seen in the binding capacities of 4S (+)-anthelvencin and 4R-(-)-enantiomers[11] as shown in fig(8a) and (8b). NMR studies show that both enantiomeric forms of anthelvencin bind to the sequence 5'-AATT-3' of the DNA. NOE studies show that the (4S) enantiomer is twisted between the two pyrrole moeities which provides the structural flexibility for the drug to bind along the 5'-AATT sequence in the minor groove of DNA. The conformation of the (4R) enantiomer is closer to planar when it is bound to DNA. The chirality of the (4S) enantiomer with its (4S) proton pointing out of the minor groove provides favorable electrostatic interactions between the positively charged 2-amino-1-pyrrolinium group and the DNA. On the other hand, the (4R) enantiomer behaves like an achiral lexitropsin and binds centrally between the 5'-AATT sequence. The steric interaction between the (4R) portion and the DNA interfaces with the electrostatic attraction between the positively charged 2-amino-1-pyrrolinium group and the DNA.

3) The relative position and distances between hydogen bond-accepting and donating sites is very important for effective recognition and binding of ligands. This effect can be observed in a series of monocationic lexitropsins which contain an N-formyl group in place of

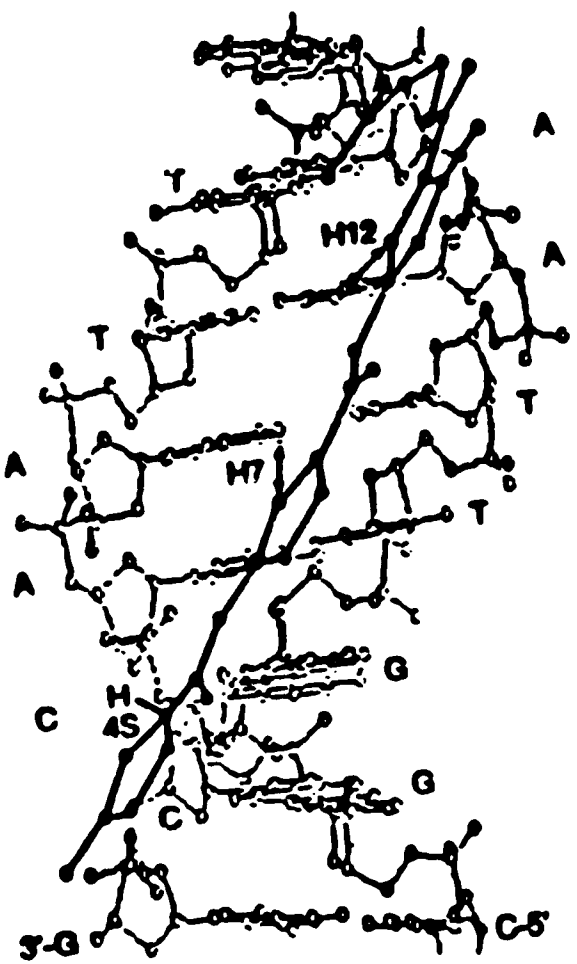


fig (8a)

Anthelvencin (4S)-(+)-a

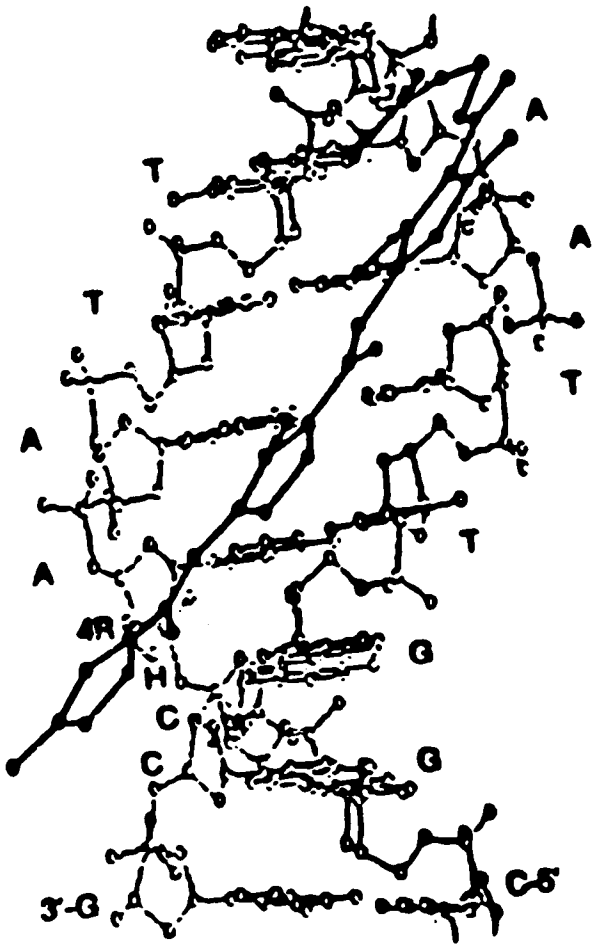


fig (8b)

Anthelvencin (4R)-(-)-b

guanidinium moeity normally present in netropsin, fig(7)[12]. By replacing one or two N-methylpyrrole rings by N-methylimidazole moeities, the binding of these dipeptides are altered significantly in the sense that these compounds are able to recognize AT as well as GC sequences of DNA. The reason for this altered specificity is due to hydrogen bonding between the nitrogen of imidazole and the hydrogen of the 2-amino group guanine located in the minor groove of DNA. Also, the relative position of the imidazole and pyrrole moieties are found to be important. When the pyrrole moeity is located between the charged propylamidinium moiety abd the N-methylimdiazole, electostatic effects associated with the amidinium and the hydrogen bonding between the imidazole and guanine help the contact between the pyrrole and DNA. However, when the pyrrole moiety is placed on the amino terminus of the ligand, the pyrrole-DNA contact is insufficient and the binding is reduced.
4) Charge influences the selectivity of the binding ligand. Pullman and coworkers have shown that the binding of these drugs is not only due to hydrogen bonding but also to electrostatic effects such as the interaction of a positively charged ligand with the great negative potential present in the minor groove of B-DNA[13]. This can be seen from the fact that B-DNA has an overall negative charge due to its phosphate groups. The negative charge distribution of double stranded DNA is uneven because major and minor groups places the phosphate at uneven distances, with the negative charge in minor groove more concentrated due to narrowness of minor grooves.

Many theoretical studies have been performed in order to investigate the structure and binding of lexitropsins. Some of these studies are reported by Pullman.

One of his research studies is centered on the binding to B-DNA of a series of lexitropsins. These are netropsin derivitive in which one or both of the pyrrole rings have been replaced by imidazole[13] . These are termed Lex "A", "B" and "AB", fig (9). In lex "A", N-methylpyrrole residue is located between the charged propylamidinium moiety and the N-methylimidazole while in lex "B", the relative position of the imidazole and pyrrole are reversed . Lex "AB" contains two N-methylimidazole residues. The details of the hydrogen

bonding for netropsin with poly(dG).poly(dC) show that the peptide groups form hydrogen bonds with Cytosine O2 or Guanine N3 atoms as shown in fig 10(a). Pullman found that when a pyrrole moiety was replaced by imidazole, the formation of a new hydrogen bond between the nitrogen atom of imidazole and hydrogen of amino group of guanine is observed as shown in fig (10b)-(10d). He determined that this new hydrogen bond stabilizes the lexitropsins complex by an increase in the oligonucleotide-ligand interaction energy term and by the reduction of the ligand distortion energy. However, he concluded that these charges were not sufficient to produce a preference of lexitropsins for GC sequence. From his calculation on the difference in complexation energy of these ligands for AT and GC sequences, he drew two conclusions. First, the preference for AT sequences was conserved for all these ligands, in agreement with the experimental results. This preference originated from the difference in the DNA-ligand interaction energy and especially from its electostatic component. These results showed that the origin of the AT sequence preference for these cationic ligands was related to the distribution of the electrostatic potential which was most negative in the minor groove of AT sequences[14-15]. Second, the AT preference decreased when pyrrole rings were substituted by imidazole which was due to the stabilization of GC complexes through the formation of the additional hydrogen bonds.

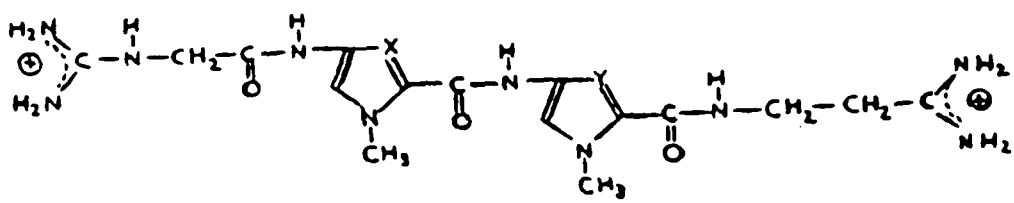
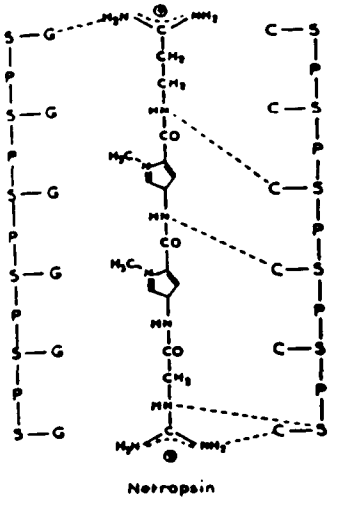
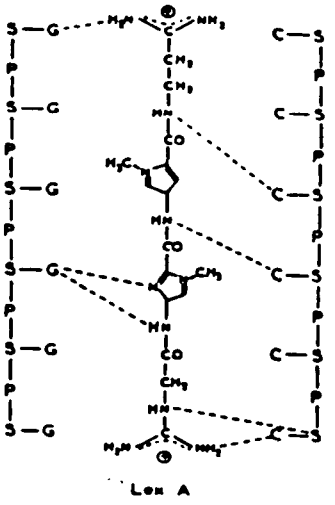


Fig. (9)

Netropsin	X = CH , Y = CH
Lex A	X = N , Y = CH
Lex B	X = CH , Y = N
Lex AB	X = N , Y = N



fig(10a)



fig(10b)

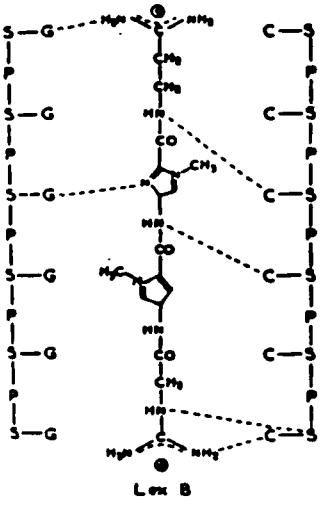


fig (10c)

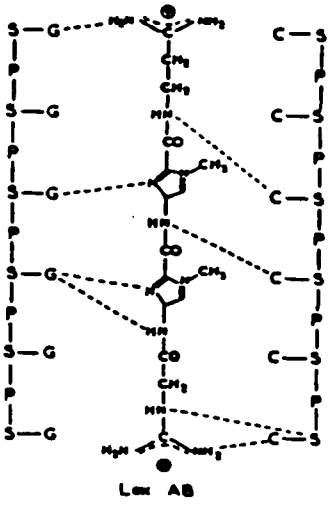
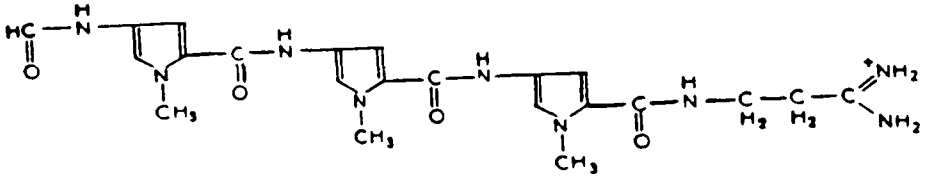


fig (10d)

Pullman and coworkers also showed that there were structural factors that affected the binding of netropsin and distamycin A[16] fig (4) and fig (11), to different types of double stranded nucleic acid. The results of their investigation showed that the affinity of B-DNA and A-DNA towards netropsin and distamycin A were correlated with the location of the negative potential in their minor groove. Also, as a complementary condition, hydrogen bonding sites N3 of the purine and O2 of the pyrimidines should be available to these drugs. The importance of these factors can be seen from the molecular electrostatic potential calculation that were performed by Pullman on different types of DNA, related to their interaction with netropsin and distamycin A. For example, in poly(dG).poly(dC) of DNA, the most negative potential was -632kcal/mole which was concentrated in the major groove while in the minor groove, the potential was only -603 kcal/mole . In poly(dA).poly(dT) of B-DNA, the most negative potential was -625kcal/mole in the minor groove while the potential in the major groove is -598 kcal/mole. From his results, Pullamn concluded that the specificity of netropsin and distamycin A for poly(dA).poly(dT) or for AT sequences of B-DNA was related to the location of the potential which was strongest in the minor grooves of B-DNA. The potential as well as the steric factors do not allow the lexitropsin to bind tightly to the major groove. Also, the reduced affinity of these antibiotics for the alternating copolmer poly(dA-dT).poly(dA-dT) with respect to the homopolymer poly(dA).poly(dT) was explained as before. In the alternating B-form, the potential minimum in the minor groove was only -605 kcal/mole while in the case of complementary homopolymer was -625 kcal/mole. The potential at the "anchoring" sites, N3 (A) and O2 (T) in different types of DNA showed that their values in the alternating -B form of poly(dA-dT).poly(dA-dT) were smaller than in poly(dA).poly(dT) while they were greater than the latter in the B-form of poly(dA-dT).poly(dA-dT). The accessibilities to N3 (A) and O2 (T) were satisfactory in all these helices and slightly greater for the alternating -B form of poly(dA-dT).poly(dA-dT) than poly(dA).poly(dT), however tha accessibility factor played a secondary factor in determining the order of affinities. For poly(dG-dC).poly(dG-dC), only the B-DNA

conformation was reported. The results showed that the location of minimum potential was in the major groove of this alternating copolymer and it also had a very low accessibility to the "anchoring" sites, N3 (G) and O2 (C), therefore, this copolymer would not interact with netropsin and distamycin A.



Distamycin A

fig (11)

In addition, these antibiotics did not react with single stranded DNA (no grooves) or with A-DNA which had a high accessibility to the mentioned hydrogen bonding sites but the negative potential was in the major grooves. This was due to the fact that the base pairs in A-DNA were away from the helical axis, toward the minor groove, so this groove became extremely shallow while the major groove was narrow and deep.

Additional studies were done by Pullman and his team on the binding of minor groove base reading polymeric ligands, called "Isolexins"[17] as shown in fig (12a-12f). They were composed of the furan-pyrrole-furan sequence and were joined by appropriate linkers such as C=O and N-H. Some isolexins had positively charged ligands such as propioamidinium groups at both ends while the neutral ones had either methyl, amino or carbonyl group at both ends.

Pullman discovered that the formation of hydrogen bonds between a ligand and the nucleic acid bases was not sufficient to ensure its binding specificity which were determined largely by the electrostatic factors. The importance of this finding was shown for two

fig (12a)

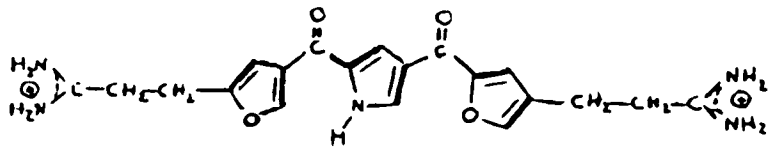


fig (12b)

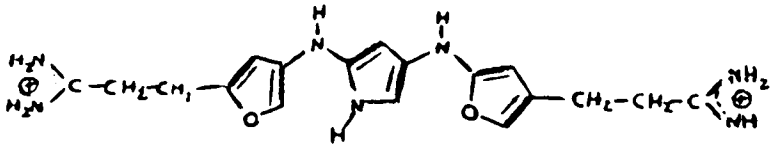


fig (12c)

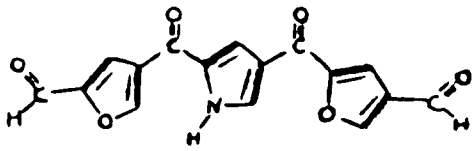


fig (12d)

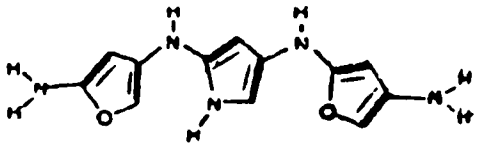


fig (12e)

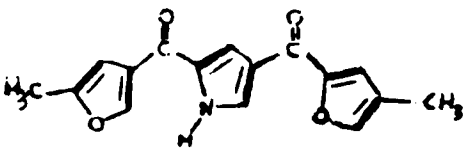
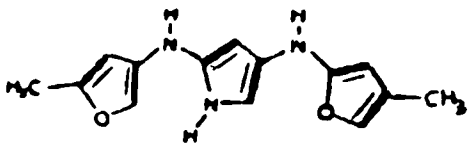


fig (12f)



dicationic isolexins. These were composed of the furan-pyrrole-furan sequence that were joined by either C=O or two N-H linkers. Their terminals consisted of two positively charged groups ($R=CH_2-C(=NH_2)-NH_2)^+$ which were similar to the charged ends of netropsin, fig 12a-12b. From his computational results on the binding of these two dicationic isolexins with various sequences of DNA, he concluded that these isolexins prefer to bind mainly to AT sequences despite their structural design to bind to GAG sequences via hydrogen bonding. In the case of neutral isolexins, the nature of the groups forming the linkers was a major factor in defining the specificity, although these groups did not participate directly in the interaction with DNA. These facts could be observed for two neutral isolexins, one with two C=O linkers and two HC=O terminals, fig (12c), which prefered to bind to AT sequences while the other neutral isolexin with two N-H linkers and two NH_2 terminals only showed binding preference for GAG sequences of DNA, fig (12d).

Further studies were done on neutral methylated isolexins in order to distinguish between the effects of the ligand terminal groups and that of the linkers. He considered two neutral isolexins with methyl groups at both ends but the first one had two C=O linkers while the second one had two N-H linkers, fig (12e-12f). His results showed that isolexin with C=O linkers preferred AT sequences although the difference in complexation energy was smaller for this isolexin than its derivative, fig (12c), while the other one with two N-H linkers preferred GC sequences. He concluded that the binding specificity of these ligands to DNA depends on the electrostatic component of the interaction energy as well as groups or parts of the ligands (the linkers).

Pullman also studied isolexin-like prototypes which were designed for specific binding to the minor grooves of GC sequences of B-DNA [18]. The structure of this compound composed of three furan ring, connected by two NH linkers with two neutral methyl groups at both ends, fig (13a). He showed that only a very limited improvement could be obtained by increasing the proton accepting capabilities of the heteroaromatic ring systems of the prototype even though these rings would interact directly with the proton donating NH_2

fig (13a)

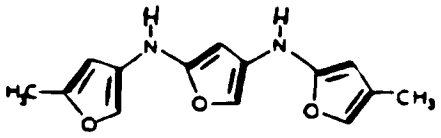


fig (13b)

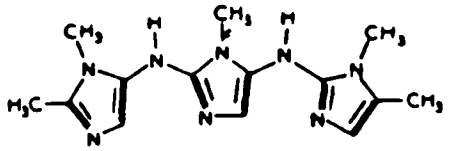
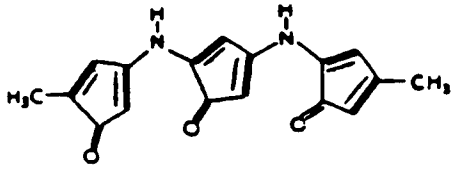


fig (13c)



group of guanine. In order to increase GC specificity of this isolexin-like prototype, the furan rings were replaced my methylimidazole, fig(13b) or cyclopentadienones, fig(13c). However, only a moderate increase in GC binding was obtained.

Further attempts were made to increase GC binding of these isolexins to DNA by replacing their N-H linkers by C=C bonds. These modified ligands were termed "vinylexins". They consisted of either three furan rings connected by C=C linkers, fig(13d) or three imidazole rings connected by C=C linkers, fig(13e). The results showed an increase in the GC specificity of these ligands. From these results he concluded that "specificity" depended on all the steric and energetic components of complex formation. Also, Pullman studied the effects of introducing cationic end groups on the binding of vinylexins. These were derivatives of neutral vinylexins, fig.(13d&e) in which one methyl end group had been replaced by a cationic CH_2-CH_2-$C(NH_2)=NH_2^+$ group. His results for imidazole and furan vinylexins, fig(13f&g) indicated an increase in the complexion energy of these ligands with DNA as well as a moderate reduction in the GC specificity. However, in the case of the monocationic cyclopentadienone derivative, fig(13h), there was a great decrease in GC binding which resulted in producing a compound which showed no specificity and no binding at all. Strength of binding to GC versus AT is specificity. According to Pullman, this was due to the presence of the exocyclic oxygen of the cyclopentadienone ring which caused the ligand to be further away from the minor groove than furan or imidazole vinylexins and the charged end of the ligand could not go deeply into the minor groove. The effect of introducing two charged end groups, CH_2-CH_2-$C(NH_2)=NH_2^+$, on cyclopentadienone derivative, fig(13i), was examined. The results showed an increase in the complexion energy but also a great reduction in the GC binding. He concluded that monocationic vinylexins with non-dipolar vinylic linkers were the best choice in perserving GC specificity.

Pullman also showed the significance of the vinylic linkers that produced GC minor groove binding ligands. He replaced the vinylic linkers of monocationic furan vinylexins, fig(13f), by peptidic bonds to produce a new ligand called "peptilexin", fig(13j). This

fig (13d)

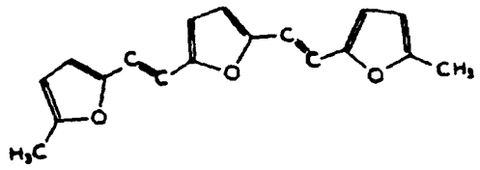


fig (13e)

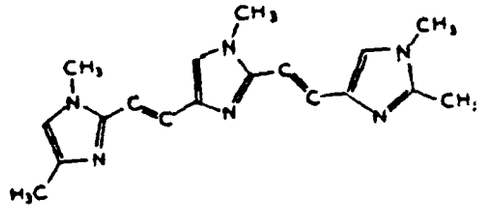


fig (13f)

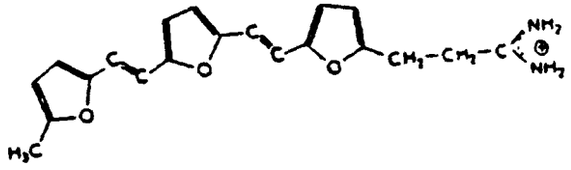


fig (13g)

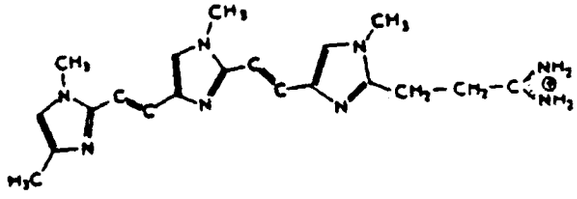
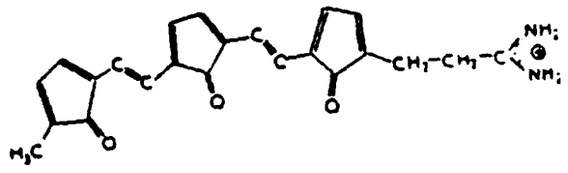


fig (13h)



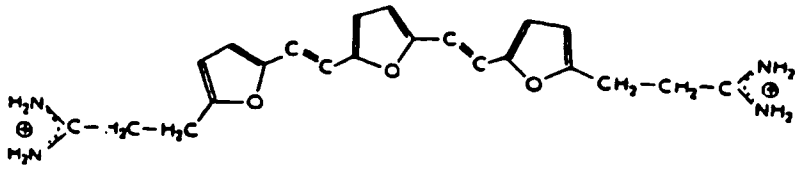


fig (13i)

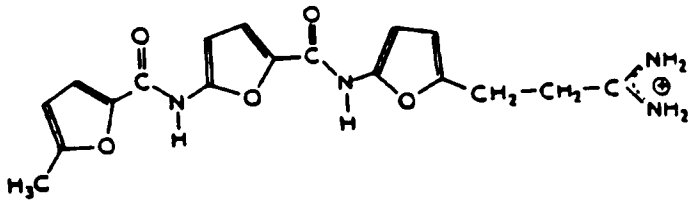


fig (13j)

compound was composed of three furan rings that were connected in-between by peptidic bonds and a cationic CH_2-CH_2-$C(NH_2)=NH_2^+$ group at one end. His results showed a complete disappearance of specificity which was due to a great increase in binding energy with AT sequences.

Next, Pullman reversed in the calculation the orientation of the peptide linkers so that their carbonyl bonds would be directed towards the DNA receptor and their N-H bonds toward the exterior of the complex while before the N-H bonds were directed toward the bases. The two new ligands of this type were called "isopeptilexins". They were composed of either furan rings connected in-between by peptidic bonds or imidazoles connected in-between by peptidic bonds, fig(13k&l). His results indicated that they bound "normally" to the minor groove of the GC sequence and they also refused to remain in the minor groove of the AT sequence and preferred to bind non-specifically to the backbone strands by forming hydrogen bonds between their peptide NH groups and anionic oxygens of the phosphates.

In addition, computations were performed on two monocationic vinylexins. The first compound, fig(13m), was composed of three pyrrole that were connected by C=C and a cationic group, CH_2-CH_2-$C(NH_2)=NH_2^+$, at one end. The orientation of the pyrrole rings for this ligand was such that its pyrrole rings could participate in hydrogen bonding with DNA receptors while the orientation of the pyrrole rings of the second compound, fig(13n), was similar to netropsin or distamycin A that were unable to form hydrogen bonding. These results showed a "remarkable" AT specificity for both compounds but somewhat greater in the case of the first compound hich could participate in hydrogen bonding. He concluded that the presence of hydrogen bonds was not necessary for binding these cationic ligands although it would increase the complexion energy as well as AT specificity. In addition , Pullman and coworkers studied five antibiotics which would bind selectively to AT sequences of B-DNA and he calculated their binding energy to B-DNA.[19] These ligands were SN-18071, NSC-101327, distamycin-2, distamycin-3 and netropsin, fig(14a-d) and fig(4).

fig (13k)

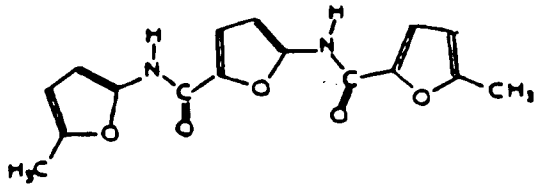
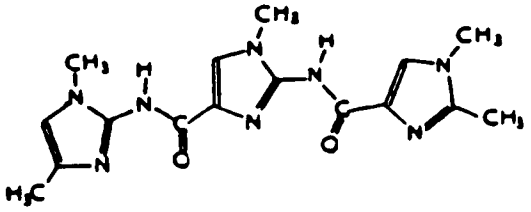


fig (13l)



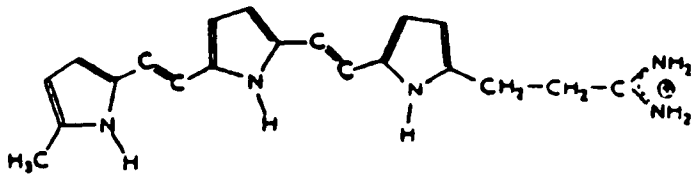


fig (13m)

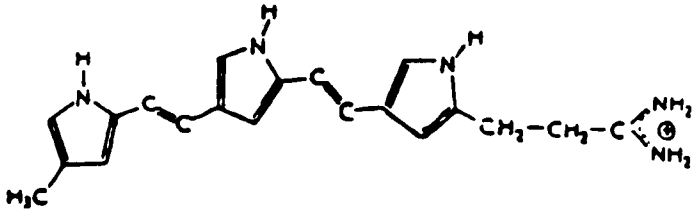
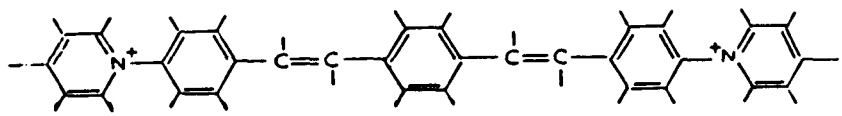


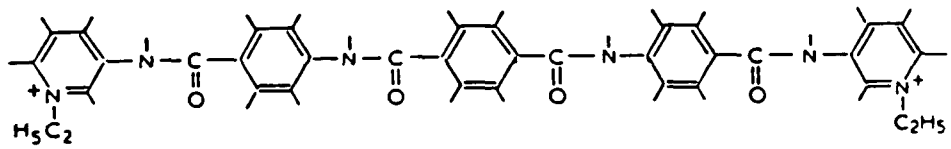
fig (13n)

fig (14a)



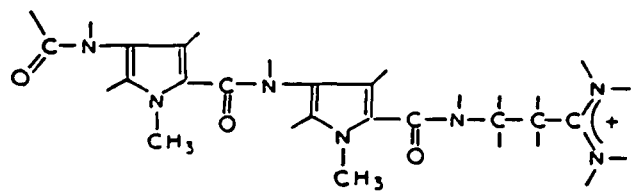
SN-18071

fig (14b)



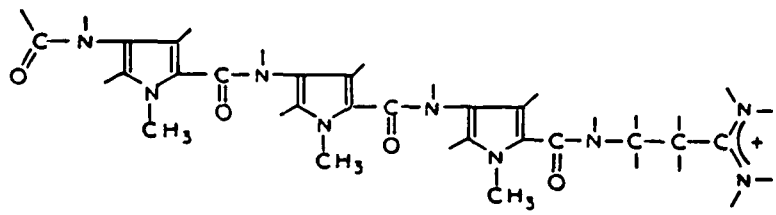
NSC-101327

fig (14c)



Distamycin - 2

fig (14d)



Distamycin-3

SN-18071 ans NSC-101327 are composed of benzene rings linked together by vinylic C=C group and peptidic bonds respectively. Distamycin-2, distamycin-3 and netropsin were consisted of N-methyl pyrrole rings that are connected by peptidic bonds. Pullman observed several features regarding these ligands. First, in terms of hydrogen bonding, SN-18071 is unable to form any hydrogen bonds with DNA while NSC-101327 can form hydrogen bonds through its peptidic groups and the rest of the ligands such as distamycin-2, distamycin-3 and netropsin can participate in hydrogen bonding through their peptidic groups as well as their terminal groups. Second, in terms of ligand charge, three ligands, (SN-18071, NSC-101327 and netropsin) are dicationic while distamycin-2 and distamycin-3 are monocationic. Third, in terms of geometry of ligand, SN-18071 and NSC-101327 have linear conformations due to para substitutions of their rings while distamycin-2, distamycin-3 and netropsin formed curved conformations which would fit better to the shape of the grooves of DNA. His computation on DNA-ligand interaction energies correlated with the properties of these antibiotics. For example, for SN-18071 and NSC-101327, which could not easily bend to fit the shape of DNA grooves and little chance of engaging in hydrogen bonding and none in the case of SN-18071, the ligand-distortion energy with poly(dA) and poly(dT) were -49.3 and -69.1 kcal/mole respectively. However, in the case of the remaining antibiotics, distamycin-2, distamycin-3 and netropsin, their value (-70.7,-85.2, and -87.2 kcal/mole) were slightly greater due to the geometry of ligands which could easily bend to fit into the minor groove of DNA and their ability to hydrogen bond.

Considering all of the components of the interaction energy such as Lennard-Jones, electrostatic and polarization, he concluded that the final order of the optimal complexation energies with the minor groove of poly(dA), poly(dT) was: SN-18071<< distamycin-2< NSC-101327<< netropsin< distamycin-3. Further theoretical studies were done by Sapse and coworkers on two prototype lexitropsins, using the method of ab initio calculation (Hartree-fock)[20]. These DNA minor groove binding agents were related to the antibiotic drug, netropsin, in which each of the two N-methyl-pyrrole moieties of netropsin was

successively replaced in turn by 1-methylimidazole and therefore could recognize AT as well as GC sequences of DNA because of the introduction of an additional nitrogen in the ring, which could accept an hydrogen from the NH_2 group of guanine, fig(15a) and fig(15b).

fig (15a)

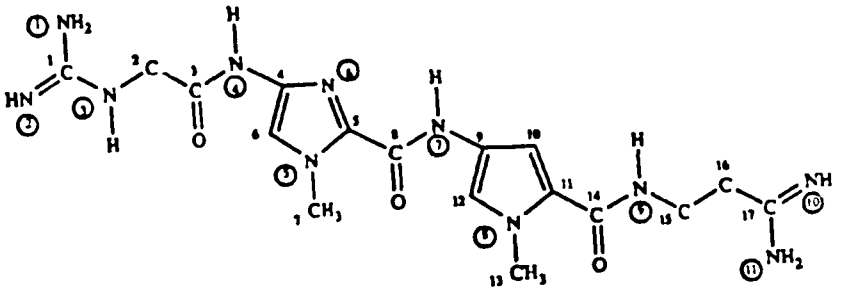
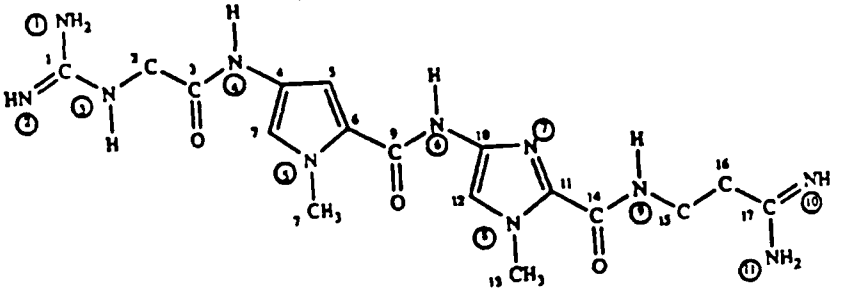


fig (15b)



The calculation was only done for one of the prototype lexitropsin as shown in fig(15a). For the second prototype lexotropsin, the results for the first prototype lexotropsin were used by reversing the positions of the pyrrole and imidazole rings. Due to large sizes of these molecules separate fragments of these lexitropsins were used for calculation. The dihedral angles between the fragments and within the fragments were obtained by optimizing the complete molecules at the STO-3G level, while all the bond lengths and angles were maintained at the 6-31G values obtained by fragement optimization. These dihedral angles were N3C2C3N4, C10C9N7C8, C16C15N9C14, C17C16C15N9, N10C17C16C15.

The optimized structure corresponded to conformation "α", fig(16a). For comparison purposes, two other conformations were taken into consideration by calculating their energy as a single point at the STO-3G level. They called this totally coplanar conformation "β", fig(16b). In this conformation, all of the dihedral angles were set at 180.0°. The third conformation was corresponded to "ϑ" in which the value of the dihedral angle, C10C9N7C8, that defined the dihedral angle between the two rings was 63° and all other dihedral angles were set at 180.0°, fig(16c). These conformations were chosen to determine the effects of the molecular bending on the binding. Sapse and her group concluded that the most stable conformation was "α". They also studied drug-DNA interaction. For this purpose, they used a B-DNA segment $[GCGAATTCGC]_2$. The site targeted for the insertion of the drugs was the AATT sequence. They found that conformation "α" bound relatively poorly due to steric hinderance caused by the guanidinium and amidinium moieties. Conformation "ϑ" was found to insert with the greatest facility into the minor groove of DNA closely followed by conformation "β". The reason for the greatest facility of conformation "ϑ" was found to be related to its dihedral angle between the hetercycles, θ=63 which assisted in isohelical matching of the right handed helix receptor as well as the lack of steric factors within the interior of the minor groove with the terminal guanidinium and amidinium group in this conformation. The energies of the structures are electrostatic.

fig (16a)

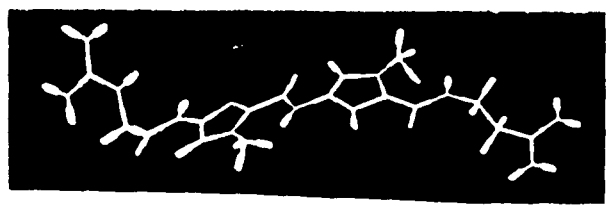


fig (16b)

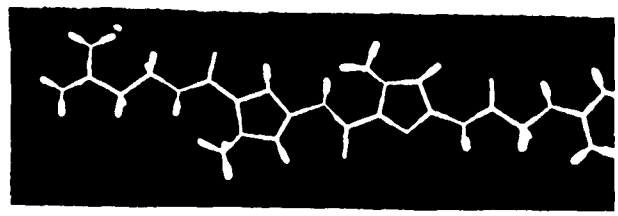
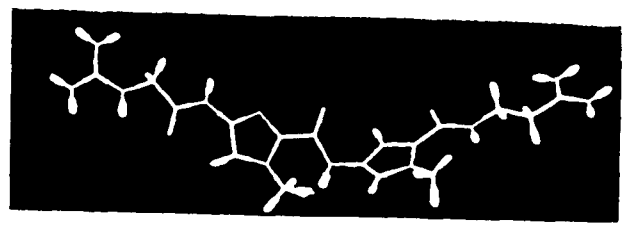


fig (16c)

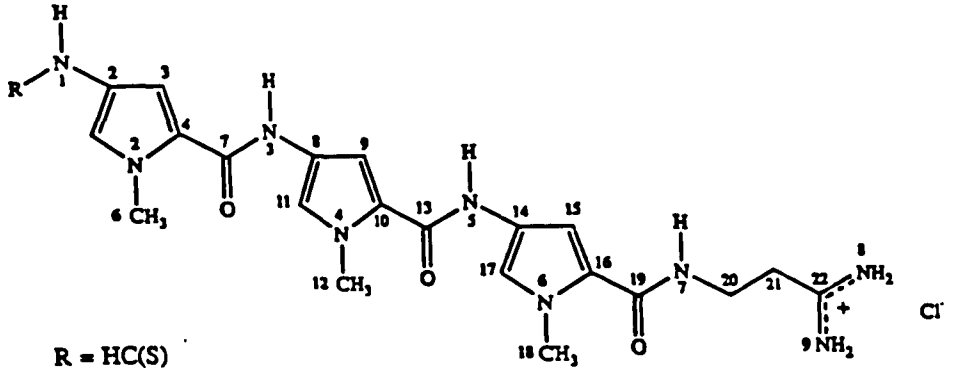


Additional studies were done by Sapse and her group on an amide isostere of the antibiotic distamycin, thioformyldistamycin-3 as shown in fig(17)[21]

NMR experiments on the free drug showed the existance of two conformations at the formyl group. In the Z conformation the sulfur atom was directed away from the minor groove while in the E conformation it was directed toward the minor groove of DNA, as shown in fig(18a&b). The E-conformation predominated in fresh solution while the Z-conformation predominated when the solution was maintained at room temperature for several hours.

The method of quantum chemical ab initio calculation at Hartree-fock level was used to study the preferred conformation of this drug. For this purpose, Gaussian-86 computer program applying a 3-21G* basis set which added d orbitals on the sulfur was used. Since it was difficult to do an ab initio calculation with a valence-split basis set including polarization for the whole molecule, a model, 3-(thioformylamino)-N-methylpyrrole, in the E and Z conformations was used. All the parameters of the molecule were optimized while keeping the other parameters at the values obtained by the optimization of the E structure. The starting value for this angle was set at 100°. After performing optimization, the angle adopted a value of 180°. The E conformation showed a greater stability than the Z conformation by 17.3 kcal/mole.

fig (17)



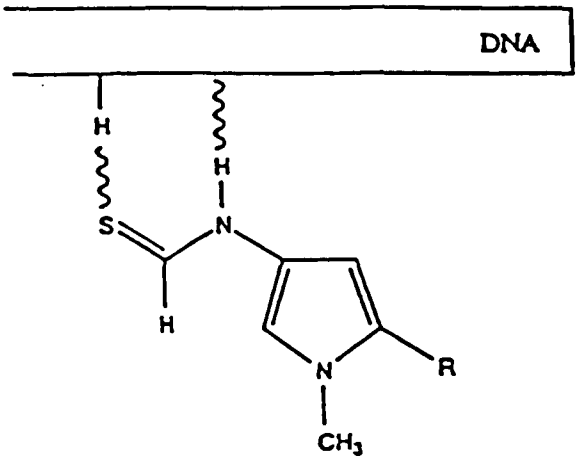


fig (18a)

E - configuration

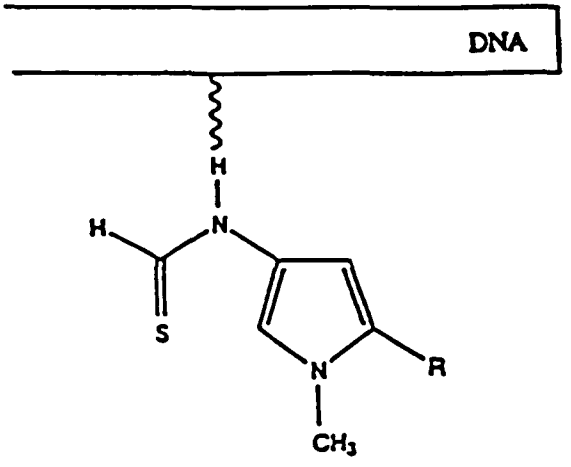


fig (18b)

Z- configuration

Next in order to estimate a possible rotational barrier to the conversion of E to Z, the SC1N1C2 angle was set at 90° and all other parameters of the molecule were optimized. From the calculation, they concluded that the two conformations of thioformyl distamycin were separated by a rotational barrier of more than 26 kcal/mole. Therefore two different types of binding were expected assuming that each conformation would bind separately to DNA.

Experimentally it was found that regardless of the initial conformation of this drug upon approach to the DNA , the molecule would adopt the Z conformation in the minor groove of B-DNA in the complex. Also, footprinting studies showed that both the parent molecule, distamycin and thioformyl distamycin had similar binding strength and affinities for DNAs.

Sapse and her group studied the guanidinium ion $C(NH_2)_3^+$ which is a part of the antibiotic, netropsin.[22]

The Hartree-Fock approximation, using the Gaussian-70 computer program with an STO-6-31G basis set was used.

The lowest energy conformation for the guanidinium ion corresponded to the planar geometry of symmetry D_{3h} with a total energy of -204.45 au, as shown in fig(19). A value of 1.33 Å was found for the CN bond length which lies between the values of 1.29 and 1.47 Å respectively for single and double CN bond distances.

Planer D_{3h}

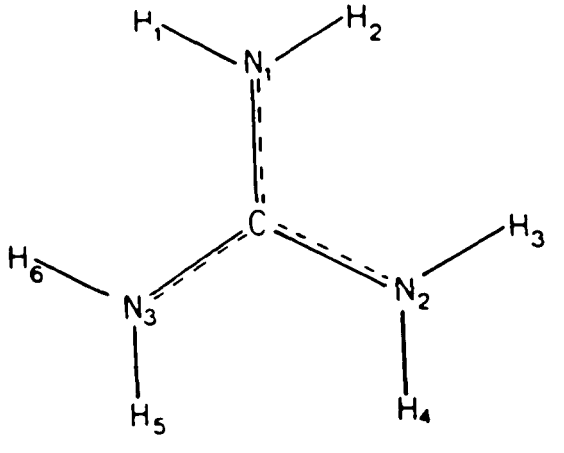


fig (19)

They determined that the ion in a planar geometry was stabilized by delocalized π bonds associated with the p-type atomic orbitals orthogonal to the plane at carbon and the nitrogens. Therefore, it was energetically favorable for guanidine to accept a proton due to this delocalization.

Next, they determined the rotational barrier associated with the single, double and triple rotation of NH2 groups about the corresponding CN bond lines. To obtain these values, the energy difference for the molecule's planar geometry and that corresponding to a 90° rotation of one (two or three) of the NH2 groups about the CN bonds were taken. The calculated rotational barriers for single, double and triple rotation were found to be 14.73, 45.34 and 111.63 kcal/mole respectively. From their results, they concluded that the first barrier was a measure of the energy required to break the π bonding in the CN group about which the rotation was made and the remaining CN bonds not involved in rotation became more strongly π bonded than before. Also, a second rotation was energetically more costly than a first rotation due to the fact that it involved twisting about a π bond whose strength had been increased as a reaction to the first twisting. Also, this was the case for the third rotation.

Next, they also studied the geometries of the guanidinium ion associated with the rotated conformations. Their results showed that rotating a NH2 group about a CN bond line by 90° resulted in the elongation of CN bond distance from 1.33 to 1.38 A and at the same time it shortened the CN bonds with NH2 groups that remained in the plane of the molecule to a value of 1.31 A. They found that twisting about any two CN bonds had the effect of lengthening these bonds but shortening the remaining CN bond even further than before.

Summary of Research

In this research study, the method of quantum chemical ab initio calculations using the Hartree-Fock method with 6-31G basis set or 3-21G basis set as implemented by the Gaussian-90 and Gaussian-86 program are used.

In the first part of this research work, 6-31G basis set implemented by the Gaussian-90 program is used to do ab initio calculations of the geometry and energy of a small lexitropsin, called "amidinomycin". The second part contains proton affinities calculations on heteroatomic rings such as imidazole, oxazole and thiazole in order to estimate their ability to undergo hydrogen bonding with the amino group of guanine or cytosine. The last part of this work compares the binding energies of uracil which was taken as a model for thyamine to either guanidinium ion present in some lexitropsins such as netropsin or to the aminopyrrolidium ion moiety as present in other lexitropsins such as Anthelvencin, Kikumycin and Noformycin.

II) METHOD

Molecular quantum calculations can be classified into two groups : 1) semiempirical calculation 2) ab initio calculation. Both semiempirical and ab initio calculation use the Born-Oppenheimer approximation that the electronic wave function is unaffected by nuclear motion. However, in semiempirical calculations, one uses a simpler Hamiltonian rather than "correct" molecular Hamiltonian and include the experimental data or parameters that can be adjusted to experimental data into the calculation. An example of this type of calculation is the Hückel molecular orbital treatment of conjugated hydrocarbons which uses a one electron Hamiltonian and takes the bond integrals as adjustable parameters rather than quantities to be calculated. In an ab initio calculation, one uses the "correct" Hamiltonian and seeks a solution without using experimental data other than the geometry of the molecule. The theory involves expanding the atomic or molecular orbitals as a linear combination of basis function.[23]

$$\psi_i = \sum_{\mu=1}^{K} C_{\mu i}\, \Phi_\mu \qquad i = 1, 2, \ldots\ldots, K \qquad \text{eq.(1)}$$

Where "ψ" is a molecular or atomic orbital and "Φ_μ" are selected slater type, Gaussian or some other functions that are referred to as a basis set.

The closed-subshell Hartree-Fock molecular orbitals satisfy :

$$\hat{f}(r_1)\psi_i(r_1) = \varepsilon_i\psi_i(r_1) \qquad \text{eq.(2)}$$

Where "ε_i" is the orbital energy, "ψ_i" is the unknown molecular orbitals and "f" is the Hartree-Fock operator. Substitution of this expression, eq.(1) into the Hartree-Fock equation, eq.(2) and using index "v" gives :

$$f(1)\sum_{v} C_{vi}\,\phi_v(1) = \varepsilon_i \sum_{v} C_{vi}\,\phi_v(1) \qquad \text{eq.(3)}$$

Multiplication by $\phi_\mu(1)$ on the left and integrating gives :

$$\sum_{v} C_{vi} \int dr_1 \phi^*_\mu(1) f(1) \phi_v(1) = \varepsilon_i \sum_{v} C_{vi} \int dr_1 \phi^*_\mu(1) \phi_v(1) \qquad \text{eq.(4)}$$

Where "S_{uv}" is the overlap marix which has elements $S_{uv} = \int dr_1 \phi^*_\mu(1)\phi_v(1)$ eq.(5)

and "F_{uv}" is the fock matrix which has elements $F_{uv} = \int dr_1 \phi^*_\mu(1) f(1) \phi_v(1)$ eq.(6)

Both overlap matrix "S_{uv}" and fock matrix "F_{uv}", are K x K Hermitian (real and symmetric) matrix. Using the definition of "F_{uv}" and "S_{uv}" leads to Roothaan equations[24]:

$$\sum_{v} F_{uv} C_{vi} = \varepsilon_i \sum_{v} S_{uv} C_{vi} \qquad \text{eq.(7)} \quad i = 1,2,\ldots,K$$

F_{uv} is the fock matrix which can be written as :

$$F_{uv} = H^{core}_{uv} + \sum_{\lambda\sigma} P_{\lambda\sigma}[(uv \,/\, \lambda\sigma) - \tfrac{1}{2}(u\lambda \,/\, v\sigma)] \qquad \text{eq.(8)}$$

Where "$P_{\lambda\sigma}$" is the density matrix :

$$P_{\lambda\sigma} = 2 \sum_{i}^{occ} c_{\lambda i} c_{i} \qquad \text{eq.(9)}$$

H^{core}_{uv} is the matrix of the core Hamiltonian (Kinetic energy and potential in the field of fixed nuclei), and $(uv \,/\, \lambda\sigma)$ is the two-electron integral :

$$(uv \,/\, \lambda\sigma) = \int\int \phi_\mu(1)\phi_v(1) \frac{1}{r_{12}} \phi_\lambda(2)\phi'(2) dr_1 dr_2 \qquad \text{eq.(10)}$$

Using the Gaussian expansions, all the integrals can be evaluated by the methods due to Boys.[25] Equation (7) for LCAO coefficients "C_{vi}" can then be solved in an iterative manner.

leading to calculated energies, density matrices, atom populations and electric dipole moments.

The calculations are performed on computer, using the Gaussian programs as mentioned before. The input consists of specification of the required basis set, the charge and multiplicity of the molecule, as well as its initial geometry. The program uses this information to calculate the cartesian (x,y,z) coordinates of the atoms, total number of electrons and the orbital occupancies. Then, it calculates the various one and two electron integrals. Next, an initial guess is produced and the program uses this initial guess and do an iterative SCF calculation. The solution to the SCF equations is improved by going from cycle to cycle until the electronic energy is at minimum and the density matrix does not change. At this point, the calculation is defined to be converged. Next, for a geometry optimization, the gradient method is used to estimate geometry corresponding to minimum energy. When the minimum of energy is obtained, the optimization is complete and the program calculates the Mullikan Population Analysis of the optimized molecule. For a single point calculation, after obtaining the solution to the SCF equation, the program moves to population analysis which calculates the atomic charges, dipole moments, etc...

The Mullikan Population Analysis can be obtained by using the charge density expression for each occupied molecular orbital, Ψ_a, containing two electrons,

$$\rho(r) = 2\sum_{a}^{N/2} \int dr \, |\Psi a(r)|^2$$

By inserting equation (1) into the charge density expression, one can obtain the following expressions:

$$= 2\sum_{a}^{N/2}\sum_{v} c_{va}^{*} \phi_v(r) \Sigma c_{\mu a} \phi_\mu(r)$$

$$= \sum_{uv} [2\sum_{a}^{N/2} c_{ua} c_{va}^{*}] \phi_u(r) \phi^{*}_v(r)$$

$$= \sum_{uv} P_{uv} \phi_u(r) \phi^{*}_v(r)$$

Where $P_{uv} = 2\sum_{a}^{N/2} c_{ua} c^{*}_{va}$ is the charge-density bond order matrix

$$= \sum_{\mu}\sum_{\nu} P_{\mu\nu}S_{\mu\nu} = \sum_{\mu}(PS)_{\mu\mu}$$

and where $S_{\mu\nu}$ is the overlap matrix.

$(PS)_{\mu\nu}$ can be interpreted as the number of electrons to be associated with . This is called the Mulliken Population analysis. The net charge associated with an atom is then given by:

$$q_A = Z_A - \sum_{\mu \in A}(PS)_{\mu\mu}$$

where Z_A is the charge of atomic nucleus A. The index of summation indicates that only the basis function centered on A are summed.

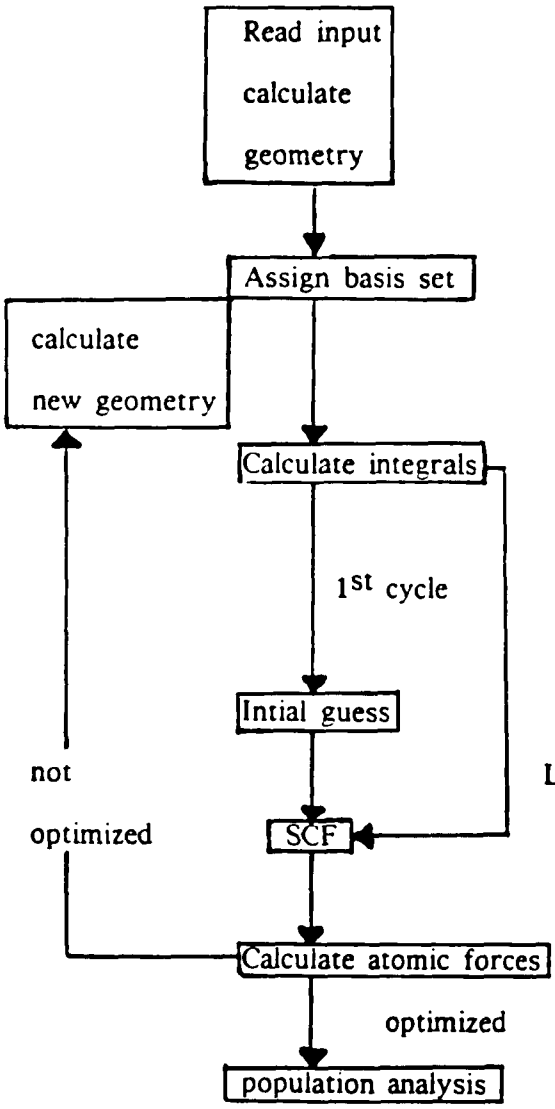


Fig (20a)

Flow charts for an ab initio optimization

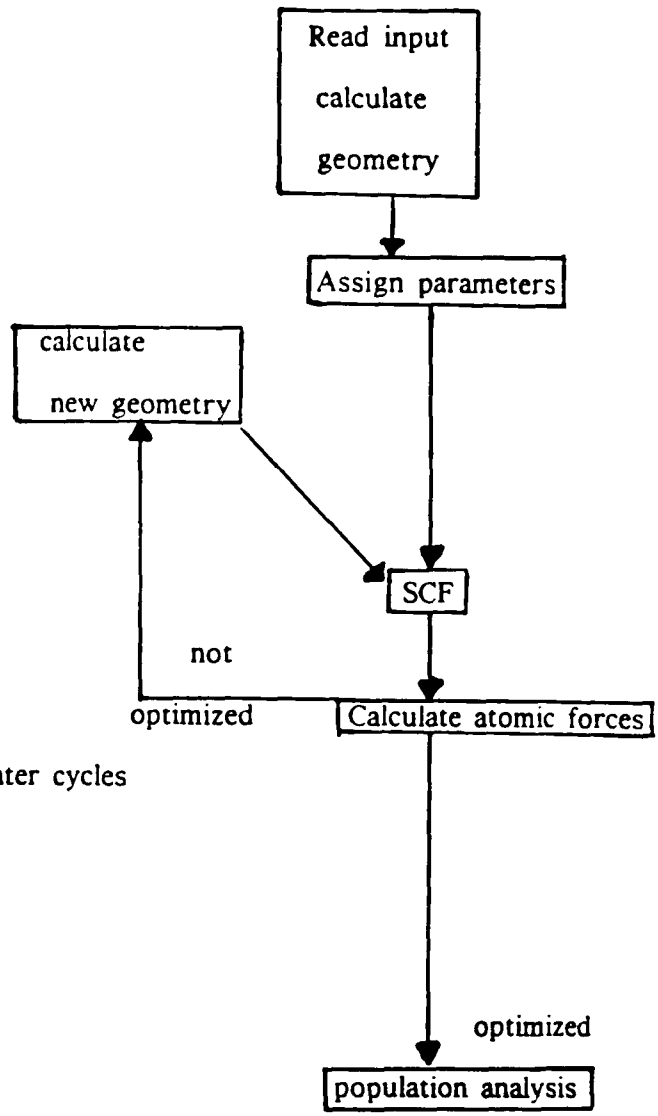


Fig (20b)

Flow charts for a semi-empirical calculation

There are many types of basis sets that can be used depending on the type of calculation. The most common type of basis sets used in an ab initio calculation are Gaussian type orbital basis sets. (GTO)[26]. In these basis set, each atomic orbital consists of a number of Gaussian probability functions that have advantages over other types of basis set. Their integrals are much faster computationally than Slater orbitals. The simplest level of basis is the minimal which corresponds to one Slater function per atomic orbital. This type of basis set is called STO-NG which stands for Slater-type orbitals[27] defined by "N" Gaussian functions each. This means that each atomic orbital consists of "N" Gaussian functions that are added together. The coefficients of the Gaussian functions are selected so that it gives as good a fit as possible to the corresponding Slater-type orbitals. One example of a minimal basis set is STO-3G. This basis set is very economical since it has one basis function or atomic orbital per hydrogen atom (the 1s), five per atom from Li to Ne : ($1s,2s,2p_x,2p_y,2p_z$) and nine per atom for the second-row elements Na to Argon ($1s,2s,2p_x,2p_y,2p_z,3s,3p_x,3p_y$,and $3p_z$). The greatest problem with this type of basis set is its inability to expand or contract its orbital to fit the molecular environment since its exponent is fixed. The next level is split-valence in which two Slater function are used for each valence atomic orbital. The split-valence basis sets are contracted which means each is a linear combination of a number of primitive Gaussian functions. A computational efficiency can be obtained if the exponents of the Gaussian primitives are shared between different basis functions. At the split-valence level, the primitive exponents between "S" and "P" functions are shared for the valence functions. A series of basis sets are defined and are named as K-LMG where K, L, M are integers. Such a basis set for a first row element (Li to Ne) consist of an s-type inner-shell function with K Gaussians, an inner set of valence s- and p-type functions with "L" Gaussians and another outer sp set with "M" Gaussians. Both valence sets have shared exponents. For hydrogen, only two s-type valence functions with "L" and "M" Gaussians are used. For atoms H and He, the basis functions are defined as :

$$\phi_S'(r) = \sum_{K=1}^{L} d_{S,k}' \, g_S(\alpha_k', r)$$

$$\Phi''_S(r) = \sum_{K=1}^{M} d''_{S,k}\, g_S(\alpha''_k, r)$$

For atoms Li to Ne, L-LMG basis functions are defined as :

$$\Phi_{1S}(r) = \sum_{K=1}^{K} d_{1s,k}\, g_S(\alpha_{1k}, r)$$
$$\Phi'_{2S}(r) = \sum_{K=1}^{L} d'_{2s,k}\, g_S(\alpha'_{2k}, r)$$
$$\Phi'_{2p}(r) = \sum_{K=1}^{L} d'_{2p,k}\, g_p(\alpha'_{2k}, r)$$
$$\Phi''_{2S}(r) = \sum_{K=1}^{M} d''_{2s,k}\, g_S(\alpha''_{2k}, r)$$
$$\Phi''_{2p}(r) = \sum_{K=1}^{M} d''_{2p,k}\, g_p(\alpha''_{2k}, r)$$

Where "g_S" and "g_p" are normalized s- and p-type Gaussian functions, respectively.

One example of a split-valence basis set is 4-31G[28] which uses inner-shell expansions of four Gaussian function and two valence shells consisting of three and one Gaussians. Another larger basis set is 6-31G[29]. In this basis set, the inner-shell functions are written in terms of a linear combination of six Gaussians and two valence shells are represented by three and one Gaussian primitives. A split-valence basis set with fewer primitive Gaussians than the 4-31G basis set is 3-21G[27] which uses three primitive Gaussians for the core orbitals and a two/one split for the valence functions. The next step in improving a basis set is addition of d-orbitals for all heavy (non-hydrogen atoms). The most common polarization basis set is 6-31G*[30] which uses six primitive Gaussians for the core orbitals, a three/one split for s- and p-valence orbitals and a single set of d-functions which is equivalent to five d- and one s-orbitals. A further improvement can be obtained by using a 6-31G**[31] basis set in which a set of p-orbitals are added to each hydrogen in the 6-31G*basis set.

When performing an ab initio calculation, it is desirable to optimize the geometry with the largest basis set possible. For larger systems, smaller basis sets have to be used and one can increase the accuracy by performing single point calculations with better basis sets.

TABLE 1.

OPTIMIZED BOND LENGTHS (A) AND ANGLES (DEGREES) FOR AMIDINOMYCIN

PARAMETERS	CALCULATED	EXPERIMENTAL[36]
Bond lengths:		
C_1C_2	1.538	1.53
C_2C_3	1.544	1.57
C_3C_4	1.557	1.52
C_5C_4	1.555	1.56
N_1C_1	1.546	1.51
C_6C_3	1.358	1.53
N_2C_6	1.340	1.34
OC_6	1.247	1.24
C_7N_2	1.457	1.47
C_8C_7	1.555	1.54
C_9C_8	1.514	1.50
N_3C_9	1.309	1.31
N_4C_9	1.317	1.33
HN_1	1.009	---
HC_1	1.081	---
HC_2	1.080	---
HC_3	1.082	---
HC_4	1.081	---
HC_5	1.081	---

TABLE 1 (continued)

OPTIMIZED BOND LENGTHS (A) AND ANGLES (DEGREES) FOR AMIDINOMYCIN

PARAMETERS	CALCULATIONS	EXPERIMENTAL[36]
Angles:		
$C_3C_2C_1$	108.3	105.0
$C_4C_3C_2$	107.6	106.0
$C_5C_4C_3$	107.9	107.0
$N_1C_1C_5$	109.9	111.0
$C_6C_3C_4$	109.3	109.0
$N_2C_6C_3$	117.5	115.0
OC_6C_3	120.7	121.0
$C_7N_2C_6$	122.9	123.0
$C_8C_7N_2$	112.9	110.0
$C_9C_8C_7$	113.4	111.0
$N_3C_9C_8$	119.3	118.0
$N_4C_9C_8$	119.3	119.0
HN_1C_1	111.5	---
$N_4C_9N_3$	---	122.0
Dihedral Angles:		
$C_5C_4C_3C_2$	-3.5	---
$N_1C_1C_5C_4$	114.2	---
$C_6C_3C_4C_2$	119.7	---
$N_2C_6C_3C_4$	103.5	---

TABLE 1 (continued)

OPTIMIZED BOND LENGTHS (A) AND ANGLES (DEGREES) FOR AMIDINOMYCIN

PARAMETERS	CALCULATIONS	EXPERIMENTAL[36]
Dihedral Angles:		
$OC_6C_3N_2$	180.0	---
$C_7N_2C_6C_3$	180.0	---
$C_8C_7N_2C_6$	73.6	---
$C_9C_8C_7N_2$	290.3	---
$N_3C_9C_8C_7$	57.4	---
$N_3C_9C_8N_3$	180.0	---
$HN_1C_1C_2$	134.5	---
$HC_1C_5C_2$	124.3	---
$HC_3C_2C_6$	109.6	---

TABLE 2. ATOMIC CHARGES WITH HYDROGENS SUMMED INTO HEAVY ATOMS (eu) FOR NONCHARGED AMIDINOMYCIN

AMIDINOMYCIN 5

C1	0.258
C2	0.057
C3	-0.091
C4	0.036
C5	0.007
N1	-0.239
C6	0.774
N2	-0.481
O	-0.650
C7	0.372
C8	-0.012
C9	0.576
N3	-0.434
N4	-0.173

TABLE 3

OPTIMIZED BOND LENGTHS (A) AND ANGLES (DEGREES) FOR STRUCTURES 1-6

PARAMETERS	STRUCTURES					
	1	2	3	4	5	6
C1-N1	1.382	1.392	-	-	-	-
C2-N1	1.309	1.303	-	-	-	-
C2-N2	1.373	1.313	1.315	1.261	1.272	1.260
N2-C3	1.377	1.405	1.384	1.432	1.396	1.442
C1-C3	1.347	1.351	1.340	1.326	1.336	1.343
C1-O1	-	-	1.390	1.463	-	-
C2-O1	-	-	1.390	1.463	-	-
C1-S1	-	-	-	-	1.806	1.786
C2-S1	-	-	-	-	1.807	1.830
N1-H5	-	0.996	-	-	-	-
O1-H5	-	-	-	0.961	-	-
S1-H5	-	-	-	-	-	1.346
C1N1C2	106.1	110.0	-	-	-	-
N1C2N2	110.7	110.0	-	-	-	-
C2N2C3	106.1	107.0	105.8	107.6	112.1	114.5
C1O1C2	-	-	104.9	106.3	-	-
O1C2N2	-	-	111.5	109.6	-	-
C1S1C2	-	-	-	-	85.4	92.4
S1C2N2	-	-	-	-	115.9	109.3
C1N1H5	-	125.3	-	-	-	-
C1O1H5	-	-	-	126.7	-	-
C1S1H5	-	-	-	-	-	-

TABLE 3 (continued)

Optimized bond lengths (A) and angles (degrees) for structures 7-10

Parameters	Structures			
	7	8	9	10
C1-N1	1.378	1.395	1.377	1.388
C2-N1	1.277	1.324	1.316	1.308
C2-N2	1.329	1.315	1.375	1.316
N2-C3	1.384	1.384	1.372	1.407
C1-C3	1.373	1.349	1.370	1.355
N2-C5	-	-	1.450	1.451
C2-C4	1.470	1.503	-	-
C4-N3	1.340	1.335	-	-
C4-O2	1.232	1.220	-	-
N1-H5	-	0.997	-	0.997
C1N1C2	107.6	109.0	106.9	110.0
N1C2N2	111.4	109.0	110.7	110.0
C1N1H5	-	124.4	-	125.2
C2N2C5	-	-	137.3	126.2
N2C2C4	121.9	119.2	-	-
C2C4N3	115.4	118.8	-	-
C2C4O2	120.3	115.3	-	-

TABLE 4

Total Energies (au)

Structure	Basis set	
	6-31g	6-31g*
1	-224.7096	-224.8129
2	-225.1023	-225.1945
3	-244.5003	-224.6246
4	-244.7758	-244.8867
5	-567.1631	-567.2829
6	-567.3816	-567.5131
7	-392.8014	-392.9716
8	-392.4331	-392.6085
9	-263.2327	-263.8434
10	-264.1206	-264.2327

TABLE 5

Proton Affinities (kcals/mole)

Structure	Basis set	
	6-31g	6-31g*
1	245.92	239.45
3	172.87	164.47
5	137.11	144.45
7	231.11	227.85
9	250.94	244.29

TABLE 6

Total atomic charges on Structures 1-6 (eu), as obtained by the use of the 6-31g* basis set.

Atom	Structures					
	1	2	3	4	5	6
C1	-0.0540	0.0276	0.0866	0.115	-0.436	-0.378
C2	0.266	0.423	0.344	0.411	-0.110	-0.0121
C3	-0.000	0.025	-0.054	-0.012	0.051	0.070
C4	-	-	-	-	-	-
H5	-	0.470	-	0.572	-	0.294
N1	-0.520	-0.686	-	-	-	-
N2	-0.725	-0.692	-0.504	-0.398	-0.436	-0.354
O1	-	-	-0.572	-0.707	-	-
S1	-	-	-	-	0.234	0.37

TABLE 6 (continued)

Total atomic charges on Structures 7-10 (eu), as obtained by the use of the 6-31g* basis set.

Atom	Structure			
	7	8	9	10
C1	-0.055	0.028	-0.051	0.022
C2	0.495	0.622	0.264	0.427
C3	0.004	0.039	0.000	0.031
C4	0.764	0.833	-0.281	-0.331
N1	-0.590	-0.783	-0.536	-0.703
N2	-0.744	-0.694	-0.588	-0.580
N3	-0.920	-0.919	-	-
O2	-0.654	-0.577	-	-
H5	-	0.490	-	0.465

TABLE 7

Bond length (A) used for complex (1) and (2) in 6-31G basis set

Bond lengths	Complex (2)	Complex (1)
N1H1	0.993	0.993
C1N1	1.374	1.374
C1O1	1.222	1.222
C1N2	1.371	1.371
N2H2	0.996	0.996
C2N2	1.392	1.392
C2O2	1.221	1.221
C2C3	1.454	1.454
C3H3	1.067	1.067
C3C4	1.334	1.334
C4H4	1.07	1.07
O2H5*	2.0054	---
N3H5	0.9938	0.9938
N3C5	1.32	1.3263
C5C6	1.504	---
C6C7	1.542	---
C7C8	1.554	---
C5N4	1.307	---
N4H6	0.997	---
N4H7	0.997	---
C6H8	1.083	---

TABLE 7 (continued)

Bond length (A) used for complex (1) and (2) in 6-31G basis set

Bond lengths	Complex (2)	Complex (1)
C6H9	1.083	---
C8H12	1.083	---
C8H13	1.083	---
O2H6*	1.9786	---
O1H5*	---	1.9849
N3H6	---	0.9938
O2H10*	---	6.5348
N5H2	---	5.0930

* Bond length optimized in the complex at 3-21G level

TABLE 7 (continued)

Angles :	Complex (1)	Complex (2)
C1N1H1	115.781°	115.781°
O1C1N1	122.595°	122.595°
N1C1N2	114.042°	114.042°
C1N2H2	116.061°	116.061°
C1N2C2	127.177°	127.177°
N2C2O2	120.393°	120.393°
C3C2N2	114.395°	114.395°
C2C3H3	118.027°	118.027°
C2C3C4	119.503°	119.503°
C3C4H4	122.734°	122.734°
H5N3H6	116.777°	---
C5N3H5	121.835°	120.0°
C5N4H8	121.835°	---
N4C5N5	119.856°	---
N3C5N5	119.856°	125.0°
C1O1H5*	140.021°	---
N3H5O1*	141.771°	---
C5N4H7	121.835°	123.3°
C5N5H10	121.835°	---
C5N5H9	121.835°	---

TABLE 7 (continued)

Angles	Complex (1)	Complex (2)
C2O2H5*	----	139.272
N3H5O2*	----	138.461
N3C5C6	----	110.6
C5C6C7	----	103.6
C6C7C8	----	104.9
C5N4H6	----	123.3
C5C6H8	----	110.1
C5C6H9	----	110.1
C6C7H10	----	110.7
C6C7H11	----	110.7
N3C8H12	----	110.0
N3C8H13	----	110.0

* Angles optimized in the complex at 3-21G level

TABLE 7 (continued)

Dihedral Angles:	Complex (1)
H1N1C1O1	0°
C1N1C2C3	0°
N2C2C3C4	0°
N2C1O1H5	0°
N5C5N3H5	0°
N3C5N4H8	0°
N3C5N5H10	0°
H1N1C1N2	180°
H2N2C1N1	180°
O1C1N2C2	180°
C1N2C2O2	180°
N2C2C3H3	180°
C2C3C4H4	180°
C1O1H5N3	180°
O1H5N3H6	180°
C5N3H5H6	180°
N3C5N4N5	180°
N3C5N4H7	180°
N3C5N5H9	180°

TABLE 7 (continued)

Dihedral Angles:	Complex (2)
H1N1C1O1	0°
C1N2C2C3	0°
N2C2C3C4	0°
N2C2O2H5	0°
C5N3H5O2	0°
N6N4C5H5	0°
N7N4C5H5	180°
H1N1C1N2	180°
H2N2C1N1	180°
O1C1N2C2	180°
C1N2C2O2	180°
H3C3C2N2	180°
C2C3C4H4	180°
C2O2H5N3	180°
C6C5N3H5	180°
C7C6C5N3	-12.7°
C5C6C7C8	19.63°
C6C5N4N3	180°
C5C6C7H8	120.5°
C5C6C7H9	-120.5°
C6C7C8H10	121°
C6C7C8H11	-121°
C7C8N3H12	120°
C7C8N3H13	-120°

TABLE 8

Total atomic charges (eu) for complex #(1) and complex #(2) in 3-21g basis set

Atom #	Complex (1)	Complex (2)
N1	-1.01	-1.01
H1	0.402	0.422
C1	1.30	1.24
O1	-0.717	-0.597
N2	-1.05	-1.06
H2	0.407	0.407
C2	0.921	0.989
O2	-0.584	-0.716
C3	-0.4	-0.470
H3	0.314	0.289
C4	0.228	0.269
H4	0.323	0.330
H5	0.431	0.421
N3	-0.962	-0.892
H6	0.389	0.433
C5	1.32	0.961
N5	-0.964	----
N4	-0.945	-0.968
H8	0.402	0.303
H7	0.402	0.396
H10	0.433	0.276
H9	0.387	0.292

TABLE 8 (Continued)

Total atomic charges (eu) for complex #(1) and complex #(2) in 3-21g basis set

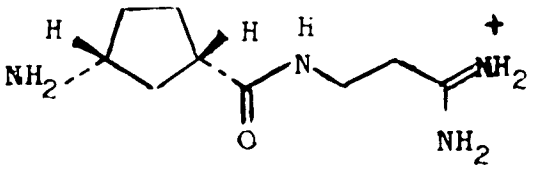
Atom #	Complex (1)	Complex (2)
C6	----	-0.511
C7	----	-0.461
C8	----	-0.169
H11	----	0.274
H12	----	0.273
H13	----	0.270

TABLE 9

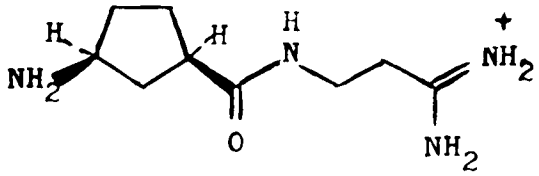
Energies and bonding energies for complex (1) and (2) in 3-21G basis set

Complex #	Energy (au)*	E binding in kcal/mole
1	-613.6050	**-26.035**
2	-674.1680	-27.798

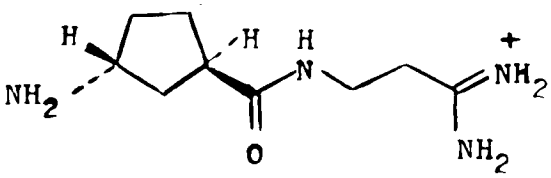
* Energies are obtained by using the 6-31G optimized geometry of the subsystems



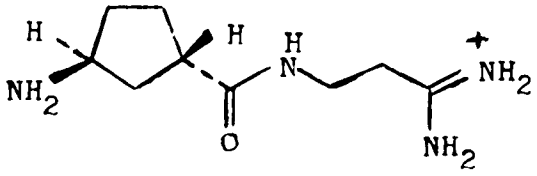
1R-3S



1S-3R



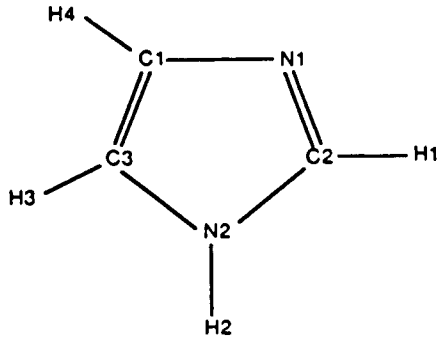
1R-3R



1S-3S

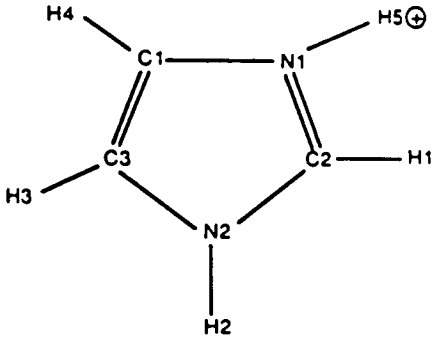
FOUR ISOMERS OF AMIDINOMYCIN

fig (21)



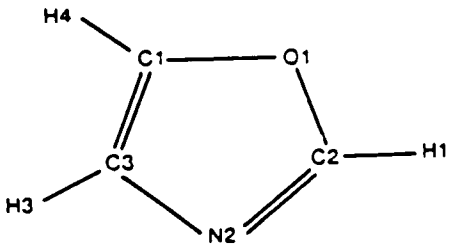
STRUCTURE 1

Fig (22a)



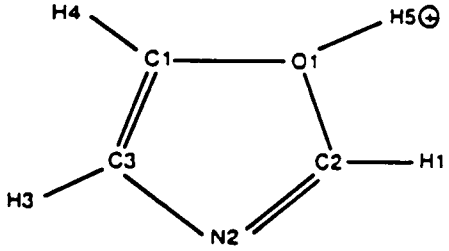
STRUCTURE 2

Fig (22B)



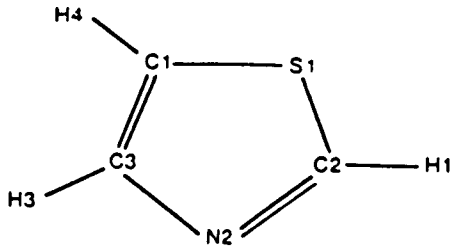
STRUCTURE 3

Fig (22c)



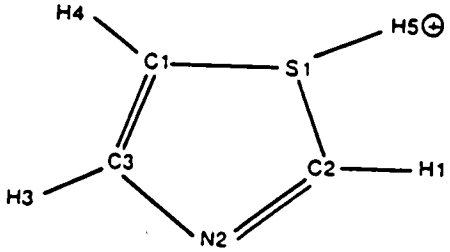
STRUCTURE 4

Fig (22d)



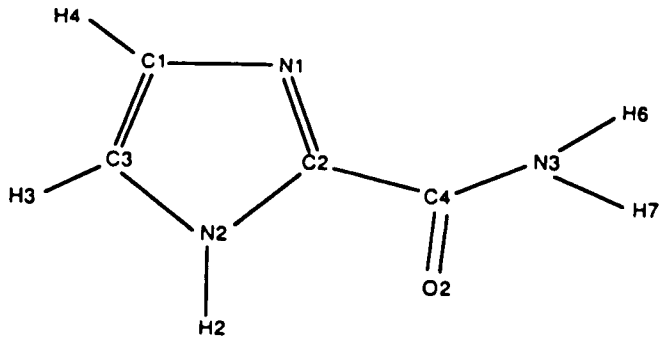
STRUCTURE 5

Fig (22e)



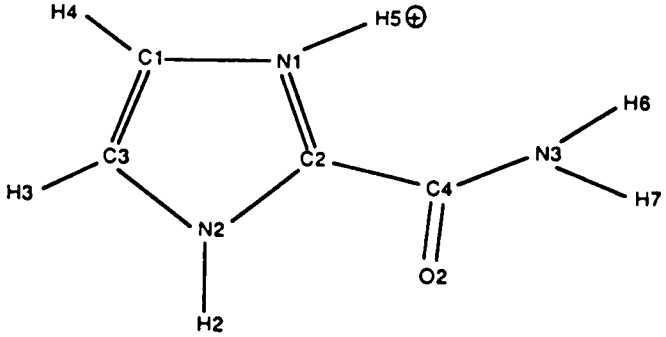
STRUCTURE 6

Fig (22f)



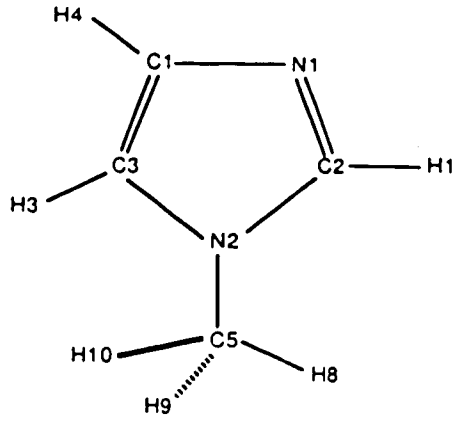
STRUCTURE 7

Fig (22g)



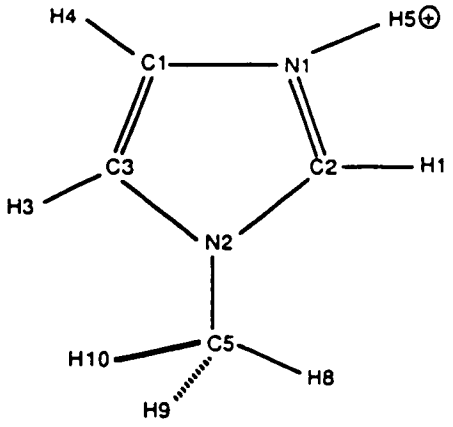
STRUCTURE 8

Fig (22h)



STRUCTURE 9

Fig (22i)



STRUCTURE 10

Fig (22j)

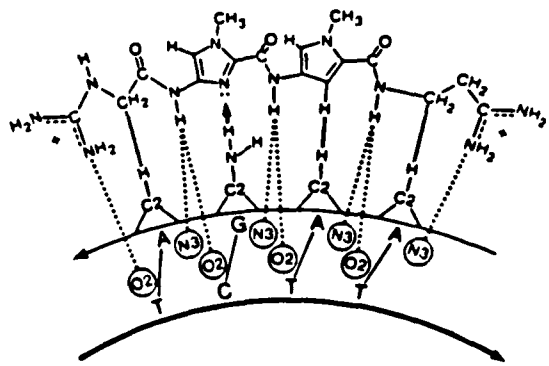


fig (23a) BINDING OF IMIDAZOLE-CONTAINING LEXITROPSINS TO AGAA DNA FRAGMENT

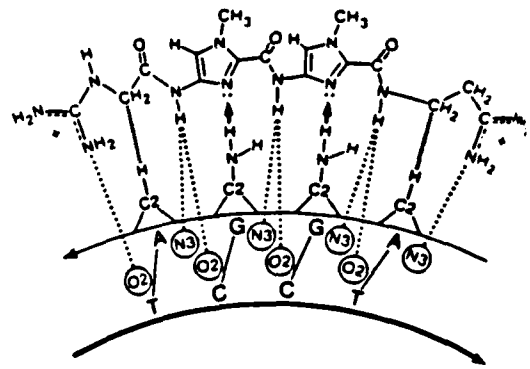


fig (23b) BINDING OF IMIDAZOLE-CONTAINING LEXITROPSINS TO AGGA DNA FRAGMENT

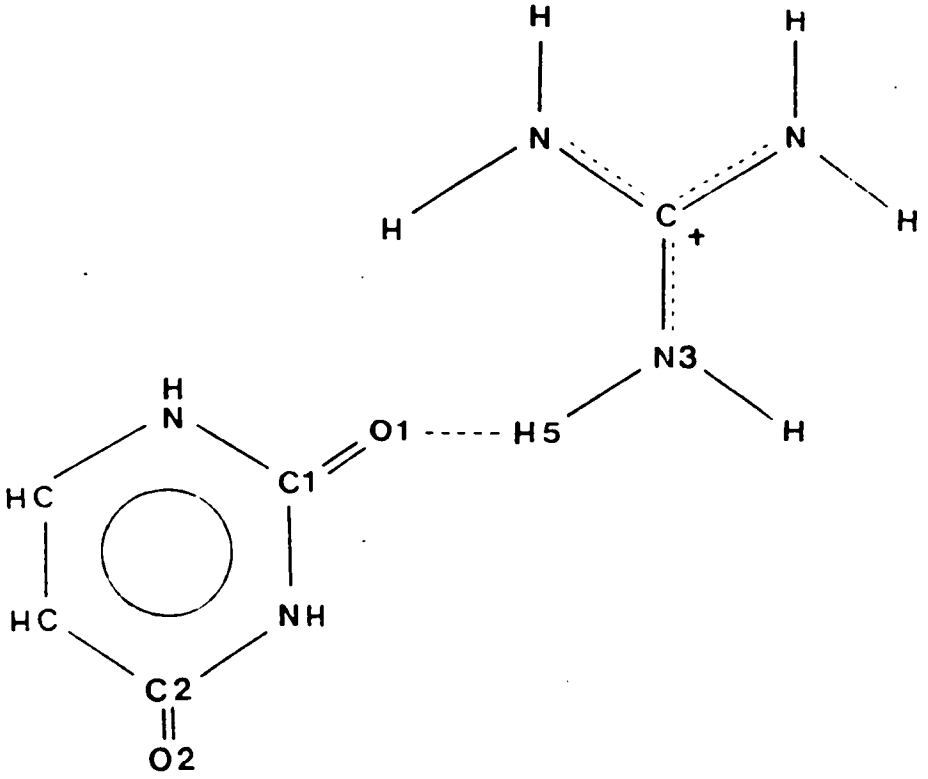
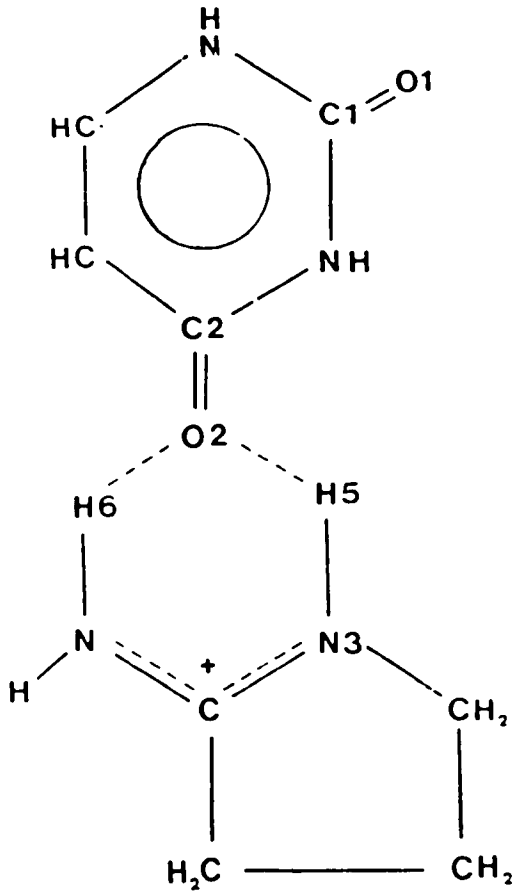
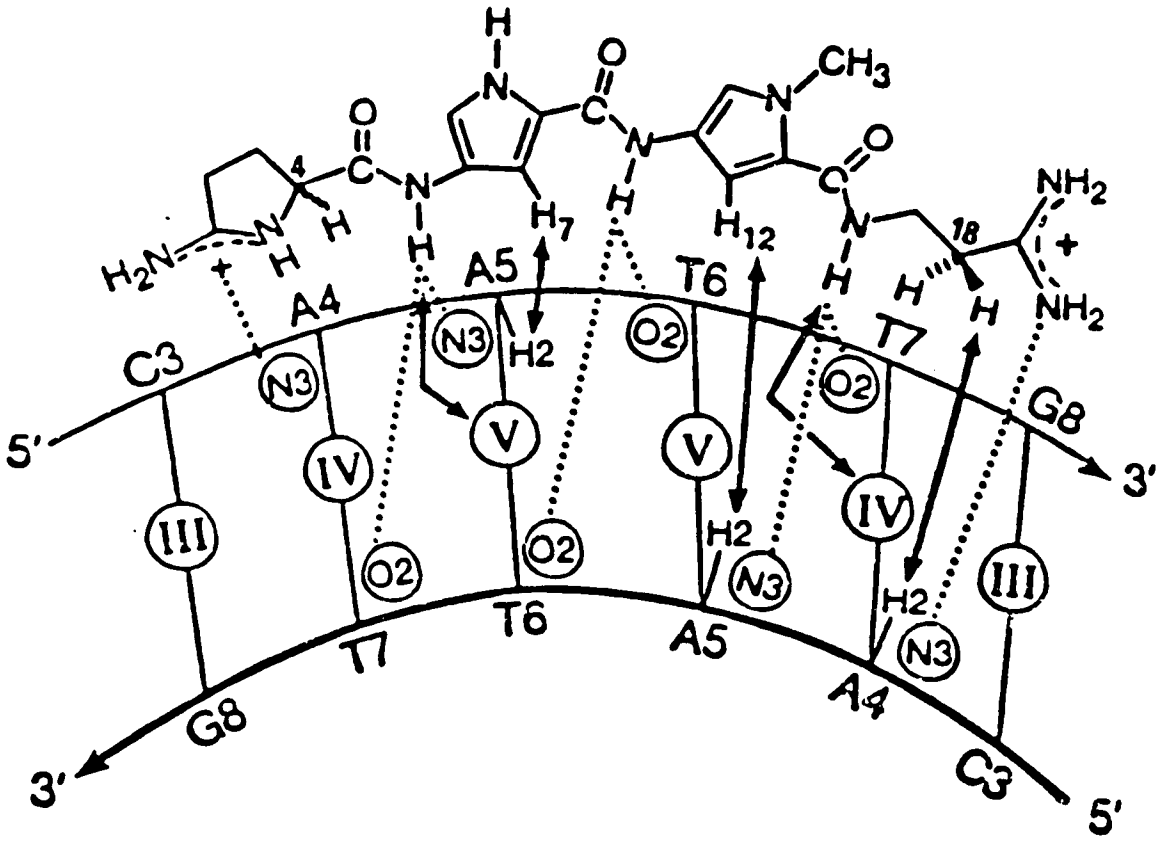


fig (24a)
COMPLEX # (1)



COMPLEX # (2)

fig (24b)

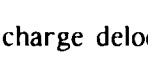


BINDING of ANTHELVENCIN to AATT of DNA

fig (25)

Ⅲ Results and Discussion

Structure of Amidinomycin :

Amidinomycin is isolated from the culture filtrate os Streptomyces flavochromogenes [33]. The product of acid hydrolysis of Amidinomycin is 3-amino-1-carboxycyclopentane and 2-amidinoethylamine. The structure of Amidinomycin is shown in fig.(5). Amidinomycin contains a cyclopentane ring with the amino group at one end and amidinium group at the other end. Under physiological conditions, it shows pKa values of 9.6 and >12.0 for its two basic sites, so it will exist in biscationic form. Also, this antibiotic has two chiral centers, C1 and C3. The four isomers of Amidinomycin are shown in fig (21). The optimized isomer is the natural-occuring one, (1R-3S). From the calculated results, Table (1), one can see that the cyclopentane ring is close to coplanar. The side chain containing the amide group is approximately 120° with respect to the ring, and the rest of the side chain is almost perpendicular to the plane of the amide group. The experimetal parameters[36] for Amidinomycin which were obtained by analyzing an x-ray crystallographic analysis of Amidinomycin sulfate, are in good agreement with the calculated values. From their experimental results, Nakamura and coworkers concluded that the cyclopentane ring has an envelope conformation, C2, C3, C4 and C5 are almost coplanar within 0.05A and C1, the flap, deviates by 0.55A from this plane in such a direction that the amino group at C1 can form an intermolecular hydrogen bond to O1 of the carbonyl group. The two bonds of the amidinium group, C9N3 and C9N4, have almost the same bond length and atoms C9, N3, N4 and C8 are also quite coplanar within 0.02 A. This is also apparent from the angles N3C9C8, N4C9C8 and N4C9C8 with the corresponding values of 118°, 119° and 122° and the amidine group may exist, as a charge delocalized cation,  Table (2) shows the ab initio net atomic charges obtained via Mulliken Population Analysis for the uncharged species. It shows that Amidinomycin features an essentially neutral cyclopentane ring and the positive charges are centered on C6 and C9 with the corresponding concentration of negative charges on the oxygen atom and nitrogen N1, N3 and N4.

Calculation of Proton affinities of heteroatomic rings, Imidazole, Oxazole and Thiozole :

Lexitropsins prefer to bind to the minor groove of DNA especially at sites consisting of AT sequences. The two reasons for this preference are the steric hinderence caused by the NH2 groups in guanine and that the positively charged drugs prefer to bind better to the more negative potential present in adenine and thymine sequences. To eventually bind GC sequences of DNA as well as AT sequences, it is desirable to synthesize drugs that have heteroatomic rings capable of accepting a proton from the NH2 group of guanine and in this way the amine group of the guanine can act as a proton donor, contributing to the binding instead of preventing it. For this purpose, the proton affinities of a series of heteroatomic rings such as imidazole, thiazole and oxazole are calculated. Also, the effect of a substituent group such as a methyl and a peptidic group, are investigated using ab initio Hartree-Fock calculations.

Table (3) shows the optimized parameters for structures 1 to 10. As can be seen from this table, both C1N1 and C1O1 increase while C2N2 is shortened upon protonation. Also, the angles C1XC2 where X = N, O, S increase which is in agreement with crystallography observations that are seen on the protonation of DNA bases[34]. The protonated structures are planar as can be seen from the fact that when the dihedral angle formed by the additional proton with the plane of the molecule are set at 120.0° and allowed to relax upon optimization, it reverted back to 120.0°. By inspecting the Hessian matrix, it is evident that all the eignvalues of the second derivative matrices are positive and therefore the optimized structures are true minima. Table (5) shows the proton affinities for these structures. Proton affinities which is the measure of the proton acceptance by the ring can indicate its ability to undergo hydrogen bonding with the amino group of guanine or cytosine. The proton affinities are defined as the difference between the energies of the protonated species and the energies of the neutral species, both at 6-31G and 6-31G* level. From examining table (5), one can see that the proton affinities decrease in the order of imidazole, oxazole and thiazole. The reason can be seen from table (6). Both nitrogen and oxygen are negatively charged and

increase their negative charge upon protonation. This causes an increase in the strength of the bond with positive hydrogen. On the other hand, sulfur, which is slightly less electronegative than carbon, is positively charged and acquires even more positive charge from the proton upon protonation. The proton affinity values obtained by using the 6-31G basis set are slightly larger than those obtained through a single point calculation at 6-31G* level, using 6-31G optimized geometry with the exception of the sulfur-containing ring. The 6-31G* value should be more reliable in this case due to the d orbitals on the sulfur. It has been shown experimentally using NMR studies that a thiazole ring-containing lexitropsin aviods the GC sequences when bound to the minor groove of DNA. This is due to the steric hindrance caused by the large size of the sulfur atom and also by its lesser ability to accept a proton from the NH_2 group of guanine. It is shown that thiazole ring-containing lexitropsin intercalates between DNA bases and will bind better to GC sequences if the nitrogen in the rings face inward, toward points of contact with the bases in the minor groove[35]. Also, it can be seen from table (5) that the presence of a methyl group which replaces the hydrogen in N2 as in structures 9, 10 increases the proton affinities of imidazole ring slightly, while a peptidic group substituted at C2 decreases it. In fact for the first case, this is due to the increased negative charge of N1 when N2 is methylated. In the second case, the proton affinities decrease due to extra stabilization caused by the conjugation with the peptidic group. In conclusion, one can expect that lexitropsin containing nitrogen heterocyclic rings can bind better to GC sequences if they contain a strong electron-donating group at N2.

Experimentally, Lown and coworkers synthesized oligopeptides that were structurally related to the antiviral anticancer antibiotic netropsin. As each pyrrole unit was successfully replaced by an imidazole moiety, the resulting oligopeptide showed an overall decrease in AT preference and an increase in GC base pairs in binding to DNA. The structures of synthetic oligopeptides are shown in fig (23a) and fig (23b). These oligopeptides have N-methylimidazole group(s) which are capable of identifying a guanine, via hydrogen bond in the binding sequence. It has been shown experimentally using DNaseI footprinting

methodology that oligopeptide corresponding to fig(23a) has a single N-methylimidazole moiety binds to a four-base pair AT-rich sequence containing a single interior GC base pair while oligopeptide corresponding to fig(23b) has two N-methylimidazole groups which binds to four base sequences having two interior GC base pair, each flanked by AT base pairs. Densitometric scans of footprinting experiments have shown that both oligopeptides of fig(23a) and fig(23b) bind to DNA with less specificity than that of netropsin and are capable of recognizing guanine in the binding sequence. However, oligopeptide corresponding to fig(23b) is less specific than oligopeptide corresponding to fig(23a).

Binding of uracil to guanidinium ion and to aminopyrrolidinium ion :

Anthelvencin and netropsin are closely related structurally. Both are tripeptides that are similar in length and have two cationic charges. Both contain amidinium groups at one end. However, anthelvencin has an aminopyrrolidinium ion at one end while netropsin has a guanidinium ion. The important role of these groups is to provide a positive potential that binds electrostatically to the negative potential present in the minor groove of DNA. In addition, they might participate in the hydrogen bonding with DNA bases, which could result in the better binding of the whole molecule to the minor groove. These two lexitropsins show a large difference in their DNA binding.

It was shown experimentally[11] that netropsin which has a guanidinum ion at the end, has a binding free energy of -53Kcal/mole while two enantiomers of anthelvencin which contain an aminopyrrolidinium ion at the end, only has -33 Kcal/mole. In order to find out how much of this difference is due to the difference in binding of the two-above mentioned groups to one of the DNA bases via hydrogen bonding, the possible geometries of two complexes (1) and (2) and their binding energies are estimated using the qunatum chemical ab initio method.

Fig (24a) shows complex (1) which is formed from the interaction of uracil with the guanidinium ion. A hydrogen bond is formed between O1 and H5 atoms of complex (1). The optimized value for N3H5O1 angle of complex (1) is found to have a value of 141.8° . An attempt was made to set this angle at 180.0°, however, at this value, the energy of complex (1) was much higher than before. Also, the O1H10 distance was decreased so that an additional hydrogen bond could be formed between uracil and guanidinium ion. But, this was not possible due to the steric hindrance caused by NH_2 group.

Fig (24b) shows the preferred geometry of complex (2) which is formed from the interaction of uracil with N-aminopyrrolidinium ion. This complex features two hydrogen bonds that are formed between O2 atom of uracil and H5 and H6 atoms of N-

aminopyrrolidinium ion. The O2H5 and O2H6 distances are very close in length with a difference of only 0.026A. The double binding is obtained as the NHO angle is decreased to a value of 138°. Also, from table (9), one can see that the binding energies for these two complexes are very close to each other. Complex (2) is slightly more stable by 1.8Kcal/mole. Indeed, NMR studies show that the three amide hydrogens of anthelvencin are expected to be involved in bifurcated hydrogen bonds to adenine-N3 and thyamine O2.[11] The sequence-reading process involves close contacts between the pyrrole hydrogen H7 and H12 and adenine hydrogen on C2 as shown in fig(25). This indicates that although anthelvencin has the structural option of directing the pyrrole N-H toward DNA, this does not not occur. This would force the loss of a hydrogen bond between the amide located on the amino terminus of the dipeptide and DNA and potentially replace it with a new hydrogen bond between the pyrrole and DNA. Apparently, the remaining hydrogen bond and van der waals contacts present in the ligand-DNA interface make this hydrogen bond exchange energetically unfavorable and it does not occur.

These results show that if there is a difference in the binding of the lexitropsins containing guanidinium and those containing aminopyrrolidinium ions, this difference could not be due to the difference in the binding of these two species to the DNA bases via hydrogen bonds. However, the value of the binding energy for both species is high and it is possible that the two moieties do bind to DNA. Accoring to Lown and coworkers, the large netropsin binding preference is entirely enthalpic in origin (-46.9kj/mole for the netropsin binding enthalpy versus -9.2 to -9.6 kj/mole for the anthelvencin binding enthalpies).

Conclusion :

Ab initio methods were used for gathering information about lexitropsins.

Amidomycin was geometry optimized and the parameters so obtained agree with experimental data.

One of the problems that was addressed is the ability of imidazole to be protonated more than oxazole and thiazole and the fact that that aminopyrrolidinium ion does not bind much better to thymine than guanidinium ion so that the contribution of this binding to the total binding of such lexitropsins as anthelvencin and netropsin to DNA is not of primary importance.

References:

1. Carter, S.K. et al, *Methods in Development of New Cancer Drugs*. pp. 13-16. Bethesda, Maryland, U.S. Dept. of Health, Education and Welfare. Public Health Service, National Institute of Health, National Cancer Institute, Washington. For sale by the Supt. of Docs., U.S. Govt. Print. of Oss., 1977.
2. Bennett L.L. Jr., Montgomery JA: *Design of anticancer agents, problems and approaches in methods in cancer research* (Busch H, ed) vol. 3, ch. 9, N.Y., Academic, 1967, p. 549
3. Hahn, F.E. *Antibiotics III. Mechanism of action of antimicrobial and antitumor agents*, 1975, pp 79-100
4. Arcamone, F.; Orezzi, *Chim. Ital.*, **1967**, *97*, 1097.
5. Zimmer, C.; Marck, C. *Nucleic Acid Res.*, **1979**, *6*, 2831.
6. Neidle, S. : Waring, *Molecular Aspects of anticancer drug action. Topic in Molecular and Structural Biology*, Macmillan : London, 1983; vol.3, pp 35-55, 127-156, 157-181
7. Lown, J.W.; Krowicki, K.; Ganapothi Bhat, U., *Biochemistry*, **1986**, *25*, 7408
8. Lown, J.W.; Kissenger, K.; Krowicki, K.; Dabrowiak J.C., *Biochemistry*, **1987**, *26*, 5590
9. Lown, J.W. *In Molecular Basis of Specificity in Nucleic Acid-Drug Interaction*, 1990, p.103
10. Zimmer, *Biochem. Biophys. Acta.*, **1983**, *741*, 15
11. Lown, J.W.; Lee, M.; Shea, R.G.; Hartley, J.A.; Kissinger, K.; Pon, R.T.; Vesnaver, G.; Breslauer, K.J.; Dadrowiak J.C. *J. Am. Chem. Soc.*, **1989**, *111*, 345
12. Lown, J.W. *J. Am. Chem. Soc.*, **1987**, *26*, 18
13. Pullman, B.; Zakrzewska, K.; Lavery, R. *J. Bimol. Struct. Dyn*, **1987**, *4*, 833
14. Pullman, B. *Quart. Rev. Biophys.*, **1981**, *14*, 289

15. Pullman, B. *J. Biochem.* **1982**, *124*, 229

16. Pullman, B.; Pullman, A. *Studia Biophysica*, **1981**, *86(2)*, 95

17. Zakrzewska, K.; Pullman, B. *J. Bimol. Struct. Dyn.*, **1988**, *5*, 1043

18. Zakrzewska, K.; Randrianarivelo M.; Pullman, B. *J. Bimol. Struct. Dyn.*, **1988**, *6(2)* 331

19. Lavery, R.; Zakrzewska, K.; Pullman, B. *J. Biomol. Struct. Dyn.*, **1986**, *3(2)*, 1155

20. Mazurek, P.; Feng, W.; Shukla, K.; Sapse, A.M.; Lown, J.W. *J. Biomol. Struct. Dyn.*, **1991**, *9(2)*, 299

21. Lown, J.W.; Sapse, A.M. *J. Biomol. Struct. Dyn.*, **1991**, *9(3)*, 599

22. Massa, L.J.; Sapse, A.M. *J. Org. Chem.*, **1980**, *45(4)*, 719

23. Pople, J.A. *J. Chem. Phys.*, **1972**, *56(2)*, 2257

24. Roothaan, C.C.J. *Rev. Mod. Phys.*, **1951**, *23*, 69

25. Boys, S.F. *Proc. Roy. Soc.*, **1950**, *A200*, 542

26. Hendre, W.J.; Stewart, R.F.; Pople, J.A. *J. Chem. Phys.*, **1z969**, *51(6)*, 2657

27. Brinkley, J.S.; Pople, J.A.; Hendre, W.J. *J. Am. Chem. Soc.*, **1980**, *102*, 939 20.

28. Ditchfield, R.; Hendre, W.J.; Pople, J.A. *J. Chem. Phys.*, **1971**, *54*, 724

29. Pople, J.A. *J. Chem. Phys.*, **1982**, *77*, 3654

30. Pople, J.A. *J. Chem Phys. Lett.*, **1972**, *66*, 217

31. Sirsch, M.J. et. al. Gaussian inc., Pittsburgh, Pa. 1990

32. Sirsch, M.J. et. al. Gaussian 86, Carnegie-Mellon Quantum Chemistry Publishing Unit, Pittsburgh, Pa. 1984

33. Nakamura, S. *Chem. Pharm. Bull.*, **1961**, *9*, 641

34. Taylor, R.; Kennard, O. *J. Mol. Struct.*, **1982**, *78*, 1

35. Lown, J.W.; Sapse, A.M. *J. Bimol. Struct. Dyn.*, **1990**, *8(1)*, 99

36. Kaneda, M.; Iitaka, Y. *J. Antibiot.*, **1980**, *XXXIII(7)*, 778