Rational recognition of nucleic acid sequences

Hirokazu Iida, Guofeng Jia and J William Lown*

Recent progress in synthetic and computational chemistry has made it possible to develop certain novel drug candidates. Drug candidates for genetic diseases, such as cancer, may also be designed on the basis of structural information obtained using X-ray analysis and NMR, as well as evidence from biological techniques applied to natural products – DNA (or RNA) complexes and conjugates. The resulting designed drug candidates exhibit promising performance based on the recognition on nucleic acid sequences.

Addresses
Department of Chemistry, University of Alberta, Edmonton, Alberta, T6G2G2, Canada
*e-mail: annabelle.wiseman@ualberta.ca

Current Opinion in Biotechnology 1999, **10**:29–33

http://biomednet.com/elecref/0958166901000029

Abbreviation
PNA peptide nucleic acid

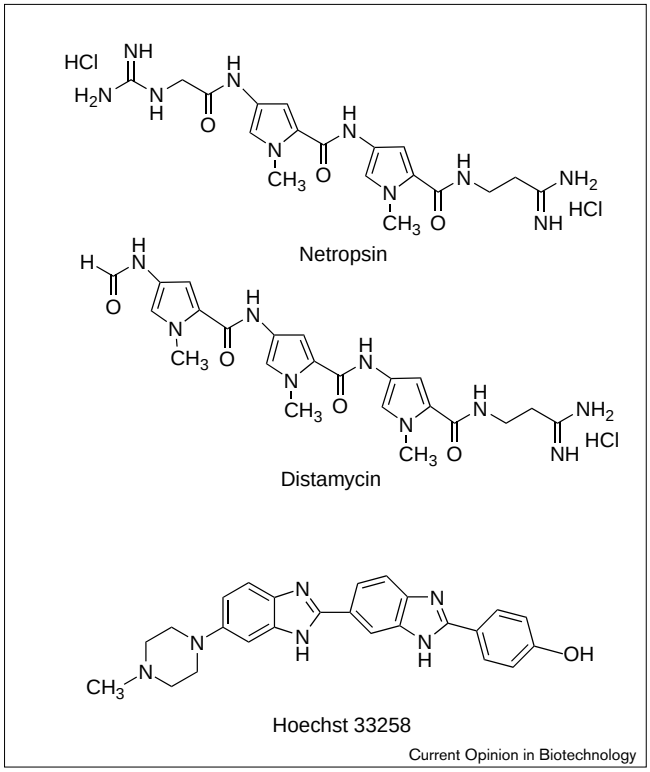
Introduction

Many drug candidates have been designed and synthesized in recent years. For example, structure-based drug design has yielded remarkable progress in the synthesis of inhibitors of certain enzymes, such as HIV type I integrase [1,2]. Because there is considerable interest currently in the development of DNA-sequence-specific or selective agents for genetic targeting, for the control of gene expression, and for applications in diagnosis or ultimately in therapy, progress in the design of drugs for these purposes has been accelerating. Until recently, the antisense strategy was the only method available for these purposes. Antisense strategy is the method that introduces a strand of DNA that has the sequence complementary to a target mRNA, thereby inhibiting translation [3]. The complementary antigene strategy has been vigorously pursued and remarkable progress has been made in the past five years. This strategy targets specific sequences of the DNA double helix directly (inhibition of transcription) [4]. In this review, we will focus on recent progress of synthetic compounds that play import roles in the rational recognition of nucleic acid sequences. These compounds may prove to be applicable to antigene strategy.

Drug–DNA complexes and conjugates elucidated by X-ray analysis

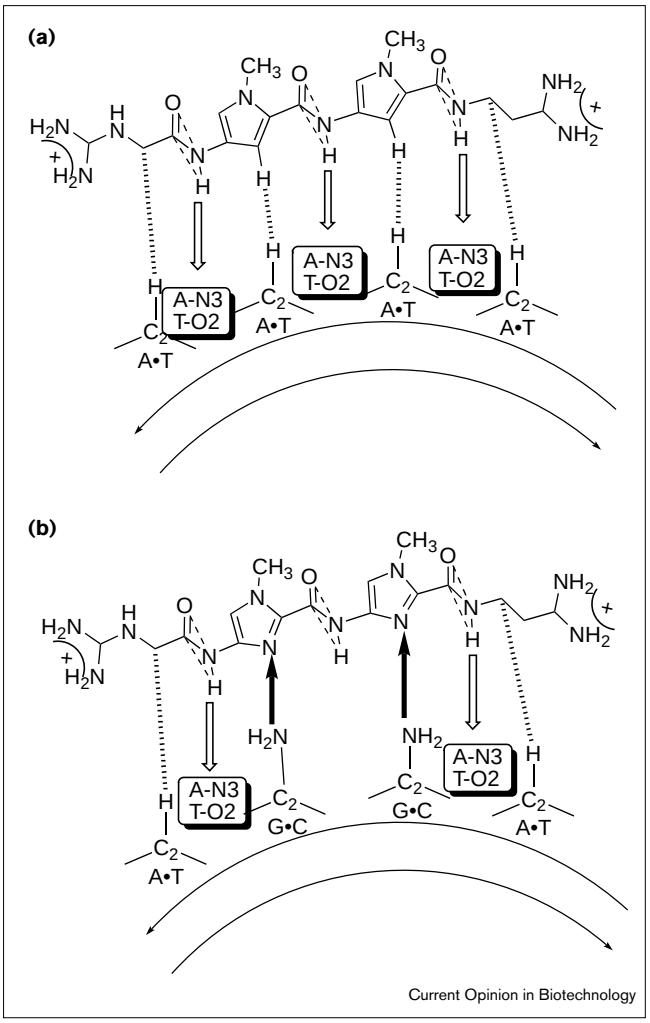
Some natural products have features that permit them to recognize and bind to specific DNA sequences. These compounds provide unique lead compounds for the antigene strategy, so many studies have been performed on natural product–DNA (or RNA) conjugates using several techniques, including X-ray analysis and NMR. Among these lead compounds, X-ray results of DNA complexed with minor groove binders, such as netropsin, distamycin and Hoechst 33258 (Figure 1), have provided considerable insight into the factors controlling how such ligands recognize nucleic acid sequences. Netropsin and distamycin are naturally occurring antibiotics which have polypyrrole amide structures and adopt crescent-shaped conformations strictly complementary to the spiral of the DNA helix. These natural products possess AT sequence selectivity. Since the first X-ray results were reported by Dickerson and co-workers in 1985 [5], several groups succeeded in the analysis of crystals of netropsin–DNA or distamycin–DNA complexes [6,7]. The selectivity for AT-rich regions was explained by Coll *et al.* [7]: firstly, AT basepairs are associated with a narrow minor groove and so afford a tighter fit to the elongated distamycin molecule; secondly, the presence of the amino group N-2 of guanine serves as a major steric block preventing the pyrrolamide chain from docking fully to the floor of the minor groove in GC-rich segments. Studies on netropsin, distamycin and related compounds have led to the concept of texitropsins, or information-reading oligopeptides.

Figure 1



Netropsin, Distamycin and Hoechst 33258.

Figure 2



Representation of the binding to DNA of **(a)** netropsin and **(b)** imidazole–netropsin. Open arrows represent existing hydrogen bonds from ligand to DNA. Closed arrows indicate new hydrogen bonds from DNA to lexitropsin. Heavy arrows are hydrogen bonds, from donor to acceptor. Dash lines mark close van der Waals nonbonded contacts between ligands and DNA.

Hoechst 33258 is a synthetic bis-benzimidazole dye agent, containing benzimidazole repeating structural motifs [8–13]. The bis-benzimidazole moiety of Hoechst 33258 shows the same specificity of AT-rich regions as netropsin, but the bulky and non-polar piperazine ring demands the wider groove that only the GC region can provide.

The 1,3-bis(4-phenylaminidinium)triazene compound berenil is an anti-trypanosomal agent and has cytotoxic and antiviral properties. The X-ray analysis of berenil–DNA conjugates suggests that berenil should bind well to AATT sites [14].

Pentamidine (1,5-bis[4-amidinophenoxy]pentane) has established antiprotozoal activities and exhibits modest antiviral activity, and also has specificity to a 5′-AATT region. X-ray analysis of complexes with DNA has suggested that the strongly cationic ends of the drug are positioned deep in the minor groove rather than close to the phosphate groups, in common with berenil, where bis(phenylamidinium) groups are present [15].

Factors contributing to the molecular recognition of DNA binders

Although the structures of Hoechst 33258 and netropsin are ostensibly quite dissimilar, examination of the molecular recognition surface components of the two ligands directed towards the DNA minor-groove receptor reveals significant similarities. X-ray analysis studies have provided valuable data regarding the role of various functional groups of these lead compounds in molecular recognition of DNA. Based on not only X-ray analysis but also NMR and biological techniques of assessing DNA–minor-groove-binder interactions, progress has been made in understanding some of the factors contributing to the molecular recognition processes: firstly, the ability of certain hydrogen bond accepting heterocylic moieties towards specific basepair recognition; secondly, the influence of ligand cationic charge in sequence-selective binding; thirdly, certain van der Waals, contacts in 3′-terminal basepair recognition; and finally, electrostatic interactions between the polyanionic DNA and the cationic compounds, which are sequence dependent [16•,17•].

Rational modification of minor groove binders to alter sequence recognition

Modification of lead compounds was explored systematically according to the factors contributing to the molecular recognition processes mentioned above. The most successful examples were the modifications of distamycin. Rational modifications have been made to this natural product to change its sequence specificity and binding characteristics. For example, replacement of one of the pyrrole rings with an imidazole, or other appropriate heterocycle, which can accept a hydrogen bond from G-2-NH_2, allows GC basepairs to be recognized [18,19]. Further studies on lexitropsins revealed some of the factors contributing to the molecular recognition process, such as structural, stereochemical, conformational, electrostatic and phasing factors [20] (Figure 2).

Other successful examples of rational structural modifications are those of Hoechst 33258. For example, replacement of one (or both) imidazole unit of the parent Hoechst 33258 by pyridoimidazole, which can accept a hydrogen bond from C-2-NH_2, allows GC basepairs to be recognized (Figure 3). These two examples clearly suggest that rational design of the compounds should make it possible to impart different basepair preferences and to design compounds that have no memory of the basepair preferences of the parent compound [16•].

Peptide nucleic acids

One of the most exciting examples for DNA sequence recognition is peptide nucleic acid (PNA). PNA is a

DNA mimic, based on a pseudo-peptide backbone to which the DNA nucleobases are attached. Whereas the conformation of PNA is similar to synthetic DNA, the behavior of PNA towards complementary DNA is completely different: sequence-specific binding to double-stranded DNA takes place to homopurine tracts via strand displacement to form an internal PNA_2/DNA triplex — a P-loop [21,22••]. PNA is capable of hybridizing to complementary single strand DNA, double strand DNA, RNA, and PNA targets, so several potential applications of PNA may be envisaged. PNA may be important as both antisense and antigene drugs. Excitingly, it was recently shown that PNA can function as a template for chemical synthesis of RNA and vice versa, thereby illustrating the principle that information transfer between PNA and RNA is possible [23]. And also, because only the nucleobases of DNA have been retained in PNA, and because the backbone of PNA is uncharged, PNA was used to separate the effects of the nucleobases and backbone of DNA in a study of minor groove binders and intercalators [24].

Hairpin compounds

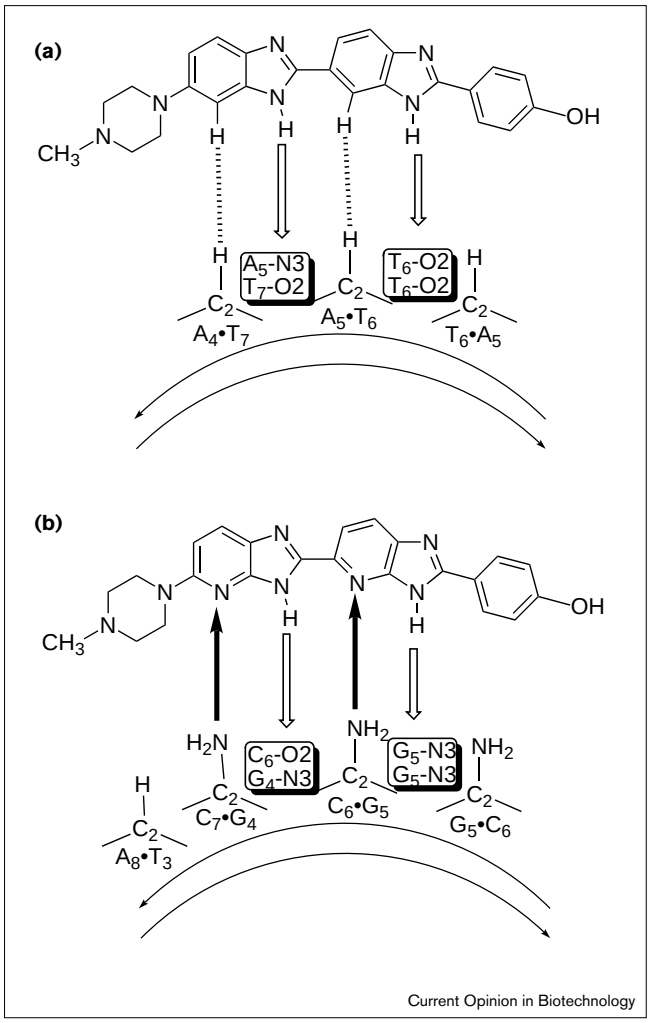
A distinctive example of structure-based drug design was reported by Baird and Dervan [25]. They developed solid-phase synthetic methods for extended lexitropsin synthesis. Using this method, they extended the lexitropsin concept to design lexitropsin dimers that show hairpin structure when these compounds bind to DNA. For example, they prepared two eight-ring pyrrole-imidazole polyamides differing in sequence by a single amino acid and showed that they bind specifically to respective six-basepair target sites which differ in sequences by a single basepair. Binding is observed at subnanomolar concentrations of ligands [26]. In a recent application, Dervan and co-workers [27,28,29••–31••] designed extended hairpins in order to bind to the transcription factor IIIA-binding site that regulates the expression of the 5S RNA gene.

Another example of the extended lexitropsin concept is the synthesis of lexitropsin dimers using *interalia* alkyl long chains, aromatic compounds, cyclopropane, and cyclobutane as linkers [17•,32•]. Certain dimers exhibited high potency as inhibitors of HIV type 1 integrase.

Further modifications of DNA minor groove binders and biological significance: conjugates with alkylating agents

Even though alkylating agents historically have been useful drug candidates for cancer treatment, their apparent lack of selectivity has hampered their use medically. In order to improve their selectivity, the synthesis of alkylating agent–minor groove binder conjugates has been examined. Lexitropsin was mainly used for this purpose and lexitropsin–alkylating agents, such as nitrogen mustards [33,34] or CC-1065 [35,36] conjugates, were synthesized [32•].

Figure 3



Representation of the binding to DNA of **(a)** Hoechst 33528 and **(b)** pyridoimidazole and analogues. Open arrows represent existing hydrogen bonds from the Hoechst structure to DNA. Closed arrows indicate new hydrogen bonds from DNA to the structurally modified ligand. Dash lines indicate van der Waals nonbonded contacts between ligands and DNA.

Application of the lexitropsin concept to CC-1065 analogues led to certain cyclopropylpyrroloindole–lexitropsin conjugates, which exhibit up to 10,000 times higher potency than CC-1065 itself against KB human cancer cells. In this case, the sequence selectivity of the lexitropsin unit may play an important role for these conjugates to exhibit high or alternative expressions of bioactivity [37].

Conclusion

Systematic modifications of lead DNA-binding compounds have made it possible to understand some of the factors contributing to the molecular recognition processes. As a result, compounds can be designed that have high selectivity to certain DNA sequences. In order to advance this field, additional versatile synthetic methods must be

developed. The approach of combinatorial synthesis is also being applied [38]. The sequence selectivity of minor-groove-binder units may play an important role for the above conjugates to exhibit high or alternative expressions of bioactivity. Further rational design, synthesis and biological evaluation by alternative molecular architecture to new gene targets will extend the possibility of this area.

Acknowledgements

The research in the authors' laboratory is supported by grants from the National Cancer Institute of Canada, the Natural Sciences and Engineering Research Council of Canada, and the University of Alberta.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Hubbard RE: **Can drugs be designed?** *Curr Opin Biotechnol* 1997, **8**:696-700.

2. Hermann T, Westhof E: **RNA as a drug target: chemical, modelling, and evolutionary tools.** *Curr Opin Biotechnol* 1998, **9**:66-73.

3. Altmann K-H, Fabbro D, Geiger T: **Antisense- and antigene-based drug design strategies in oncology.** In *Advances in DNA Sequence-Specific Agents.* London: JAI Press Inc.; 1998:227-266.

4. Chaires JB: **Drug–DNA interactions.** *Curr Opin Struct Biol* 1998, **8**:314-320.

5. Kopka ML, Yoon C, Goodsell D, Pjura P, Dickerson RE: **Binding of an antitumor drug to DNA, netropsin and C-G-C-G-A-A-T-T-^{Br}C- G-C-G.** *J Mol Biol* 1985, **183**:553-563.

6. Goodsell DS, Kopka ML, Dickerson RE: **Refinement of netropsin bound to DNA: bias and feedback in electron density map interpretation.** *Biochemistry* 1995, **34**:4983-4993.

7. Coll M, Frederick CA, Wang AH, Rich A: **A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin.** *Proc Natl Acad Sci USA* 1987, **84**:8385-8389.

8. Pjura PE, Grzeskowiak K, Dickerson RE: **Binding of Hoechst 33258 to the minor groove of B-DNA.** *J Mol Biol* 1987, **197**:257-271.

9. Teng MK, Usman N, Frederick CA, Wang AH: **The molecular structure of the complex of Hoechst 33258 and the DNA dodecamer d(CGCGAATTCGCG).** *Nucleic Acids Res* 1988, **16**:2671-2690.

10. Quintana JR, Lipanov AA, Dickerson RE: **Low-temperature crystallographic analyses of the binding of Hoechst 33258 to the double-helical DNA dodecamer C-G-C-G-A-A-T-T-C-G-C-G.** *Biochemistry* 1991, **30**:10294-10306.

11. Spink N, Brown DG, Skelly JV, Neidle S: **Sequence-dependent effects in drug–DNA interaction: the crystal structure of Hoechst 33258 bound to the d(CGCAAATTTGCG)2 duplex.** *Nucleic Acids Res* 1994, **22**:1607-1612.

12. Wood AA, Nunn CM, Czarny A, Boykin DW, Neidle S: **Variability in DNA minor groove width recognised by ligand binding: the crystal structure of a bis-benzimidazole compound bound to the DNA duplex d(CGCGAATTCGCG)2.** *Nucleic Acids Res* 1995, **23**:3678-3684.

13. Clark GR, Squire CJ, Gray EJ, Leupin W, Neidle S: **Designer DNA-binding drugs: the crystal structure of a meta-hydroxy analogue of Hoechst 33258 bound to d(CGCGAATTCGCG)2.** *Nucleic Acids Res* 1996, **24**:4882-4889.

14. Brown DG, Sanderson MR, Skelly JV, Jenkins TC, Brown T, Garman E, Stuart DI, Neidle S: **Crystal structure of a berenil-dodecanucleotide complex: the role of water in sequence-specific ligand binding.** *EMBO J* 1990, **9**:1329-1334.

15. Edwards KJ, Jenkins TC, Neidle S: **Crystal structure of a pentamidine-oligonucleotide complex: implications for DNA-binding properties.** *Biochemistry* 1992, **31**:7104-7109.

16. • Lown JW: **DNA sequence recognition altered bis-benzimidazole minor-groove binders.** In *Advances in DNA Sequence-Specific Agents.* London: JAI Press Inc.; 1998:67-95.
Rational structural modifications of Hoechst 33258, structural analysis of ligand–DNA interactions, and cellular and pharmacological effects of synthetic compounds were discussed. Rational modification of Hoechst 33258 made it possible to change selectivity from AT to GC.

17. • Bailly C: **Sequence-specific recognition and modification of double-helical DNA by minor-groove binding conjugates structurally related to netropsin and distamycin.** In *Advances in DNA Sequence-Specific Agents.* London: JAI Press Inc.; 1998:97-156.
Rational structural modifications of netropsin and distmysin, and the factors contributing to the molecular recognition of DNA by lexitropsins were reviewed.

18. Goodsell DS, Ng HL, Kopka ML, Lown JW, Dickerson RE: **Structure of a dicationic monoimidazole lexitropsin bound to DNA.** *Biochemistry* 1995, **34**:16654-16661.

19. Kopka ML, Goodsell DS, Han GW, Chiu TK, Lown JW, Dickerson RE: **Defining GC-specificity in the minor groove: side-by-side binding of the di-imidazole lexitropsin to C-A-T-G-G-C-C-A-T-G.** *Structure* 1997, **5**:1033-1046.

20. Lown JW: **Synthesis of sequence-specific agents: lexitropsins.** In *Molecular Aspects of Anticancer Drug–DNA Interactions.* Boca Raton: CRC Press; 1993:322-355.

21. Dueholm KL, Nielsen PE: **Chemistry, properties and applications of PNA (peptide nucleic acid).** *New J Chem* 1997, **21**:19-31.

22. •• Nielsen PE: **Sequence-specific recognition of double-stranded DNA by peptide nucleic acids.** In *Advances in DNA Sequence-Specific Agents.* London: JAI Press Inc.; 1998:267-278.
The concept and biological significance of PNAs were described. PNA's unique behavior in forming PNA_2–DNA triplexes – a P loop with high sequence specificity – is explained in detail.

23. Böhler C, Nielsen PE, Orgel LE: **Template switching between PNA and RNA oligonucleotides.** *Nature* 1995, **376**:578-581.

24. Wittung P, Kim SK, Buchardt O, Nielsen P, Nordèn B: **Interactions of DNA binding ligands with PNA–DNA hybrid.** *Nucleic Acids Res* 1994, **22**:5371-5377.

25. Baird EE, Dervan PB: **Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids.** *J Am Chem Soc* 1996, **118**:6141-6146.

26. Trauger JW, Baird EE, Dervan PB: **Extended hairpin polyamide motif for sequence-specific recognition in the minor groove of DNA.** *Chem Biol* 1996, **3**:369-377.

27. Gottesfeld JM, Neely L, Trauger JW, Baird EE, Dervan PB: **Regulation of gene expression by small molecules.** *Nature* 1997, **387**:202-205.

28. Pilch DS, Poklar N, Gelfand CA, Law SM, Breslauer KJ, Baird EE, Dervan PB: **Binding of a hairpin polyamide in the minor groove of DNA: sequence-specific enthalpic discrimination.** *Proc Natl Acad Sci USA* 1996, **93**:8306-8311.

29. •• Swalley SE, Baird EE, Dervan PB: **Discrimination of 5′-GGGG-3′, 5′ GCGC-3′, and 5′-GGCC-3′ sequences in the minor groove of DNA by eight-ring hairpin polyamides.** *J Am Chem Soc* 1997, **119**:6953-6961.
Eight-ring hairpin polyamides containing pyrrole and imidazole amino acids were designed for recognition of six basepair DNA sequences containing four contiguous GC basepairs.

30. •• White S, Baird EE, Dervan PB: **Orientation preferences of pyrrole-imidazole polyamides in the minor groove of DNA.** *J Am Chem Soc* 1997, **119**:8756-8765.
Equilibrium association constants (Ka) were determined for a series of six ring hairpin polyamides which differ with respect to substitution at the amino and carboxyl-termini. A positive charge at the carboxyl-terminus and no substitution of an amino-terminal imidazole was preferred for optimal orientation and sequence specficity.

31. •• Turner JM, Baird EE, Dervan PB: **Recognition of seven base pair sequences in the minor groove of DNA by ten-ring pyrrole-imidazole polyamide hairpins.** *J Am Chem Soc* 1997, **119**:7636-7644.
Ten-ring hairpin polyamides containing pyrrole and imidazole amino acids were designed for recognition of seven basepair sequences in the minor groove of DNA.

32. Sondhi SM, Reddy BSP, Lown JW: **Lexitropsin conjugates: action on DNA targets.** *Curr Med Chem* 1997, **4**:313-358.
• A variety of lexitropsin conjugates with a DNA intercalator, a DNA alkylating agent, or a DNA cleaving agent were reviewed. Biological results of selected compounds confirm the usefulness of lexitropsin units.

33. Arcamone FM, Animati F, Barbieri B, Conliacchi E, D'Alessio R, Geroni C, Giuliani FC, Lazzari E, Menozzi M, Mongelli N: **Synthesis, DNA-binding properties, and antitumor activity of novel distamycin derivatives.** *J Med Chem* 1989, **32**:774-778.

34. Ciucci A, Manzini S, Lombardi P, Arcamone F: **Backbone and benzoyl mustard carrying moiety modifies DNA interactions of distamycin analogues.** *Nucleic Acids Res* 1996, **24**:311-315.

35. Wang Y, Gupta R, Huang L, Lown JW: **CC-1065 functional analogues possessing different electron-withdrawing substituents and leaving groups: synthesis, kinetics, and sequence specificity of reaction with DNA and biological evaluation.** *J Med Chem* 1993, **36**:4172-4182.

36. Wang Y, Gupta R, Huang L, Luo W, Lown JW: **Design, synthesis, cytotoxic properties and preliminary DNA sequencing evaluation of CPI-N-methylpyrrole hybrids. Enhancing effect of a trans double bond linker and role of the terminal amide functionality on cytotoxic potency.** *Anticancer Drug Des* 1996, **11**:15-34.

37. Iida H, Lown JW: **Design and synthesis of cyclopropylpyrroloindole (CPI) – lexitropsin conjugates and related structures.** *Recent Res Dev Synth Organic Chem* 1998, **1**:17-46.

38. Myers PL: **Will combinatorial chemistry deliver real medicines?** *Curr Opin Biotechnol* 1997, **8**:701-707.