TOPICS IN CURRENT CHEMISTRY 253

Volume Editors M.J. Waring · J.B. Chaires

DNA Binders and Related Subjects

Springer

253
Topics in Current Chemistry

Topics in Current Chemistry

Recently Published and Forthcoming Volumes

Anion Sensing
Volume Editor: Stibor, I.
Vol. 255, 2005

Organic Solid State Reactions
Volume Editor: Toda, E.
Vol. 254, 2005

DNA Binders and Related Subjects
Volume Editors: Waring, M.J., Chaires, J.B.
Vol. 253, 2005

Contrast Agents III
Volume Editor: Krause, W.
Vol. 252, 2005

Chalcogenocarboxylic Acid Derivatives
Volume Editor: Kato, S.
Vol. 251, 2005

New Aspects in Phosphorus Chemistry V
Volume Editor: Majoral, J.-P.
Vol. 250, 2005

Templates in Chemistry II
Volume Editors: Schalley, C.A., Vögtle, F., Dötz, K.H.
Vol. 249, 2005

Templates in Chemistry I
Volume Editors: Schalley, C.A., Vögtle, F., Dötz, K.H.
Vol. 248, 2005

Collagen
Volume Editors: Brinckmann, J., Notbohm, H., Müller, P.K.
Vol. 247, 2005

New Techniques in Solid-State NMR
Volume Editor: Klinowski, J.
Vol. 246, 2005

Functional Molecular Nanostructures
Volume Editor: Schlüter, A.D.
Vol. 245, 2005

Natural Product Synthesis II
Volume Editor: Mulzer, J.
Vol. 244, 2005

Natural Product Synthesis I
Volume Editor: Mulzer, J.
Vol. 243, 2005

Immobilized Catalysts
Volume Editor: Kirschning, A.
Vol. 242, 2004

Transition Metal and Rare Earth Compounds III
Volume Editor: Yersin, H.
Vol. 241, 2004

The Chemistry of Pheromones and Other Semiochemicals II
Volume Editor: Schulz, S.
Vol. 240, 2005

The Chemistry of Pheromones and Other Semiochemicals I
Volume Editor: Schulz, S.
Vol. 239, 2004

Orotidine Monophosphate Decarboxylase
Volume Editors: Lee, J.K., Tantillo, D.J.
Vol. 238, 2004

Long-Range Charge Transfer in DNA II
Volume Editor: Schuster, G.B.
Vol. 237, 2004

Long-Range Charge Transfer in DNA I
Volume Editor: Schuster, G.B.
Vol. 236, 2004

Spin Crossover in Transition Metal Compounds III
Volume Editors: Gütlich, P., Goodwin, H.A.
Vol. 235, 2004

Spin Crossover in Transition Metal Compounds II
Volume Editors: Gütlich, P., Goodwin, H.A.
Vol. 234, 2004

Spin Crossover in Transition Metal Compounds I
Volume Editors: Gütlich, P., Goodwin, H.A.
Vol. 233, 2004

DNA Binders and Related Subjects

Volume Editors: Michael J. Waring · Jonathan B. Chaires

With contributions by
B. A. Armitage · D. P. Arya · C. Bailly · C. M. Barbieri · J. B. Chaires
P. B. Dervan · N. Dias · B. S. Edelson · C. Escudé · E. J. Fechter
M. Kaul · A. Lansiaux · D. S. Pilch · A. T. Poulin-Kerstien · J.-S. Sun
H. Vezin · L. D. Williams

Springer

The series *Topics in Current Chemistry* presents critical reviews of the present and future trends in modern chemical research. The scope of coverage includes all areas of chemical science including the interfaces with related disciplines such as biology, medicine and materials science. The goal of each thematic volume is to give the nonspecialist reader, whether at the university or in industry, a comprehensive overview of an area where new insights are emerging that are of interest to a larger scientific audience.

As a rule, contributions are specially commissioned. The editors and publishers will, however, always be pleased to receive suggestions and supplementary information. Papers are accepted for *Topics in Current Chemistry* in English.

In references *Topics in Current Chemistry* is abbreviated Top Curr Chem and is cited as a journal.

Visit the TCC content at springerlink.com

Library of Congress Control Number: 2004109782

ISSN 0340-1022
ISBN 3-540-22835-7 **Springer Berlin Heidelberg New York**
DOI 10.1007/b99428

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in The Netherlands

Cover design: KünkelLopka, Heidelberg/design & production GmbH, Heidelberg
Typesetting: Fotosatz-Service Köhler GmbH, Würzburg

Printed on acid-free paper 02/3141 – 5 4 3 2 1 0

Volume Editors

Professor
M. J. Waring
Department of Pharmacology
University of Cambridge
Tennis Court Road
Cambridge CB2 1PD, UK
mjw11@cam.ac.uk

Dr. Jonathan B. Chaires
James Graham Brown Cancer Center
Department of Medicine
Health Sciences Center
University of Louisville
529 South Jackson St.
Louisville, KY 40202, USA
j.chaires@louisville.edu

Editorial Board

Topics in Current Chemistry also Available Electronically

For all customers who have a standing order to Topics in Current Chemistry, we offer the electronic version via SpringerLink free of charge. Please contact your librarian who can receive a password for free access to the full articles by registration at:

springerlink.com

If you do not have a subscription, you can still view the tables of contents of the volumes and the abstract of each article by going to the SpringerLink Homepage, clicking on "Browse by Online Libraries", then "Chemical Sciences", and finally choose Topics in Current Chemistry.

You will find information about the

– Editorial Board
– Aims and Scope
– Instructions for Authors
– Sample Contribution

at springeronline.com using the search function.

Preface

A little over forty years ago, the study of drug interactions with DNA at the molecular level was born in Cambridge with the publication of the intercalation model for binding of mutagenic aminoacridines to double-helical DNA. Prior to that, there had been pioneering studies on drug-nucleic acid interaction which set the stage for the intercalation concept as well as other modes of binding, but it was Leonard Lerman's insight into the likely molecular nature of proflavine–DNA interaction that heralded the age of detailed studies on drug binding to nucleic acids leading to precise understanding of molecular contacts. Eventually the field of molecular recognition, meaning the mechanisms underlying specific binding of small molecules to macromolecular receptors, acquired an identity of its own and the work on drug binding to DNA became increasingly important because it furnished striking examples of molecular recognition in action.

As always, advances in technology and the application of new methods to outstanding problems have played a major part in the development of ideas about drug-nucleic acid recognition. The field has undergone an explosive diversification as wider and wider problems became accessible to study using the new ideas and techniques. This volume reflects that diversification by offering accounts of selected areas that illustrate recent advances in the study of ligand–nucleic acid binding over disparate areas of the subject. There are chapters dealing specifically with the invention and application of new methodology, and a particularly thoughtful essay on the interpretation of X-ray diffraction data which may not be as straightforward as is often imagined. Other chapters illustrate the diversity and complexity of drug-DNA binding from several perspectives, referring to particular groups of related compounds or the potential attractions of the less-preferred DNA major groove as a target for nucleotide sequence recognition by ligands. An account of specific polyamide binding to the minor groove provides a timely insight into one of the most promising avenues for developing useful gene-directed drugs by rational design, contrasting nicely with another account of the remarkable success of searches for naturally-occurring drugs acting at the level of DNA. Finally, two related chapters deal with the interaction between important members of the aminoglycoside group of antibiotics and ribosomal RNA, a tremendously important target for antibacterial chemotherapy, illus-

trating exciting new developments in our understanding of how the catalytic workings of RNA can be highly specifically modified by drug binding – a topic undreamed of when the first nucleic acids (always DNA) were originally identified as bona fide drug targets.

The selection of subjects presented here has stemmed from personal contacts, reading of the literature, friendships, perseverance and perhaps a measure of inspiration. It does not amount to a balanced compilation of hot topics of today. It does, however, attempt to offer food for thought as well as some useful glimpses of science at work in a field which every one agrees holds enormous promise for the future of biology and medicine firmly rooted in good chemistry. It is in that spirit that the editors offer it in response to the kind invitation of the publishers to make a contribution to their excellent series of *Topics in Current Chemistry*.

Cambridge and Louisville, October 2004 M. J. Waring, J. B. Chaires

Contents

Contents of Volume 247
Collagen

Volume Editor: J. Brinckmann, H. Notbohm, P. K. Müller
ISBN 3-540-23272-9

Top Curr Chem (2005) 253: 1–31
DOI 10.1007/b100440

Regulation of Gene Expression by Synthetic DNA-Binding Ligands

Peter B. Dervan (✉) · Adam T. Poulin-Kerstien · Eric J. Fechter · Benjamin S. Edelson

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA
dervan@caltech.edu

Abstract During the past 20 years, polyamides have evolved from the natural product distamycin to a new class of programmable heterocyclic oligomers that bind a broad repertoire of DNA sequences with high affinity and specificity. This chapter details recent advances in this field of research, focusing on molecular recognition of DNA, and biological applications such as modulating gene expression by small molecules. Work presented here represents efforts towards the modulation of specific cellular function by small molecules in an addressable fashion within the context of live cells.

Keywords Polyamide · DNA · Gene regulation

1
Polyamides: Sequence-Specific Recognition of DNA

1.1
From Natural Roots: Distamycin

The natural product distamycin contains three *N*-methylpyrrole (Py) amino acids and binds in the minor groove of DNA at A,T tracts four to five base pairs (bp) in size [1, 2]. Distamycin inhibits DNA-dependent processes, including transcription, and has antibacterial [3], antimalarial [4], antifungal [5], and antiviral activities [6], but is of limited use because of toxicity [7].

Structural studies of distamycin–DNA complexes revealed that the crescent-shaped molecule binds A,T tracts in both 2:1 and 1:1 ligand:DNA stoichiometries (Fig. 1) [8–10]. These structures revealed key ligand–DNA interactions, such as a series of hydrogen bonds between pyrrolecarboxamides and the edges of the nucleobases on the adjacent DNA strand. These studies revealed that both the shape complementarity and the specific hydrogen-bonding profile of distamycin account for its affinity and specificity toward B-form DNA. Over the nearly 20 years since the original structural work, analogues have been created and characterized that bind a large number of different DNA sequences in a programmable fashion [11]. We now have a set of five- (and six-membered) heterocyclic amino acids that can be combined as modular,

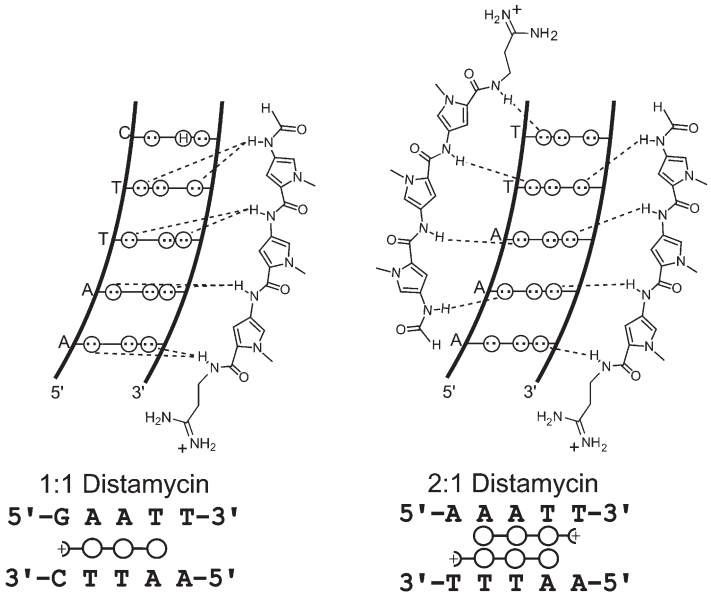


Fig. 1 Schematic representation of the two DNA-binding modes of distamycin [8–10]. Hydrogen bonds are shown as *dashed lines*. *Circles with dots* represent lone pairs of N3 of purines and O2 of pyrimidines. *Open circles* below represent pyrrole amino acid units

antiparallel ring pairs in the minor groove of DNA to recognize predetermined sequences of DNA, with affinities and specificities comparable to DNA-binding proteins [12]. Presented here are recent advances in the field of DNA recognition by synthetic minor groove-binding polyamides.

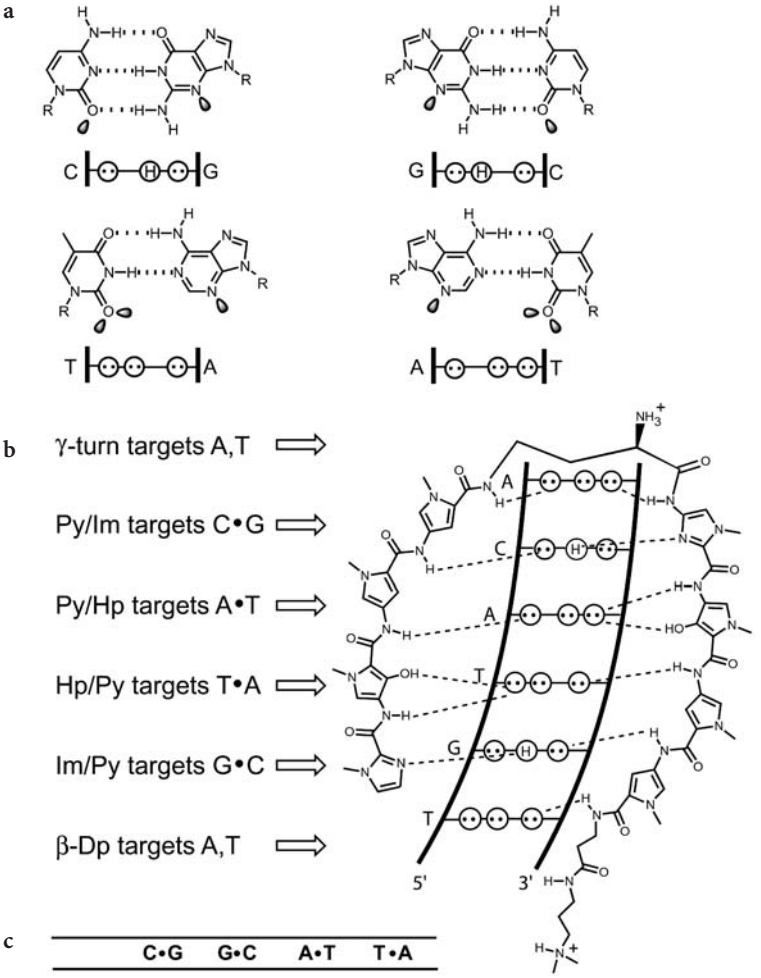
1.2 Pairing Rules

In a formal sense, the four Watson–Crick base pairs can be differentiated on the minor groove floor by the specific positions of hydrogen bond donors and acceptors, as well as by differences in molecular shape and electronic potential surfaces (Fig. 2a) [12]. The exocylic amine of guanine presents an unsymmetrical hydrogen bond donor "bump" on the minor groove edge of a G•C base pair. A key study in the early 1990s demonstrated that the *N*-methylimidazole (Im)-containing polyamide ImPyPy bound to the five bp sequence 5′-WGWCW-3′ (where W=A or T) [13]. This result was rationalized in terms of the formation of a 2:1 polyamide–DNA complex in which an antiparallel ring pairing of Im stacked against Py could specifically distinguish a G•C from the other three base pairs (Figs. 2 and 3).

The Im/Py pair has been explored by extensive studies, including analyses of binding in hundreds of different sequence contexts. Crystal structures confirmed the existence of a hydrogen bond between the Im nitrogen and the exocyclic NH_2 of guanine when the Im/Py pair binds opposite the G•C base pair [14]. The preference for a linear hydrogen bond, coupled with the unfavorable angle to an Im over the cytosine side of the base pair, provides a basis for the ability of an Im/Py pair to discriminate G•C from C•G (Fig. 3). Thermodynamic investigations dissected binding free energies into enthalpic and entropic contributions, revealing that the sequence selectivity of the Im/Py pair is driven by a favorable enthalpic contribution [15].

Discrimination of A•T from T•A base pairs was achieved 6 years later. The A•T base pair appears fairly symmetrical, with both adenine and thymine presenting a hydrogen bond acceptor to the floor of the minor groove (Fig. 2a). However, closer inspection reveals that a small asymmetric cleft is formed between the thymine O2 and adenine C2. Furthermore, the N3 of adenine presents only one lone pair while the O2 of thymine presents two lone pairs capable of hydrogen bonding. Informed by high resolution crystallographic data from a polyamide–DNA complex, *N*-methyl-3-hydroxypyrrole (Hp) was designed, and subsequently proved to be a thymine-selective recognition element when paired across from Py (Figs. 2 and 3) [16].

Crystal structures of two different Hp-containing polyamides, as their 2:1 complexes with DNA, have been determined at high resolution [17, 18]. The specificity of an Hp/Py pair was shown to arise from a combination of specific hydrogen bonds between the hydroxyl and the thymine O2, along with shape-selective recognition of the asymmetric cleft (Fig. 3). Hp polyamides bind with lower affinities than their Py counterparts [16], and a recent computational



c

	C•G	G•C	A•T	T•A
Py/Im	+	−	−	−
Im/Py	−	+	−	−
Py/Hp	−	−	+	−
Hp/Py	−	−	−	+

Fig. 2a–c **a** Structures of the four Watson–Crick base pairs. The R group represents the sugar–phosphate backbone of DNA, and shaded orbitals represent electron lone pairs projecting into the minor groove. *Circles with dots* represent lone pairs of N3 of purines and O2 of pyrimidines. *Circles with an H* represent the exocyclic 2-amino group of guanine. **b** Schematic model for the hairpin polyamide ImHpPyPy-(R)^{H2N}*γ*-ImHpPyPy-*β*-Dp bound to its match site 5′-TGTACA-3′ as determined by the pairing rules for recognition of all four Watson–Crick base pairs of DNA in the minor groove by polyamides. Putative hydrogen bonds are indicated as *dashed lines*. **c** Table indicating the code for minor groove recognition by polyamides. *Plus* and *minus* signs indicate favored and disfavored interactions, respectively

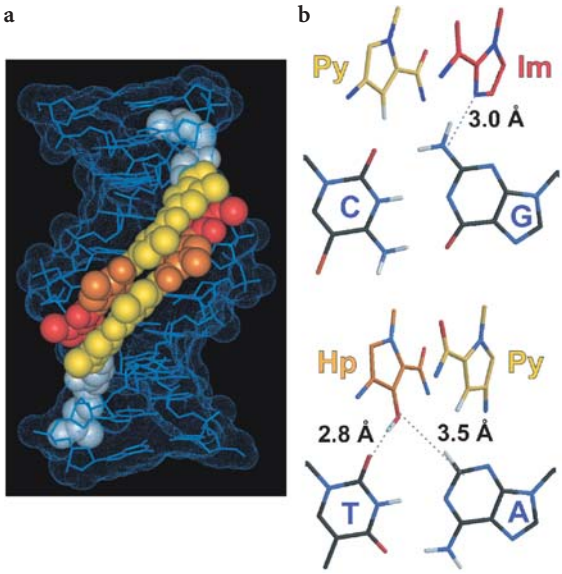


Fig. 3a,b **a** X-ray crystal structure of ImHpPyPy-*β*-Dp (Dp=dimethylaminopropylamine) bound in a 2:1 complex with its target DNA site, 5′-AGTACT-3′ (PDB code 407D) [16]. Im residues are *darkly shaded/red*, Hp residues are *lightly shaded/orange*, Py residues are *white/yellow*. **b** Detail of the Py/Im pair interacting with the C•G base pair (*top*) and of the Hp/Py pair interacting with the T•A base pair. *Dashed lines* indicate interatomic distances between carbon, nitrogen, and oxygen atoms emphasizing the close interactions responsible for specificity

study argues that desolvation of the hydroxyl group upon insertion into the minor groove accounts for the energetic penalty [19]. Together, three rings – Py, Im, and Hp – can be combined as unsymmetrical pairs to recognize specifically each of the four Watson–Crick base pairs: Im/Py is specific for G•C and Hp/Py for T•A (Fig. 2b). These interactions can be conveniently described as pairing rules (Fig. 2c). The pairing rules should be considered as guidelines only. Antiparallel polyamide dimers bind B-form DNA, and there are limitations regarding sequences targeted due to the sequence-dependent microstructure of DNA.

2
Expanding and Refining DNA Recognition

2.1
Improving Affinity and Specificity

Covalent linkage of the two antiparallel polyamide strands results in molecules with increased affinity and specificity (Fig. 4a,b). Currently, the “standard” motif is the eight-ring hairpin oligomer, in which a *γ*-aminobutyric acid linker

(γ-turn) connects the carboxylic terminus of one polyamide to the amino terminus of its antiparallel partner. Compared to the unlinked homodimers, hairpin oligomers display ~100-fold higher affinity, with the γ-turn demonstrating selectivity for A,T over G,C base pairs (Fig. 2b), presumably due to a steric clash between the aliphatic turn unit and the exocyclic amine of guanine [20]. Eight-ring hairpins, which bind six bp, were shown to have affinities and specificities similar to DNA-binding proteins (i.e., K_d<1 nM) [21]. NMR studies confirmed that the γ-turn locks the register of the ring pairings, preventing the ambiguity of slipped dimers [22]. Hairpin polyamides retain the orientation preferences of unlinked antiparallel polyamides, aligning N→C with respect to the 5′→3′ direction of the adjacent DNA strand [23]. The β-alanine-dimethylaminopropyl amine (β-Ala, Dp) tail substituent at the C-terminus of many hairpin polyamides is an A,T-specific element (Fig. 2b), again presumably due

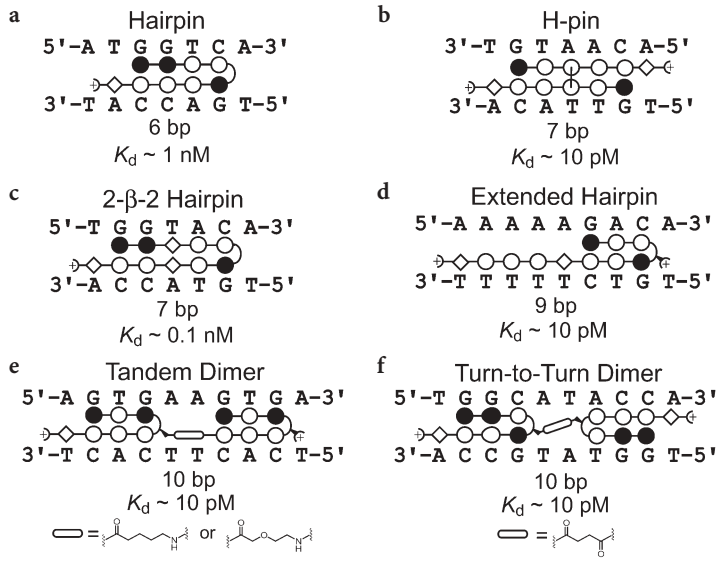


Fig. 4a–f Polyamide–DNA binding motifs with approximate dissociation constants (K_d). **a** Hairpin:standard motif targeting six bp with high affinity and sequence specificity [20]. **b** H-pin: covalent linkage of two polyamide strands is achieved at a position that is sequence-neutral [39, 40]. **c** β-Ala-containing polyamides are able to bind longer sequences because the flexible aliphatic residue relaxes the curvature of the polyamide [31]. **d** Extended hairpin: incorporates both 1:1 and 2:1 binding motifs to target longer sequences of DNA. **e**, **f** Hairpin dimers: linked either turn-to-tail (**e**) or turn-to-turn (**f**), these molecules are able to target long sequences of DNA with moderate specificity and high affinity [33–35]. Optimized linkers for each motif are shown below. The *black* and *open circles* represent Im and Py rings, respectively; *diamonds* represent β-Ala residues; and *plus signs next to diamonds* represent Dp residues. A *curved line connecting the sides of two circles* represents the γ-aminobutyric acid turn, and a *curved line with a wedge and a plus sign* represents the chiral $(R)^{H2N}$γ-turn. For the H-pin, *curved lines connecting the centers of two circles* represent alkyl linkers attached to the *N*-methyl positions of the aromatic rings

to an unfavorable steric interaction between the aliphatic chain and the exocyclic amine of guanine [24]. The tail is thought to play a role in the orientation preference of polyamides [25].

For some hairpins, however, "reversed binding" (a C→N alignment of the polyamide with respect to the 5′→3′ direction of the adjacent DNA strand) has been observed as the preferred orientation [23]. By introducing an amino substituent at the α position of the γ-turn, reversed binding is disfavored because of a steric clash between the amino substituent and the floor of the minor groove [26]. Not only does the chiral turn maintain the specificity of hairpins, it increases the overall affinity, either by the addition of a positive charge, which interacts favorably with the negatively charged backbone or minor groove floor of the DNA polymer, or by sterically reducing the conformational freedom of the polyamide in solution [26].

In other cases, polyamides containing aliphatic residues such as the γ-turn or β-Ala have been shown to favor an extended 1:1 binding mode. Depending on the stoichiometry, the *same* polyamide may bind *different* sequences (Fig. 4) [27]. For example, a hairpin containing internal β-Alas might bind in two different confirmations, a "folded" versus an "extended" conformation. An amino substituent on the γ-turn also disfavors the extended binding mode, and serves to lock the polyamide into the hairpin conformation [28]. A larger, *N*-acetyl group increases this effect, with an eight-ring, *N*-acetyl-substituted polyamide favoring the hairpin conformation over extended binding by >25,000-fold. Substitution of β-Ala for Py has also been shown to influence the tendency of a polyamide to bind in a hairpin conformation [29]. In each case, it is presumably a steric fit between the substituent and the wall of the minor groove that drives the equilibrium toward the hairpin conformation.

2.2 Binding Site Size

While the standard eight ring hairpin recognizes DNA with high affinity and specificity, it targets only six bp. For biological applications, binding site size may be critical because longer sequences would be expected to occur less frequently in a gigabase-size genome. Early attempts to increase the targeted site size by simply extending the number of aromatic rings resulted in polyamides with decreased affinity [30]. Crystal structures of polyamide–DNA complexes have shown that the polyamide rise-per-residue matches the pitch of the B-DNA helix, that is, the spacing of the pyrrole and imidazole rings matches the spacing of the DNA base pairs [14, 17, 18]. However, the inherent crescent-shaped curvature of polyamides is slightly tighter than the curvature of the minor groove. Beyond five contiguous rings, the shape of a polyamide is no longer complementary to DNA, and the resulting loss of specific contacts accounts for the observed loss in affinity and specificity [14].

The hypercurvature of polyamides can be overcome by the introduction of the flexible β-Ala residue. When introduced as a Py replacement [31], its

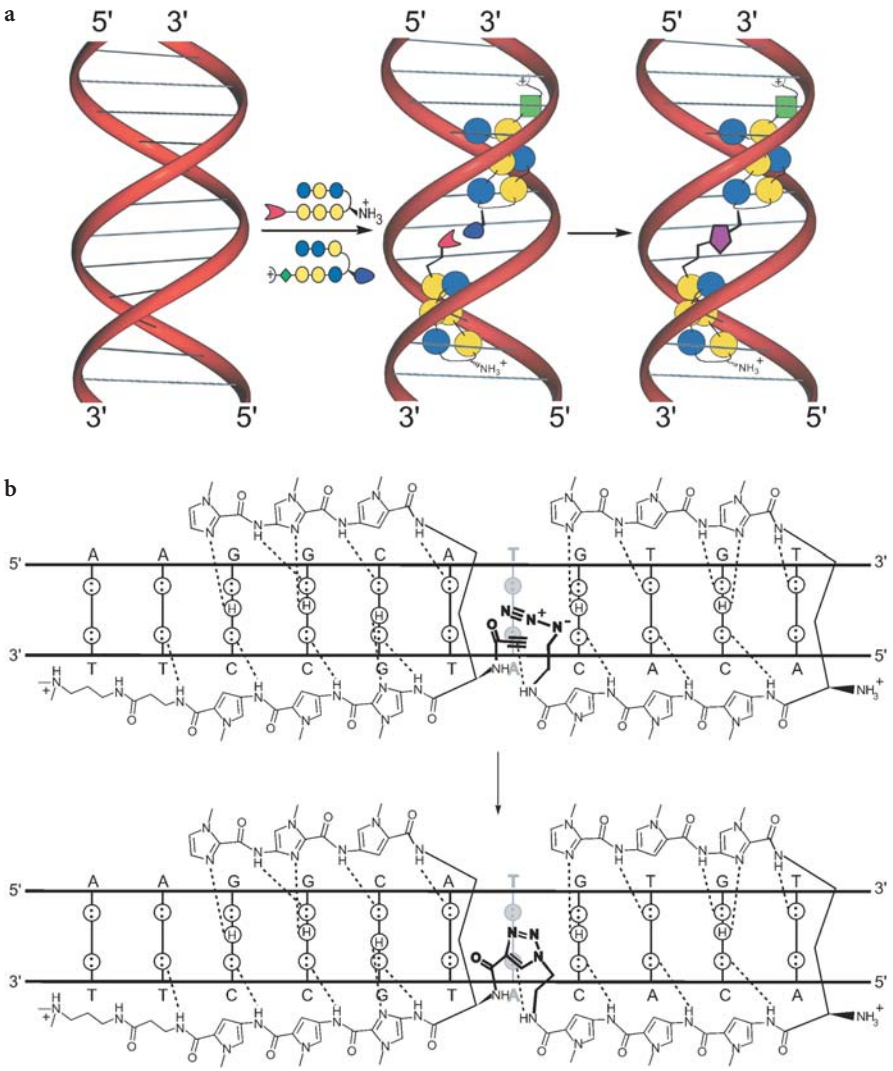


Fig. 5a,b **a** Schematic model of the DNA-templated dimerization of two functionalized hairpin polyamides. **b** Model of the specific interactions between azide- and alkyne-functionalized hairpins to form a 1,4-triazole-linked turn-to-tail tandem hairpin [38]. Symbols are defined in Figs. 2 and 4

flexibility relaxes the curvature of the polyamide, restoring complementarity to DNA. β-Ala-containing polyamides have been used to target up to 16 bp with high affinity and specificity (Fig. 4c,d) [32].

Another strategy to target longer sequences of DNA is to covalently link two hairpin polyamides with a flexible linker. These dimeric polyamides, linked both turn-to-tail and turn-to-turn have been shown to bind longer sequences

with high affinity (Fig. 4e,f) [33–35]. Both turn-to-turn and turn-to-tail dimers with optimized linkers showed good selectivity for a ten bp site (over 11 and 12 bp sites) but exhibited poor specificity (expressed in terms of affinity for match over single base pair mismatch sites). Nonetheless, an impressive application of the tandem motif was demonstrated by Laemmli and coworkers, who used tandem-hairpins to stain insect or vertebrate telomeres, $(TTAGG)_n$ or $(TTAGGG)_n$ repeats, respectively, with remarkable selectivity in fixed cells and chromosome spreads [36].

A kinetic dissection of polyamide–DNA interactions revealed that the association rates of linked polyamides, such as hairpin polyamides, are essentially diffusion limited [37]. Furthermore, discrimination between match and mismatch DNA sites is based in large part on differences in dissociation rates from these sites. With respect to hairpin dimers, their high affinities for mismatch sites equate to slow dissociation rates. In a biological context, when exposed to potentially thousands of mismatch sites, these polyamides may encounter problems equilibrating to their designed match sites in a reasonable time course. Recently it was shown that the minor groove of the DNA double helix could be used to template the formation of tandem dimers from functionalized hairpins (Fig. 5) [38]. The templated reaction proved to be specific for a ten bp site (over 11 or 12 bp sites) and dependent upon the DNA sequence (reaction was more efficient at fully matched DNA versus single- or double-base pair mismatch sites). The dimer product showed high affinity for its match sequence of DNA. Thus, high affinity, larger polyamides that target long sequences of DNA can be made in situ from smaller hairpin polyamides that display better kinetic profiles.

2.3 H-Pin and U-Pin Motifs

Polyamides also can be linked, via the ring nitrogens, with an alkyl spacer that projects away from the minor groove. When placed in the center of a polyamide, the resultant branched molecule has been termed an H-pin (Fig. 4b); when placed at the end, a U-pin. H-pin polyamides bind with high affinity and good specificity [39]. Recent efforts to improve the synthetic methods for H-pins using alkene metathesis on a solid support have enabled a detailed study of the optimal alkyl linker length, demonstrating that four and six methylene units provide the highest affinities [40]. U-pin polyamides behave similarly [41]. The affinity of an eight-ring U-pin is most comparable to a six-ring hairpin polyamide, most likely due to a loss of two hydrogen bond donors upon removal of the γ-turn element. Thus, the dimeric Py-Im U-turn element may be thought of as a C•G-specific replacement for the γ-turn. In combination with removal of the β-Ala-Dp tail, H-pin and U-pin polyamides could potentially bind purely G,C sites, a sequence type that has been difficult to target with other polyamide motifs.

2.4
1:1 Polyamide–DNA Complexes

Homopurine tracts have been a challenging target for polyamide binding because these DNA sequences have particularly narrow minor grooves [42]. While the width of the minor groove of these DNA sequences may be too narrow for hairpin polyamides, a single Py-Im-β-containing strand may be accommodated at such sequences.

In 1:1 polyamide–DNA complexes, β-linked polyamides appear to prefer a single orientation, N→C with respect to the 3′→5′ direction of the purine-rich strand [43]. Footprinting of 1:1 complexes has shown that Im residues do not distinguish G,C from A,T whereas Py and β residues prefer A,T over G,C base pairs [43].

A high-resolution 1:1 solution NMR structure of Im-β-ImPy-β-Im-β-ImPy-β-Dp elucidated the role of β-Ala in minor groove recognition (Fig. 6) [44]. β-Ala allows both Im rings in the Im-β-Im subunit to adapt to the relatively large dihedral required for hydrogen bonding. Additionally, close contacts of

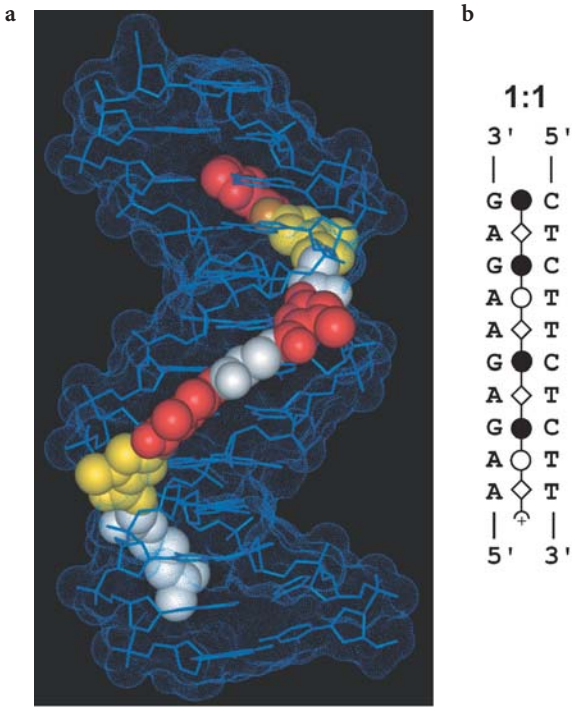


Fig. 6a,b **a** NMR structure of a 1:1 polyamide–DNA complex [44]. The DNA backbone and bases are shown in *black*. Im residues are *darkly shaded*, Py residues are *lightly shaded*, β-Ala and Dp residues are *white*. **b** Schematic model of the same interaction. Symbols are defined in Fig. 4

β-Ala to the floor of the minor groove provided a structural explanation for its observed A/T specificity.

While a four bp recognition code in the 1:1 motif has not been fully elucidated, this type of binding mode is uniquely suited to targeting homopurine sequences [43, 45]. Laemmli and coworkers reported a striking example of a 1:1 motif polyamide targeted to the satellite regions of *Drosophila melanogaster* being able to induce specific gain- and loss-of-function phenotypes when fed to developing flies [46].

2.5 Improving Synthetic Methodology

The investigation of minor groove-binding polyamides was greatly accelerated by the implementation of solid-phase synthesis [47]. Originally demonstrated on Boc-β-Ala-PAM resin with Boc-protected monomers, it was also shown that Fmoc chemistry could be employed with suitably protected monomers and Fmoc-β-Ala-Wang resin [48]. Recently, Pessi and coworkers used a sulfonamide-based safety-catch resin to prepare derivatives of hairpin polyamides [49]. Upon activation of the linker, resin-bound polyamides were readily cleaved with stoichiometric quantities of nucleophile to provide thioesters or peptide conjugates.

While allowing rapid preparation of a range of polyamides, these resins instal a T,A selective β-Ala residue at the C-terminus, which places limits on the DNA sites that can be targeted [24]. The shortest tail available from these resins

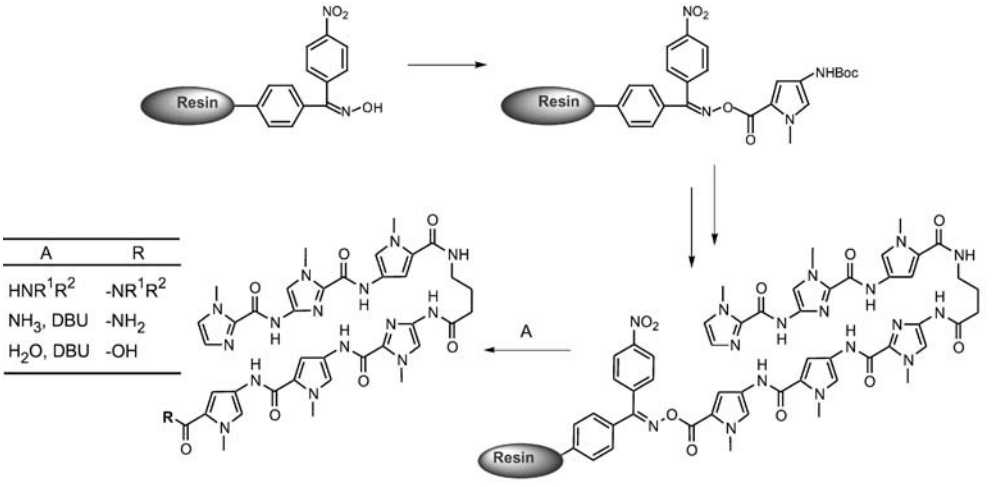


Fig. 7 Scheme for synthesis of polyamides on the Kaiser oxime solid-support resin. Cleavage from resin with various reagents (*A*) can result in polyamides with shorter C-terminal groups than molecules prepared on β-Ala-PAM resin [50]. The amine HNR^1R^2 may be a primary or secondary alkyl amine

is a propanolamide, obtained by reductive cleavage. Polyamides prepared on Boc-Gly-PAM resin can be reductively cleaved to obtain ethanolamide tails, but it was expected that further truncation of the C-terminus would be necessary for tolerance of G,C at the tail position [24]. The Kaiser oxime resin was therefore adapted to polyamide synthesis, allowing the preparation of polyamides with shorter C-termini (Fig. 7). These molecules display the desired tolerance for G,C bases while maintaining high affinities [50].

2.6 *N*-Terminal Monomers

The specificity of cofacial aromatic amino acid pairings is highly dependent upon their position within a given polyamide. For example, an Im/Py pairing is specific for G•C at both internal and terminal positions [13]. In contrast,

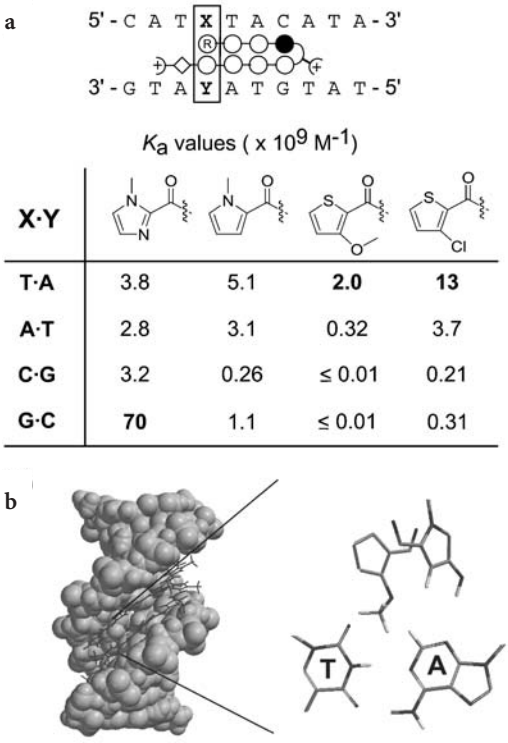


Fig. 8a,b **a** Table indicating the equlibrium association constants (K_a) for polyamides containing the indicated ring at the *N*-terminal position of the eight-ring hairpin polyamide pictured above at sequences containing each of the four Watson–Crick base pairs. When paired with Py, Im shows a preference for guanine while 3-substituted thiophenes show preference for thymine [51]. **b** Model created using the PC Spartan (Wavefunction) illustrating the shape-selective recognition of thymine by the 3-methoxy substituent of thiophene

Hp/Py and Py/Hp pairings, while specific for T•A and A•T, respectively at internal positions, lose all specificity when incorporated at the *N*-terminal cap position. The loss of specificity at the cap position is presumably a result of conformational freedom caused by the absence of a second "groove-anchoring" carboxamide, allowing terminal rings to bind DNA in either of two rotamers.

Recently, a library of substituted five-membered aromatic carboxylic acids was incorporated at the *N*-terminal position of an eight ring polyamide, and *N*-terminal specificity was probed [51]. It was found that 3-chlorothiophene (Ct) and 3-methoxythiophene (Mt) rings, when paired against Py, imparted a moderate degree of specificity for T•A over A•T (six- and threefold, respectively), and a large degree of specificity over G,C base pairs (>200-fold) (Fig. 8a). While the Mt/Py imparts a higher degree of specificity for T•A over A•T, the Ct/Py pair imparts modest specificity as well as higher affinity. Molecular modeling of Mt- and Ct-containing polyamides indicated that the rotamer that places the 3-substituent into the floor of the minor groove is energetically favored. This substituent fills the asymmetric cleft created by the N2 of thymine and the C2 of adenine, accounting for the observed preference for T•A (Fig. 8b).

2.7 Other Heterocyclic Ring Scaffolds

While most polyamide research has focused on five-membered aromatic ring systems, other scaffolds have been shown to bind DNA. The benzimidazole ring system represents a different structural framework which is amenable to functionalization on the six-membered ring and appears to impart a curvature that is complementary to DNA [52]. Indeed, the classic minor groove-binding Hoechst dyes are composed of benzimidazole units, and a number of derivatives of these molecules have been prepared. We have incorporated benzimidazole derivatives into the backbone of hairpin polyamides in a manner that preserves critical hydrogen bonding contacts and overall molecular shape (Fig. 9) [53, 54]. The imidazopyridine (Ip) and hydroxybenzimidazole (Hz) rings are introduced into polyamides as dimeric subunits PyIp and PyHz, respectively, in which the Py ring is directly connected to the benzimidazole derivative without an intervening amide bond [53, 54]. DNase I footprinting indicates that the Ip/Py and Hz/Py pairs are functionally identical, at least in some sequence contexts, to the analogous five-membered ring pairs Im/Py and Hp/Py [54]. Significantly, we have found that the Hp-containing polyamides can degrade over time in the presence of acid or free radicals, whereas the analogous Hz-containing compounds are chemically robust. Thus, the Hz/Py pair is a strong candidate for replacing Hp/Py in biological studies.

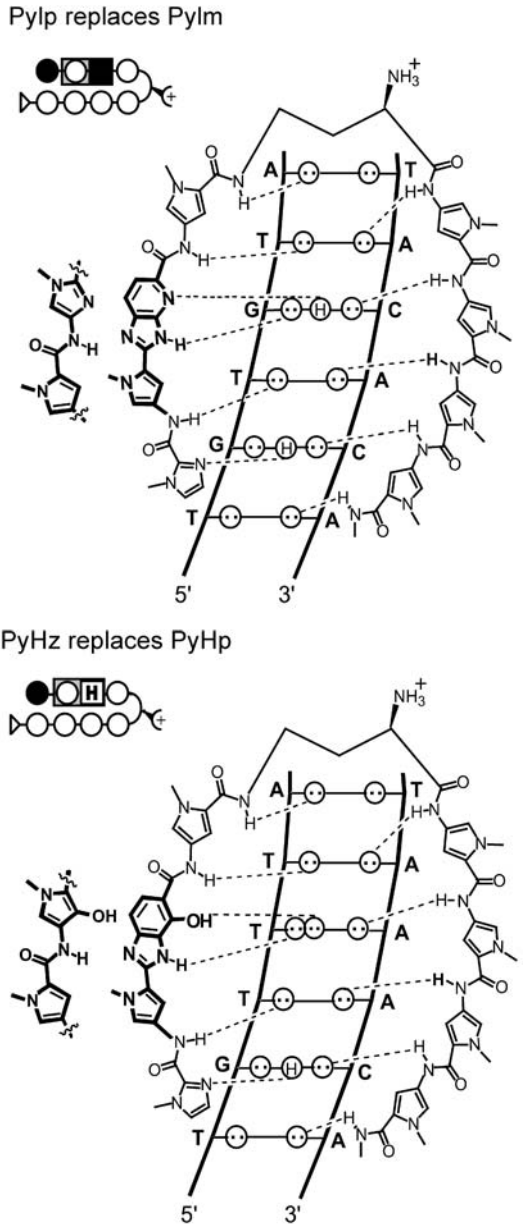


Fig. 9a,b Recognition of the DNA minor groove with benzimidazole derivatives [54]. **a** Structure of polyamide containing Py-imidazopyridine (PyIp). **b** Structure of polyamide containing Py-hydroxybenzimidazole (PyHz). The five-membered ring analogues of the dimeric benzimidazole derivatives are shown to the *right* of each model. Putative hydrogen bonds are indicated with *dashed lines*. *Rectangles containing a white circle and a black square* indicate the PyIp unit. *Rectangles containing a white circle and a square with an H* indicate the PyHz unit. Other symbols are defined in Figs. 2 and 4

3
Secondary Effects of Polyamides

3.1
Displacement of DNA-Binding Proteins

Polyamides bind with high affinity to a wide range of DNA sites and can competitively displace many proteins from DNA. This can have an effect on gene expression, as DNA-binding proteins are often involved in the regulation of transcription. One approach to modifying gene expression involves inhibition of key transcription factor (TF)–DNA complexes in a designated promoter, thus interfering with recruitment of RNA polymerases [55]. Significantly, because there are considerably fewer oncogenic TFs than potentially oncogenic signaling proteins, TF inhibition represents a uniquely promising approach to cancer treatment [56]. The transcription factor TFIIIA was chosen as a first target because it regulates a relatively small number of genes and because the contacts between the nine zinc-finger protein and the minor groove had been established. A polyamide bound in the recognition site of TFIIIA suppressed transcription of 5 S RNA genes by RNA polymerase III in vitro and in cultured *Xenopus* kidney cells [55]. Further studies used polyamides in combination with recombinant derivatives of TFIIIA subunits to elucidate essential minor groove contacts for the binding of this TF [57].

Polyamides were then used to target viral genes transcribed by RNA polymerase II. The HIV-1 enhancer/promoter contains binding sites for multiple transcription factors, including TBP, Ets-1, and LEF-1. Two hairpin polyamides designed to bind DNA sequences immediately adjacent to the binding sites for LEF-1 and Ets-1 specifically inhibited binding of each transcription factor and HIV-1 transcription in a cell-free assay (Fig. 10c,d) [58]. In human blood lymphocytes, treatment with the two polyamides in combination inhibited viral replication by 99%, with no significant decrease in cell viability. Inhibition of viral replication is indirect evidence for specific transcription inhibition by polyamides, because other modes of action could be involved, such as modulation of T-cell activation pathways. However, RNase protection assays indicated that the two polyamides did not alter the RNA transcript levels of several cytokine and growth factor genes, suggesting that polyamides may affect transcription directly.

This early biological result spurred a variety of biochemical studies of the interactions of polyamides with the basal transcription machinery and TF–DNA complexes. Two studies have used promoter scanning to identify sites where polyamide binding inhibits transcription [59, 60]. The method uses a series of DNA constructs with designed polyamide binding sites at varying distances from the transcription start site. Essential minor groove contacts were identified for a subunit of TFIIIB (possibly TBP) in a *Xenopus* tRNA promoter [60], as well as for TFIID–TFIIA and TBP in the HIV-1 core promoter [59]. The binding of the homodimeric basic-helix-loop-helix TF Deadpan was investi-

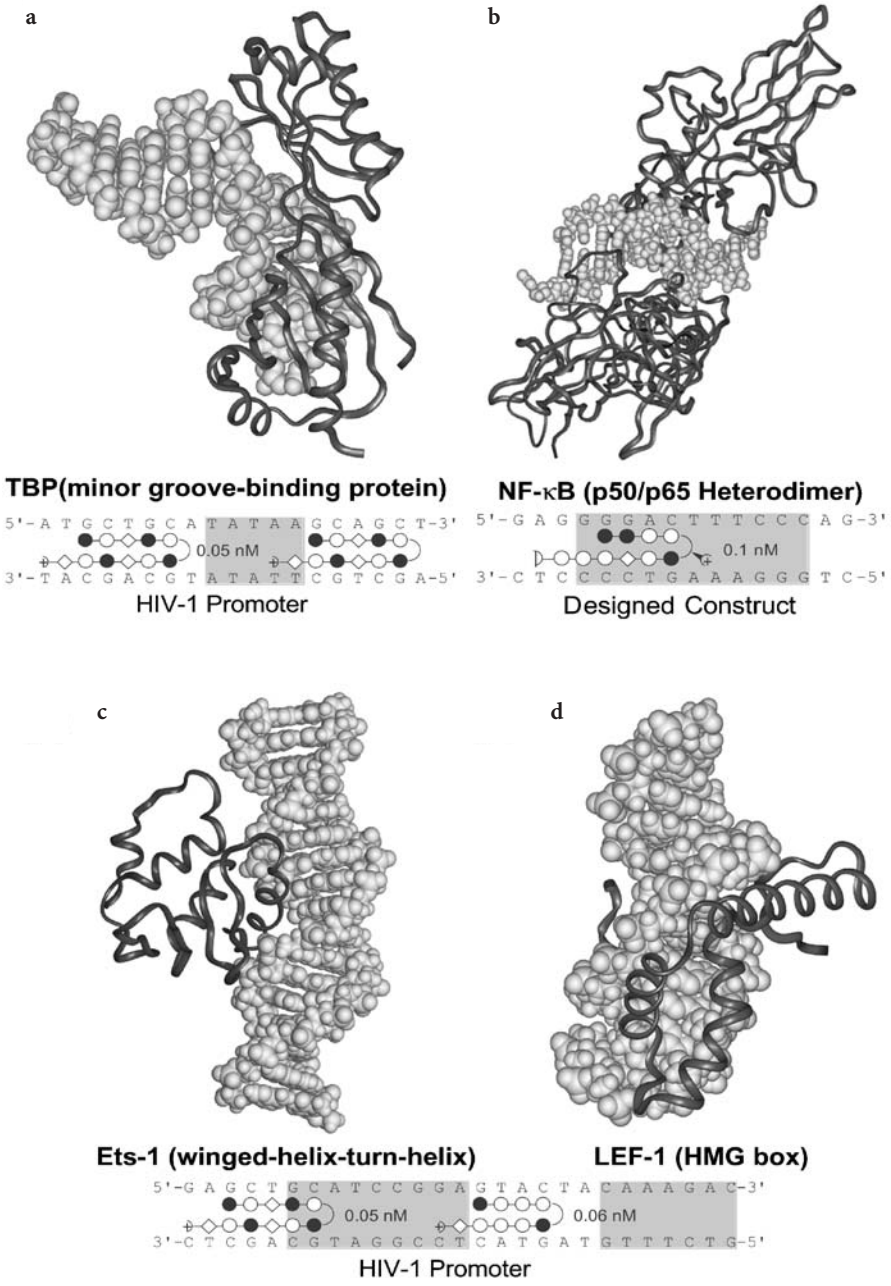


Fig. 10a–d X-ray crystal structures of four different protein–DNA complexes that have been inhibited by polyamides. Below each structure is illustrated the context for inhibition, with the protein binding sites shaded and the polyamides responsible for inhibition shown bound to their match sites [58, 64]. Listed beside each polyamide is its dissociation constant (K_d). Symbols are as defined in Fig. 4

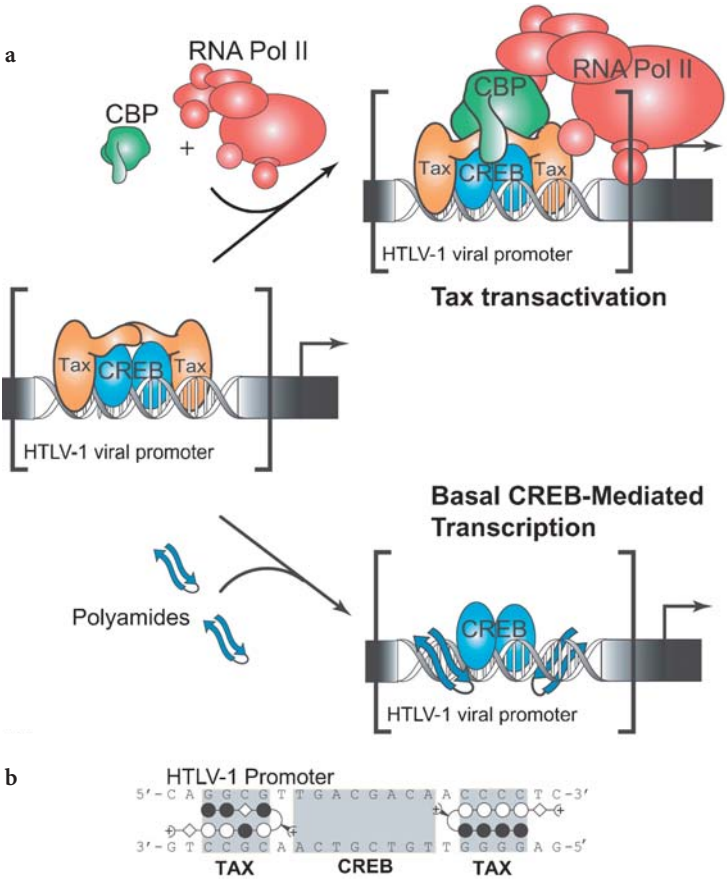


Fig. 11a,b **a** Polyamide inhibition of Tax transactivation [62]. Tax and CREB are bound to the HTLV-1 viral promoter. The trimeric complex recruits CBP and the transcriptional machinery. Polyamides specifically inhibit Tax but not CREB from binding to DNA, thus abolishing Tax transactivation while leaving basal CREB-mediated transcription unaffected. **b** Model of the sequence recognized by Tax and CREB with the structures of the polyamides used to inhibit Tax shown bound to their targeted sequences. All symbols are as defined in Fig. 4, with Tax and CREB binding sites *shaded*

gated using a variant of promoter scanning [61]. A series of duplex oligonucleotides based on a *Drosophila* neural promoter were designed, incorporating polyamide binding sites on different sides of the Deadpan recognition sequence and in different orientations. The TF–DNA complex was inhibited only by a polyamide binding upstream of the homodimer, establishing an asymmetric binding mode for this TF.

In the human T-cell leukemia virus type 1 (HTLV-1) promoter, polyamides targeted to G,C-rich regions flanking the viral CRE sites inhibited binding of the Tax protein and Tax transactivation in vitro (Fig. 11) [62]. This example

illustrates several important polyamide–DNA–protein interactions. HTLV-1 genes are regulated by the major groove-binding protein CREB (CRE binding protein). CREB-mediated transcription is enhanced by the binding of the viral protein Tax to the CRE flanking regions which makes contacts with CREB and to the minor groove. Tax then recruits CREB binding protein (CBP) via a KIX domain on CBP, which then induces transcription. Researchers found that addition of two polyamides designed to target the Tax recognition elements, inhibited Tax from associating to the CREB–DNA complex, and Tax-induced transcription was abolished. Interestingly, these polyamides bind only a few base pairs away from the CRE site, yet CREB is able to co-occupy the DNA, with CREB-mediated basal transcription remaining intact. Thus, polyamides are able to interfere very specifically with some protein–DNA interactions while leaving other nearby interactions unaffected.

Several other protein–DNA interactions have been inhibited with polyamides. Bacterial gyrase recognizes a short 5′-GGCC-3′ site, and a polyamide targeted to this sequence inhibited gyrase-catalyzed strand cleavage at nanomolar concentrations [63]. NF-κB is a TF crucial for development, viral expression, inflammation, and anti-apoptotic responses. The most common form is a p50-p65 heterodimer, which binds DNA in the major groove, making several phosphate contacts throughout the binding site. Polyamides targeted to the minor groove opposite p50, but not p65, inhibit major-groove DNA binding by NF-κB, presumably by an allosteric mechanism (Fig. 10b) [64]. This illustrates how minor-groove binding polyamides may interfere with DNA-binding proteins without actually contacting the protein [65]. In a different study, polyamides were shown to bind very near the 3′ processing end of Moloney Murine leukemia virus (M-MuLV) long terminal repeat (LTR) sequences, thereby inhibiting retroviral integration catalyzed by M-MuLV integrase (IN) [66].

The binding of Ets-1 to the HIV-1 enhancer was examined in greater detail, and polyamides were shown to inhibit the formation of a ternary Ets-1–NF-κB–DNA complex [67]. Ets-1 is a winged-helix-turn-helix TF, and its key phosphate contacts on either side of the major groove can be disrupted by a polyamide in the adjacent minor groove. The report provided evidence for cooperative DNA binding by Ets-1 and NF-κB to the HIV-1 enhancer sequence. A different Ets binding site in the HER2/*neu* promoter was targeted with hairpin polyamides that successfully blocked Ets–DNA complex formation and transcription of the HER2/*neu* oncogene in a cell-free system [68].

Other purely major groove-binding TFs, such as the basic-region leucine zipper (bZIP) protein GCN4, have been shown to co-occupy the DNA helix in the presence of polyamides [69]. Strategies employing polyamides functionalized with helix-distorting moieties have been successful at inhibiting such proteins. Polyamides with an attached Arg-Pro-Arg tripeptide can interfere with major-groove binding proteins by disrupting key phosphate contacts, distorting the DNA by charge neutralization, or sterically invading the major groove. An Arg-Pro-Arg-polyamide conjugate successfully inhibited the bind-

ing of GCN4 to DNA [69], and further optimization yielded a polyamide derivative with an alkyl diamine substituent that was tenfold more potent [70].

Polyamide-intercalator conjugates that distort the DNA at specific, targeted sequences by insertion of an intercalator have also proved potent inhibitors of major groove-binding proteins. Polyamides conjugated to the intercalator acridine disrupt the DNA microstructure via unwinding, and were shown to significantly inhibit GCN4 binding when bound to sites adjacent to the GCN4 binding site, placing their acridine moieties into the GCN4 recognition element (Fig. 12). Such molecules are promising candidates for site-selective inhibition of any DNA-binding protein [71].

Polyamides can also upregulate transcription by inhibition of a repressor protein (derepression). For example, a hairpin polyamide was shown to block binding of the repressor IE86 to DNA, thereby upregulating transcription of the human cytomegalovirus MIEP [72]. A more complex case involves derepression of the integrated HIV-1 long terminal repeat (LTR). The human protein LSF binds in the promoter region at the LTR and recruits YY1, which then recruits histone deacetylases (HDACs). HDACs subsequently maintain LTR quiescence, which has been implicated in HIV latency, by maintaining a silent stock of pathogen. Three different live cell models demonstrated that polyamides can inhibit LSF binding and increase expression of integrated HIV-1 promoter [73]. As with other systems, only polyamides matched to the correct protein binding site induced significant effects. Several existing drug treatments can reduce HIV-1 levels in the blood to below detectable amounts, yet the virus inevitably returns in infected patients. Derepression by inhibition

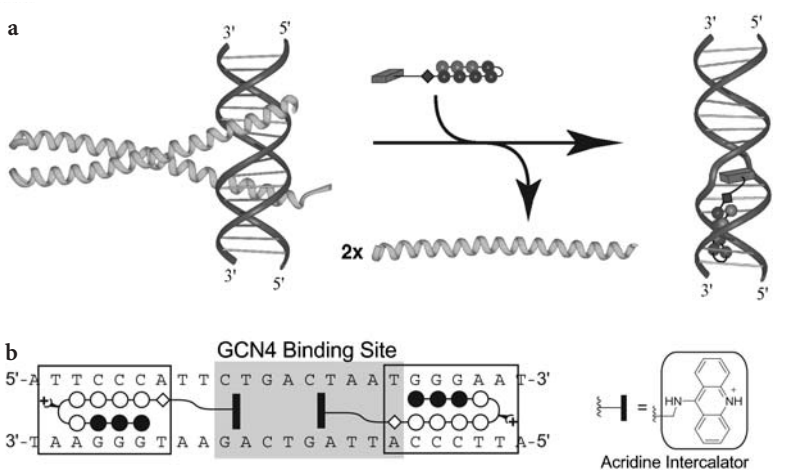


Fig. 12a,b Model for allosteric inhibition of a protein–DNA complex by a polyamide–intercalator conjugate [71]. **a** The GCN4 homodimer is displaced by the intercalating moiety of the polyamide conjugate. **b** Ball-and-stick model of the polyamide conjugate binding the target site (*boxed*) adjacent to the binding of the protein GCN4 (*shaded*). The structure of the acridine intercalator is shown at *right*. All other symbols are as defined in Fig. 4

of LSF–DNA binding may eventually allow HIV to be fully eradicated by drug treatments. This approach is particularly promising because LSF is a human protein, which could make the target less susceptible to resistance by HIV-1 mutations.

3.2 Recruitment of DNA-Binding Proteins

Polyamides have also been shown to affect DNA structure and function by recruiting proteins to specific, targeted sites. Most transcription factors have a DNA-binding domain and a separate domain that recruits the transcriptional machinery to that site (often called the activation domain). A polyamide can be thought of as an artificial DNA-binding domain that can be linked to an activation domain. Such artificial transcription factors have been synthesized and evaluated in cell-free transcription assays [74, 75]. A hairpin polyamide tethered by a 36-atom straight chain linker to the short (20-residue) peptide activation domain AH gives robust activation of transcription, with a size of only 4.2 kDa. Replacing the AH peptide with the shorter yet more potent activator VP2, derived from the activator domain of the viral activator VP16, and reducing the linker from 36 to eight atoms provided a "minimal" polyamide–peptide conjugate, 3.2 kDa in size, which activated transcription slightly more effectively than the larger analogue (Fig. 13a) [75]. Since the linker length had been shown to influence activation efficiency, a set of molecules with rigid oligoproline linkers between the polyamide and the activation domain was synthesized [76]. The oligoproline linkers act as "molecule rulers," and optimal activation was observed with a Pro_{12} linker, about 36 Å in length.

Many genes are influenced by multiple pathways and thus rely on the binding of several proteins. One example is the Hox (Homeobox) family of transcriptional regulators, which plays a vital role in the developmental fate of an organism. However, Hox proteins generally display poor affinity and sequence specificity towards DNA. Instead, they are recruited to DNA by the strong, specific binding of members of the TALE (three amino acid loop extension) class of homeodomain proteins. Recent crystal structures of one such ternary complex shows that the Hox protein Ultrabithorax (Ubx) interacts with the *Drosophila* TALE protein extradenticle (Exd) via a short docking YPWM peptide [77]. A polyamide functionalized with this YPWM peptide successfully recruited Exd at nanomolar concentrations, outperforming the natural Ubx protein (Fig. 13b) [78]. This demonstrates that cooperative interactions among functionalized polyamides, DNA, and a protein can stabilize the formation of a ternary complex on a composite DNA site in vitro.

Polyamide–camptothecin conjugates specifically recruited DNA topoisomerase I (Topo I) and induced single-strand cleavage [79]. Camptothecin is known to stabilize the cleavage complex formed between a tyrosine residue on Topo I and the 3′-phosphoryl end of the DNA backbone [80]. Using polyamide–camptothecin conjugates, this cleavage complex could be generated

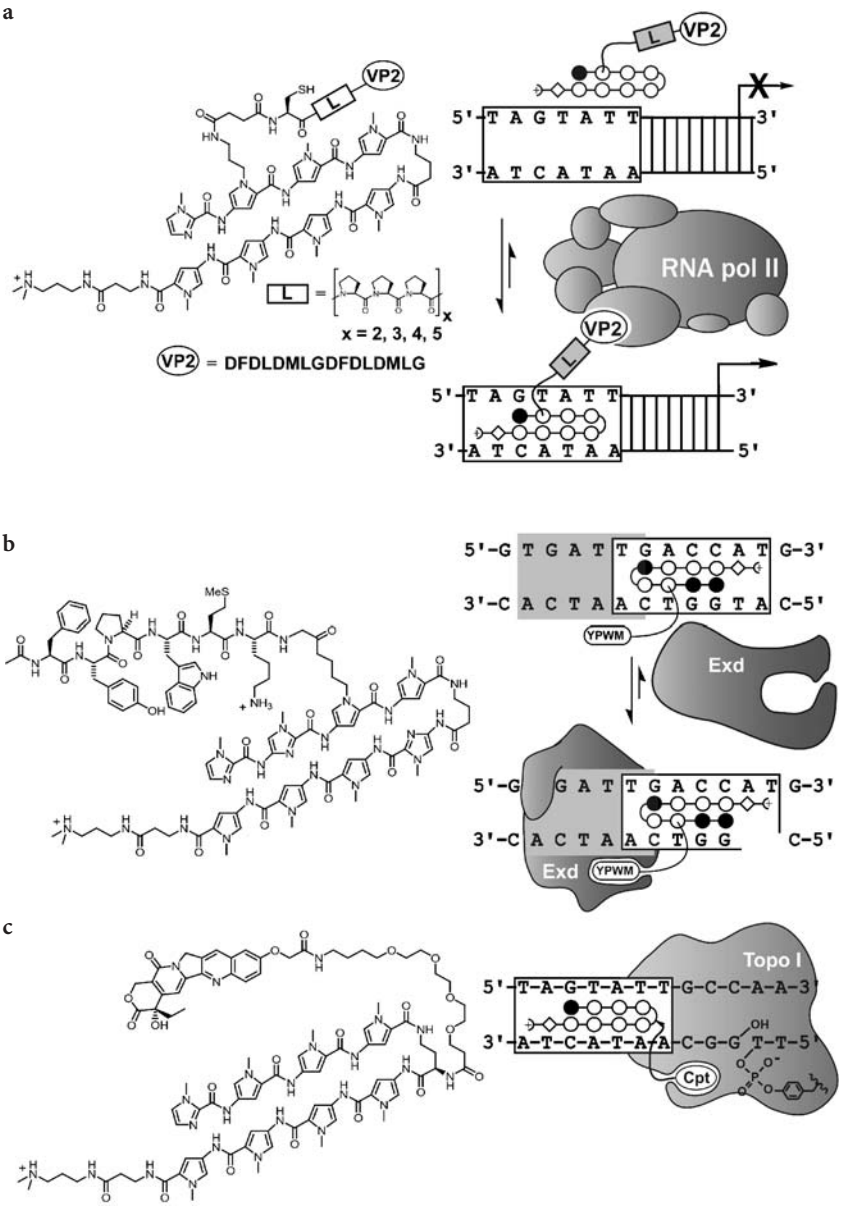


Fig. 13a–c Recruitment of cellular proteins to DNA by polyamides. **a** Polyamides conjugated to the VP2 activation domain via a rigid poly-proline linker recruit the transcriptional machinery to a targeted site [76]. **b** A polyamide functionalized with a short, YPWM peptide recruits Exd at nanomolar concentrations, changing the protein from a non-DNA-binding conformation (*top*) to one that binds the DNA–polyamide complex with high affinity (*bottom*) [78]. **c** Polyamide–camptothecin conjugates recruit topoisomerase I, inducing specific, targetable DNA strand breaks [79]. Protein binding sites are *shaded*, and polyamide binding sites are *boxed*. All other symbols are labeled or defined in Fig. 4

sequence-specifically at sites adjacent to the polyamide binding site (Fig. 13c). Since camptothecin–Topo I–DNA complexes have been shown to arrest transcription elongation [81], polyamide–camptothecin conjugates may function as sequence-specific transcription terminators.

3.3 Sequence-Specific DNA Alkylation

Another type of bifunctional polyamide is able to covalently react with the minor groove of DNA. Several classes of alkylating agents have been conjugated

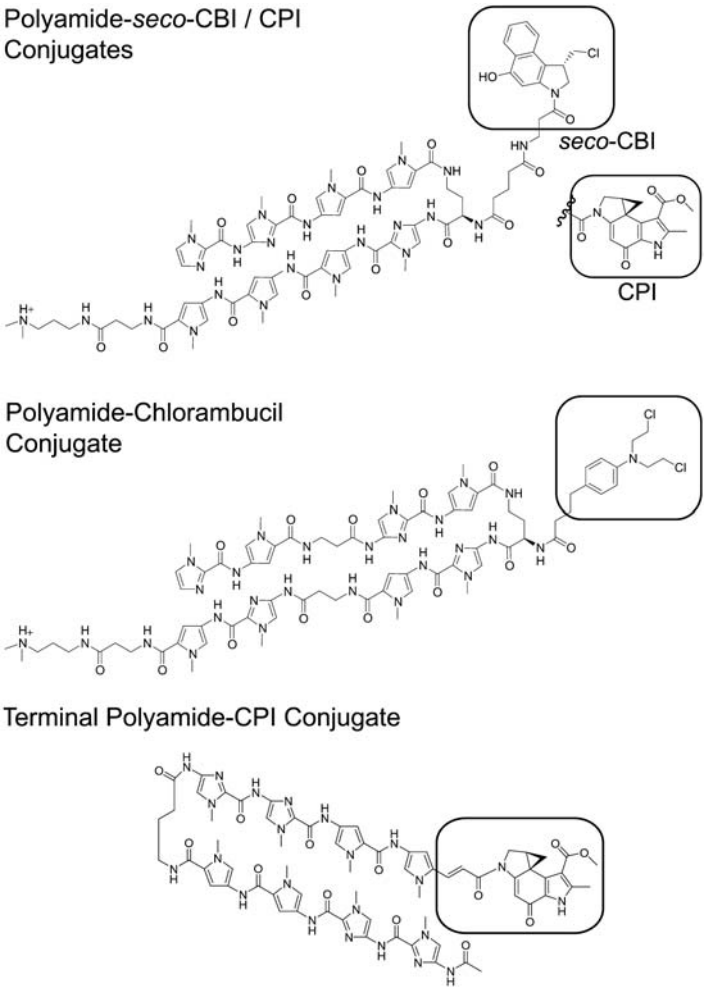


Fig. 14 Structures of several classes of polyamide–alkylator conjugates [82–88]. The alkylating moiety is *boxed*

to the C-terminal tail or hairpin "turn" unit of DNA-binding polyamides (Fig. 14) [82, 83]. Polyamide–*seco*-CBI conjugates at nanomolar concentrations showed alkylation at adenines proximal to the binding site within 12 h [82]. Polyamides conjugated to the close analogue, CPI, targeted to the human telomere repeats specifically alkylated telomere adenine sites, and were active against several tumor cell lines, in some cases outperforming cisplatin and bleomycin [84]. CPI-conjugated polyamides were recently shown to inhibit in vitro transcription elongation when targeted to a coding sequence, an effect not seen by unfunctionalized polyamides [85]. Polyamide–CPI conjugates were also shown to silence genes transfected into human cancer cell lines in a sequence-specific manner [86].

Polyamide–chlorambucil conjugates also specifically alkylate predetermined sites in the minor groove [83, 87]. It is likely that the slower rate of alkylation for the chlorambucil moiety (versus *seco*-CBI alkylation) allows for an increased specificity of alkylation to polyamide match sites. Chlorambucil-linked polyamides alkylated specific sites in the HIV-1 LTR both in vitro and in live CEM T-cells with a stably integrated copy of HIV-1 LTR [87, 88].

4
Cellular Studies

4.1
Nucleosomes

In eukaryotic cells, DNA is tightly packaged by compaction into chromatin, and changes in chromatin structure can alter the accessibility of specific sequences and affect components of the molecular machinery in the nucleus. The fundamental repeating unit of chromatin is the nucleosome, comprising a 20–80 bp DNA linker region and the nucleosome core particle (NCP) – roughly two tight superhelical turns of DNA (147 bp in length) wrapped around a disk of eight histone proteins. The ability of DNA-binding proteins to recognize their cognate sites in chromatin is restricted by the structure and dynamics of nucleosomal DNA, and by the translational and rotational positioning of the histone octamer. Using six different hairpin polyamides, it was shown that sites on nucleosomal DNA facing away from the histone octamer, or even partially facing the octamers, are fully accessible [89]. Remarkably, one section of 14 consecutive base pairs – more than a full turn of the DNA helix – was accessible for high affinity polyamide binding. The only positions very poorly bound by polyamides were sites near the amino-terminal tails of histone H3 or histone H4. Removal of either tail allowed polyamides to bind, suggesting that both the structure of the DNA and perhaps its rotational position are strongly influenced by the *N*-terminal tails of histone H3 and H4 [90].

Subsequently, the structures of three of these polyamide–NCP complexes were determined by X-ray crystallography [91]. The histone octamer is unaf-

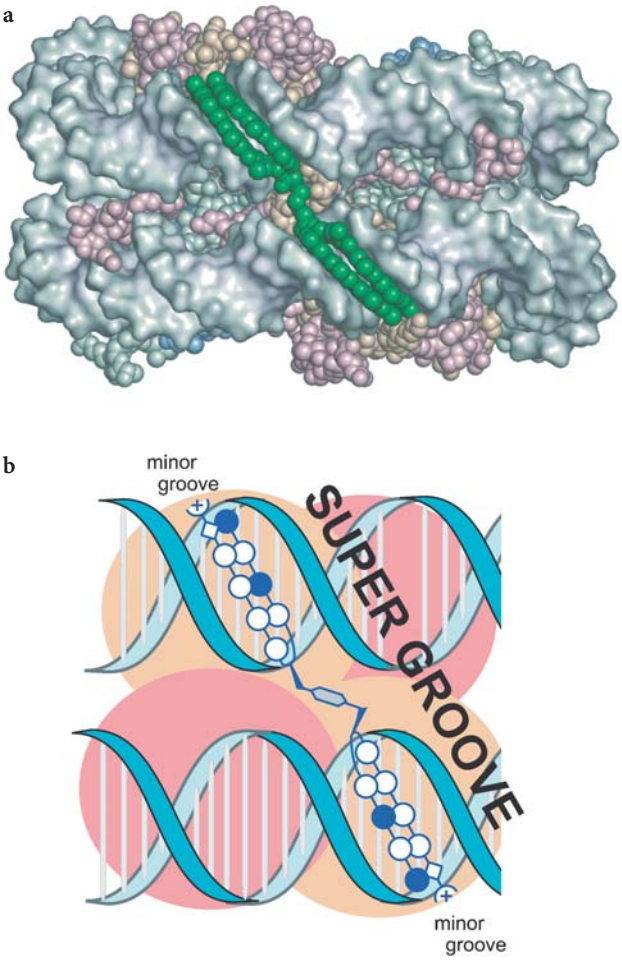


Fig. 15a,b **a** Detail of the X-ray crystal structure of the nucleosome core particle (NCP) with bound "clamp" polyamide dimer (PDB code **1S32**) [92]. DNA helices run horizontally. Protein elements of the NCP are colored *red and orange*. Polyamide bound in the "supergroove" is *green*. **b** Model for supergroove recognition by turn-to-turn linked tandem hairpin polyamide. *Dark shaded shapes* represent the protein components of the NCP. The DNA helices are *blue*

fected by polyamide binding, but the nucleosomal DNA undergoes significant structural changes at the ligand binding sites and the adjacent regions. Significantly, distortions in DNA twist can propagate over long distances without disrupting histone–DNA contacts, giving a potential mechanistic rationale for the role of twist diffusion in nucleosome translocation. Although the three polyamides display very similar affinities for their binding sites in the α-satellite nucleosome particle, only the relatively non-specific polyamide ImPyPyPy-

γ-PyPyPyPy-β-Dp inhibits temperature-induced nucleosome translocation [90]. This may indicate that ligand positioning is critical, such that a single properly placed polyamide would effectively block translocation; or that the small effects of a single bound ligand can be amplified, such that a combination of several different polyamides would block translocation.

Although polyamides can block transcription by targeting promoter elements, they do not affect transcription when bound in the RNA coding regions of DNA [90]. Presumably, the strand melting required for RNA polymerase progression disrupts the minor groove and displaces polyamides. To investigate potential effects on transcription through a nucleosome, hairpin polyamides were targeted to sites on the nucleosome positioning sequence of the sea urchin 5S gene [90]. The two molecules that blocked heat-induced nucleosomal translocation also blocked transcription by T7 RNA polymerase. Each of these polyamides binds with high affinity to a single site in the nucleosome construct, potentially implying that placement is critical. Nonetheless, the positions of these sites are distinct from those occupied by the compound ImPyPyPy-γ-PyPyPyPy-β-Dp in the crystal structure [91]. Although the precise mechanisms involved in nucleosome repositioning remain in question, it appears that, in some cases, DNA can "roll" over the histones, and certain polyamides can act as chocks to prevent the DNA from moving.

The nucleosome architecture brings regions of DNA located 80 base pairs apart on linear DNA into close proximity, forming a DNA "supergroove." One of the polyamide–NCP crystal structures revealed a striking alignment of polyamides bound along one "supergroove," where the polyamide γ-turn moieties are juxtaposed [90]. Dimer polyamide hairpins linked turn-to-turn by short PEG linkers are able to target the supergroove, effectively crosslinking the two gyres of DNA [92] (Fig. 15). Such supergroove targeting becomes relevant to biological systems, as the majority of cellular DNA is nucleosome-bound, and may offer a strategy for gene regulation.

4.2 Nuclear Uptake

DNA-binding polyamides can inhibit and influence a wide variety of protein–DNA interactions in solution, yet effectiveness in cell culture has proved to be dependent on cell type. A series of fluorescent labeled polyamides was prepared to analyze the intracellular distribution of these molecules in a panel of cell lines [93]. In cell types that had shown robust responses to polyamides, such as primary human T-cells, fluorescent polyamide–bodipy conjugates were observed to enter the nuclei of live cells [93]. However, in the majority of cell lines, polyamide–bodipy conjugates were excluded from the nucleus. Co-staining with organelle-specific dyes indicates that polyamide–bodipy conjugates are often trapped in lysosomes and other cytoplasmic vesicles [93], such that cells treated with polyamides can give a false nuclear signal upon fixing, even if they are washed extensively. Bashkin and coworkers have demonstrated that a poly-

amide–bodipy conjugate will traffic to the nucleus of a human cell line in the presence of verapamil, a P-glycoprotein inhibitor [94].

Recently, a series of fluorescein-labeled polyamides were assayed for nuclear uptake against a panel of live mammalian cells [95]. In some cases, small changes, such as the removal of a β-Ala residue at the C-terminus of a polyamide dramatically enhanced nuclear localization (Fig. 16 **1** vs **2**). Nuclear uptake of tested polyamide–fluorescein conjugates is an energy-dependent process. HeLa cells grown in energy inhibitory medium (supplemented with 2-deoxyglucose and sodium azide) displayed little to no discernable nuclear staining when treated with a fluorescein-labeled polyamide, while the same cells grown in normal medium showed clear nuclear staining. Washing to remove the inhibitory medium, and replacement with normal medium (supplemented with additional polyamide) resulted in nuclear staining.

While there are currently no general rules for cellular uptake of polyamides, determinants such as polyamide size, imidazole content (Fig. 16 **2** vs **3**), struc-

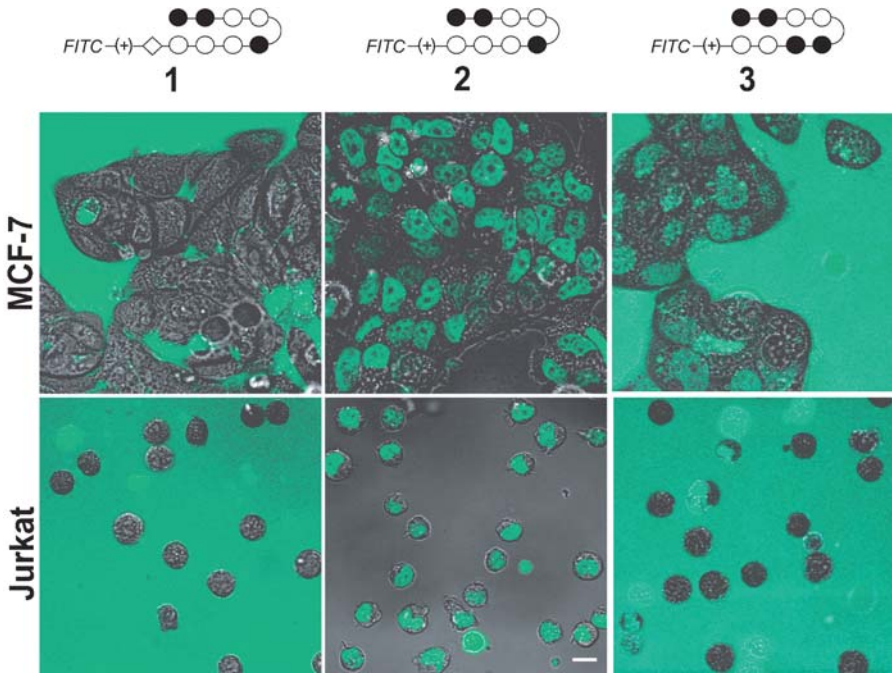


Fig. 16 Nuclear uptake of hairpin polyamides [95]. Representative confocal microscopy images of fluorescein-labeled polyamides in MCF-7 and Jurkat cells. Polyamide **1** exhibits poor uptake and is excluded from the nuclei of both cell types. Removal of the C-terminal β-Ala residue results in a polyamide **2** with excellent uptake properties, localizing to the nuclei of both cell types. Polyamide **3** differing from **2** by a single pyrrole to imidazole substitution, localizes to MCF-7 nuclei less strongly than does **2**, and **3** is completely excluded from Jurkat cells. *Green/light* areas indicate areas of polyamide–fluorescein localization

ture and attachment point of the fluorescent dye, and structure of the "tail" are important for nuclear localization. Each cell line possess a unique uptake profile such that choices of specific cell lines and compound architectures will be critical for future biological experiments.

4.3 Antifungal and Antibacterial Properties of Polyamides

In work on living bacterial and fungal cells [96–99], Bürli and coworkers reported that modification of the C- and *N*-terminal tails of a four-ring polyamide core results in compounds with vastly different antibacterial and antifungal properties. Several of the compounds were extremely effective antibacterial agents, even against Gram-positive strains resistant to vancomycin and other clinically proven antibiotics [96]. The researchers found good correlation between DNA affinity and in vitro efficacy, making the bacterial genome a potentially novel antibacterial target. Further work in this area identified polyamides with novel *N*-terminal ring systems and C-terminal tail substituents that effectively inhibit growth of several bacterial strains but have no adverse effects on mice in vivo [98]. The addition of a second positive charge at the C-terminus of a hairpin polyamide confers activity against a number of clinically relevant fungal strains in vitro and activity against *Candida albicans* in mice. Control experiments indicate that the observed antifungal activity results from a DNA-binding mechanism that does not involve DNA damage or disruption of chromosomal integrity [100].

4.4 Genetic Profiling of Polyamides

While the question of specificity on a genomic scale is still not well understood, methods such as genome-wide analysis of gene expression on oligonucleotide arrays (GeneChips) are now available to pursue the issue [101]. What will be the minimum requirements for selectively altering transcription of a single gene? It may be that multiple "small polyamides" (each targeting six bp) acting in concert will be more effective than one single large molecule (targeting 10–12 bp). Genome-wide analysis of the effects of hairpin polyamides have already begun. Researchers found that of a panel of 18,000 genes from human MT2 cells, treatment with a polyamide significantly changed the expression levels of only 21 genes [88]. Furthermore, a polyamide with a different target sequence also affected 21 genes, but only ten in common with the first polyamide (Fig. 17). Researchers went on to show that in almost every gene affected, there was a match site for the polyamide proximal to the transcription start site [88]. Work is currently ongoing toward generating a database of transcription profiles for different polyamides in human cells.

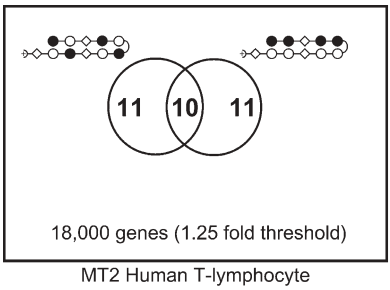


Fig. 17 Venn diagram representing the number of genes for which transcription was altered in a genome-wide screen of MT2 human cells treated with two different polyamides. Each polyamide altered the expression of only 21 out of 18,000 genes, with the threshold set at 1.25-fold change [88]

5 Conclusion

Protein engineering and triple helix-forming oligonucleotide technology continue to play important roles in the sequence-specific recognition of DNA [102–104]. Although gene suppression techniques have made extraordinary advances, most notably with the advent of RNAi technology, which targets mRNA [104, 105], DNA-binding polyamides are the only cell-permeable small molecules capable of recognizing specific, predetermined sequences of double helical DNA. However, a major challenge will be nuclear uptake of functionalized polyamide conjugates, including polyamide-bearing moieties for activation or protein recruitment. Finally, significant progress continues with regard to peptide–nucleic acids (PNA) targeting DNA [106, 107]. It may be that a basic research portfolio of different approaches to gene suppression and activation is the best path forward until one or more are sufficiently mature for the investment of time and resources necessary for drug development.

Acknowledgements We are grateful to the National Institutes of Health for research support. A.T.P.-K. is supported by an NIH Predoctoral Training Grant. E.J.F. is supported by an NIH Research Service Award and Ralph M. Parsons Fellowship. B.S.E. is supported by a predoctoral fellowship from the Howard Hughes Medical Institute.

References

1. Arcamone F, Nicolell V, Penco S, Orezzi P, Pirelli A (1964) Nature 203:1064
2. Zimmer C, Wahnert U (1986) Progr Biophys Mol Biol 47:31
3. Schuhman E, Haupt I, Thrum H, Taubenec U, May U (1974) Zeitsch Allgem Mikrobiol 14:321
4. Ginsburg H, Nissani E, Krugliak M, Williamson DH (1993) Mol Biochem Parasitol 58:7
5. Thrum H, Haupt I, Bradler G, Zimmer CG, Reinert KE (1972) Antimicrobial and antineoplastic chemotherapy. Czech Med Press, Prague

6. Verini MA, Ghione M (1964) Chemotherapy 9:144
7. Mars G, Regoli U (1968) Clin Ter 30:573
8. Kopka ML, Yoon C, Goodsell D, Pjura P, Dickerson RE (1985) Proc Natl Acad Sci USA 82:1376
9. Coll M, Frederick CA, Wang AHJ, Rich A (1987) Proc Natl Acad Sci USA 84:8385
10. Pelton JG, Wemmer DE (1989) Proc Natl Acad Sci USA 86:5723
11. Dervan PB, Edelson BS (2003) Curr Opin Struct Biol 13:284
12. Dervan PB (2001) Bioorg Med Chem Lett 9:2215
13. Wade WS, Mrksich M, Dervan PB (1992) J Am Chem Soc 114:8783
14. Kielkopf CL, Baird EE, Dervan PD, Rees DC (1998) Nature Structural Biology 5:104
15. Pilch DS, Poklar N, Baird EE, Dervan PB, Breslauer KJ (1999) Biochemistry 38:2143
16. White S, Szewczyk JW, Turner JM, Baird EE, Dervan PB (1998) Nature 391:468
17. Kielkopf CL, White S, Szewczyk JW, Turner JM, Baird EE, Dervan PB, Rees DC (1998) Science 282:111
18. Kielkopf CL, Bremer RE, White S, Szewczyk JW, Turner JM, Baird EE, Dervan PB, Rees DC (2000) J Mol Biol 295:557
19. Wellenzohn B, Loferer MJ, Trieb M, Rauch C, Winger RH, Mayer E, Liedl KR (2003) J Am Chem Soc 125:1088
20. Mrksich M, Parks ME, Dervan PB (1994) J Am Chem Soc 116:7983
21. Trauger JW, Baird EE, Dervan PB (1996) Nature 382:559
22. deClairac RPL, Geierstanger BH, Mrksich M, Dervan PB, Wemmer DE (1997) J Am Chem Soc 119:7909
23. White S, Baird EE, Dervan PB (1997) J Am Chem Soc 119:8756
24. Swalley SE, Baird EE, Dervan PB (1999) J Am Chem Soc 121:1113
25. Hawkins CA, de Clairac RP, Dominey RN, Baird EE, White S, Dervan PB, Wemmer DE (2000) J Am Chem Soc 122:5235
26. Herman DM, Baird EE, Dervan PB (1998) J Am Chem Soc 120:1382
27. Dervan PB, Urbach AR (2001) In: Quinkert G, Kisakurek MV (eds) Essays in contemporary chemistry. Helv Chim Acta, p 327
28. Urbach AR (2002) 1:1 Motif for DNA recognition by β-alanine-linked polyamides. Thesis, California Institute of Technology
29. Woods CR, Ishii T, Wu B, Bair KW, Boger DL (2002) J Am Chem Soc 124:2148
30. Kelly JJ, Baird EE, Dervan PB (1996) Proc Natl Acad Sci USA 93:6981
31. Turner JM, Swalley SE, Baird EE, Dervan PB (1998) J Am Chem Soc 120:6219
32. Trauger JW, Baird EE, Dervan PB (1998) J Am Chem Soc 120:3534
33. Herman DM, Baird EE, Dervan PB (1999) Chem Eur J 5:975
34. Kers I, Dervan PB (2002) Bioorg Med Chem Lett 10:3339
35. Weyermann P, Dervan PB (2002) J Am Chem Soc 124:6872
36. Maeshima K, Janssen S, Laemmli UK (2001) Embo J 20:3218
37. Baliga R, Baird EE, Herman DM, Melander C, Dervan PB, Crothers DM (2001) Biochemistry 40:3
38. Poulin-Kerstien AT, Dervan PB (2003) J Am Chem Soc 125:15811
39. Greenberg WA, Baird EE, Dervan PB (1998) Chem Eur J 4:796
40. Olenyuk B, Jitianu C, Dervan PB (2003) J Am Chem Soc 125:4741
41. Heckel A, Dervan PB (2003) Chem Eur J 9:3353
42. Hunter CA (1993) J Mol Biol 230:1025
43. Urbach AR, Dervan PB (2001) Proc Natl Acad Sci USA 98:4343
44. Urbach AR, Love JJ, Ross SA, Dervan PB (2002) J Mol Biol 320:55
45. Marques MA, Doss RM, Urbach AR, Dervan PB (2002) Helvetica Chim Acta 85:4485
46. Janssen S, Cuvier O, Muller M, Laemmli UK (2000) Mol Cell 6:1013

47. Baird EE, Dervan PB (1996) J Am Chem Soc 118:6141
48. Wurtz NR, Turner JM, Baird EE, Dervan PB (2001) Org Lett 3:1201
49. Fattori D, Kinzel O, Ingallinella P, Bianchi E, Pessi A (2002) Bioorg Med Chem Lett 12:1143
50. Belitsky JM, Nguyen DH, Wurtz NR, Dervan PB (2002) Bioorg Med Chem Lett 10:2767
51. Foister S, Marques MA, Doss RM, Dervan PB (2003) Bioorg Med Chem Lett 11:4333
52. Minehan TG, Gottwald K, Dervan PB (2000) Helvetica Chim Acta 83:2197
53. Briehn CA, Weyermann P, Dervan PB (2003) Chem Eur J 9:2110
54. Renneberg D, Dervan PB (2003) J Am Chem Soc 125:5707
55. Gottesfeld JM, Neely L, Trauger JW, Baird EE, Dervan PB (1997) Nature 387:202
56. Darnell JE (2002) Nature Rev Cancer 2:740
57. Neely L, Trauger JW, Baird EE, Dervan PB, Gottesfeld JM (1997) J Mol Biol 274:439
58. Dickinson LA, Gulizia RJ, Trauger JW, Baird EE, Mosier DE, Gottesfeld JM, Dervan PB (1998) Proc Natl Acad Sci USA 95:12890
59. Ehley JA, Melander C, Herman D, Baird EE, Ferguson HA, Goodrich JA, Dervan PB, Gottesfeld JM (2002) Mol Cell Biol 22:1723
60. McBryant SJ, Baird EE, Trauger JW, Dervan PB, Gottesfeld JM (1999) J Mol Biol 286:973
61. Winston RL, Ehley JA, Baird EE, Dervan PB, Gottesfeld JM (2000) Biochemistry 39:9092
62. Lenzmeier BA, Baird EE, Dervan PB, Nyborg JK (1999) J Mol Biol 291:731
63. Simon H, Kittler L, Baird E, Dervan P, Zimmer C (2000) Febs Lett 471:173
64. Wurtz NR, Pomerantz JL, Baltimore D, Dervan PB (2002) Biochemistry 41:7604
65. Nguyen DH, Ramm E, Taylor CM, Joung JK, Dervan PB, Pabo CO (2004) Biochemistry 43:3880
66. Yang F, Belitsky JM, Villanueva RA, Dervan PB, Roth MJ (2003) Biochemistry 42:6249
67. Dickinson LA, Trauger JW, Baird EE, Dervano PB, Graves BJ, Gottesfeld JM (1999) J Biol Chem 274:12765
68. Chiang SY, Burli RW, Benz CC, Gawron L, Scott GK, Dervan PB, Beerman TA (2000) J Biol Chem 275:24246
69. Bremer RE, Baird EE, Dervan PB (1998) Chem Biol 5:119
70. Bremer RE, Wurtz NR, Szewczyk JW, Dervan PB (2001) Bioorg Med Chem Lett 9:2093
71. Fechter EJ, Dervan PB (2003) J Am Chem Soc 125:8476
72. Dickinson LA, Trauger JW, Baird EE, Ghazal P, Dervan PB, Gottesfeld JM (1999) Biochemistry 38:10801
73. Coull JJ, He GC, Melander C, Rucker VC, Dervan PB, Margolis DM (2002) J Virol 76:12349
74. Mapp AK, Ansari AZ, Ptashne M, Dervan PB (2000) Proc Natl Acad Sci USA 97:3930
75. Ansari AZ, Mapp AK, Nguyen DH, Dervan PB, Ptashne M (2001) Chem Biol 8:583
76. Arora PS, Ansari AZ, Best TP, Ptashne M, Dervan PB (2002) J Am Chem Soc 124:13067
77. Passner JM, Ryoo HD, Shen LY, Mann RS, Aggarwal AK (1999) Nature 397:714
78. Arndt HD, Hauschild KE, Sullivan DP, Lake K, Dervan PB, Ansari AZ (2003) J Am Chem Soc 125:13322
79. Wang CCC, Dervan PB (2001) J Am Chem Soc 123:8657
80. Hsiang YH, Hertzberg R, Hecht S, Liu LF (1985) J Biol Chem 260:4873
81. Bendixen C, Thomsen B, Alsner J, Westergaard O (1990) Biochemistry 29:5613
82. Chang AY, Dervan PB (2000) J Am Chem Soc 122:4856
83. Wurtz NR, Dervan PB (2000) Chem Biol 7:153
84. Takahashi R, Bando T, Sugiyama H (2003) Bioorg Med Chem Lett 11:2503
85. Oyoshi T, Kawakami W, Narita A, Bando T, Sugiyama H (2003) J Am Chem Soc 125:4752
86. Shinohara K, Narita A, Oyoshi T, Bando T, Teraoka H, Sugiyama H (2004) J Am Chem Soc 126:5113

87. Wang YD, Dziegielewski J, Wurtz NR, Dziegielewska B, Dervan PB, Beerman TA (2003) Nucleic Acids Res 31:1208
88. Dudouet B, Burnett R, Dickinson LA, Wood MR, Melander C, Belitsky JM, Edelson B, Wurtz N, Briehn C, Dervan PB, Gottesfeld JM (2003) Chem Biol 10:859
89. Gottesfeld JM, Melander C, Suto RK, Raviol H, Luger K, Dervan PB (2001) J Mol Biol 309:615
90. Gottesfeld JM, Belitsky JM, Melander C, Dervan PB, Luger K (2002) J Mol Biol 321:249
91. Suto RK, Edayathumangalam RS, White CL, Melander C, Gottesfeld JM, Dervan PB, Luger K (2003) J Mol Biol 326:371
92. Edayathumangalam RS, Weyermann P, Gottesfeld JE, Dervan PB, Luger K (2004) Proc Natl Acad Sci USA 101:6864
93. Belitsky JM, Leslie SJ, Arora PS, Beerman TA, Dervan PB (2002) Bioorg Med Chem Lett 10:3313
94. Crowley KS, Phillion DP, Woodard SS, Schweitzer BA, Singh M, Shabany H, Burnette B, Hippenmeyer P, Heitmeier M, Bashkin JK (2003) Bioorg Med Chem Lett 13:1565
95. Best TP, Edelson BS, Nickols NG, Dervan PB (2003) Proc Natl Acad Sci USA 100:12063
96. Burli RW, Ge YG, White S, Baird EE, Touami SM, Taylor M, Kaizerman JA, Moser HE (2002) Bioorg Med Chem Lett 12:2591
97. Dyatkina NB, Roberts CD, Keicher JD, Dai YQ, Nadherny JP, Zhang WT, Schmitz U, Kongpachith A, Fung K, Novikov AA, Lou L, Velligan M, Khorlin AA, Chen MS (2002) J Med Chem 45:805
98. Lou L, Velligan M, Roberts C, Stevens DA, Clemons KV (2002) Curr Opin Investig Drugs 10:1437
99. Kaizerman JA, Gross ML, Ge YG, White S, Hu WH, Duan JX, Baird EE, Johnson KW, Tanaka RD, Moser HE, Burli RW (2003) J Med Chem 46:3914
100. Marini NJ, Baliga R, Taylor MJ, White S, Simpson P, Tsai L, Baird EE (2003) Chem Biol 10:635
101. Supekova L, Pezacki JP, Su AI, Loweth CJ, Riedl R, Geierstanger B, Schultz PG, Wemmer DE (2002) Chem Biol 9:821
102. Lee DK, Seol W, Kim JS (2003) Curr Topics Med Chem 3:645
103. Praseuth D, Guieysse AL, Helene C (1999) Biochim Biophys Acta 1489:181
104. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, Welchman DP, Zipperlen P, Ahringer J (2003) Nature 421:231
105. Agami R (2002) Curr Opinion Chem Biol 6:829
106. Bentin T, Nielsen PE (2003) J Am Chem Soc 125:6378
107. Nielsen PE, Koppelhus U, Beck F (2004) In: Nielsen PE (ed) Pseudo-peptides in drug discovery. Wiley-VCH, Weinheim, p 153

Top Curr Chem (2005) 253: 33–53
DOI 10.1007/b100441

Structural Selectivity of Drug-Nucleic Acid Interactions Probed by Competition Dialysis

Jonathan B. Chaires (✉)

James Graham Brown Cancer Center, Department of Medicine, Health Sciences Center, University of Louisville, 529 South Jackson Street, Louisville, KY 40202, USA
j.chaires@louisville.edu

Abstract Competition dialysis is a powerful new tool for the discovery of ligands that bind to nucleic acids with structural- or sequence-selectivity. The method is based on firm thermodynamic principles and is simple to implement. In the competition dialysis experiment, an array of nucleic acid structures and sequences is dialyzed against a common test ligand solution. After equilibration, the amount of ligand bound to each structure or sequence is determined spectrophotometrically. Since all structures and sequences are in equilibrium with the same free ligand concentration, the amount bound is directly proportional to the ligand binding affinity. Competition dialysis thus provides a direct and quantitative measure of selectivity, and unambiguously identifies which of the structures or sequences within the sample array are preferred by a particular ligand. Following the introduction of the method, competition dialysis has been used worldwide to probe a variety of ligand-nucleic acid interactions. This chapter will focus on new analytical approaches for extracting information from the database that resulted from the first-generation competition dialysis assay, in which binding data were gathered for the interaction of 126 compounds with 13 different structures and sequences. Such global analysis allows identification of compounds with unique types of binding selectivity.

Keywords DNA · RNA · Intercalation · Groove binding · Dialysis · Thermodynamics · Duplex · Triplex · Quadruplex · Z DNA · i-motif

1 Introduction

Nucleic acids in general, and DNA in particular, are underrepresented pharmaceutical targets. In a recent survey of the biochemical classes of drug targets of current therapies, only 2% of known drugs were targeted toward DNA [1]. The dearth of DNA-targeted therapeutics results, in my opinion, from our ignorance of the design principles needed to produce small molecules that can selectively bind to functionally important DNA sequences or structures, and not because DNA is an inherently unsuitable target. DNA is, in fact, a highly attractive target for new drugs, as embodied by the "antigene" strategy [2–5]. The ability to target and then down-regulate the function of a specific gene has enormous potential in the therapy of genetic-based diseases [2–4]. Direct gene targeting has the advantage that only a small number of copies of the target gene exist in the cell, so it should be possible to inhibit mRNA and protein production and the resultant gene function with fewer molecules of the therapeutic agent. Lower therapeutic doses would lessen the chance of undesirable side effects.

For the last two decades, drug-DNA interaction studies were preoccupied with the characterization of *sequence selectivity* [6]. The advent of footprinting methods in the early 1980s provided an exquisite high-resolution tool for characterizing the preferred drug-binding sites within a particular DNA fragment [7–9]. The primary aim of such studies was to elucidate the rules that govern sequence-selective drug binding to DNA, and to use these rules to develop design principles that would permit synthesis of new compounds that could be targeted to any desired sequence. Sequences that regulate the expression of particular genes could then be selectively targeted by small molecules, allowing an unprecedented ability to control gene expression. This goal was realized by the spectacular success of the Dervan group, who discovered the recognition code for the binding of hairpin polyamides in the DNA minor groove, as reviewed in Chapter 1 in this volume. With the hairpin polyamide technology, it is now possible to target any 8 bp sequence, and the target length will increase with continued development of the technology [10, 11]. The hairpin polyamides can selectively shut-off gene expression both in vitro and in vivo [12]. This class of compounds represents one (but certainly not the only) solution to the DNA sequence recognition problem, and promises to be an exciting new avenue for drug development.

DNA is polymorphic, and adopts a wide variety of secondary and tertiary structures within the genome [13–16]. Localized unique structures surely play a role in gene expression [2, 3, 13, 15, 17]. Using small molecules to target such structures represents a second avenue for drug development, one that is just beginning to be recognized and exploited [2–4, 18–20]. Small molecules can exhibit striking structural selectivity. Early studies from our laboratory [21–24] and from Krugh's laboratory [25–27] showed that intercalators strongly prefer right-handed DNA over left-handed Z-DNA. A consequence of this preference

is that intercalators act as allosteric effectors of DNA conformation, and can exert long-range effects on the conformation of DNA [23, 28–30]. Enantiomeric daunorubicin (WP900) was recently synthesized, and was found to act in a manner completely complementary to daunorubicin [31]. Whereas daunorubicin binds preferentially to right-handed DNA and allosterically converts Z-DNA to a right-handed form, WP900 binds preferentially to left-handed DNA and allosterically converts right-handed DNA to a left-handed form [31]. In addition to recognition of different duplex secondary structures, more recent efforts to target DNA were directed toward multistrand triplex and quadruplex structures [18–20]. Quadruplex DNA, in particular, is thought to be an integral feature of telomeres [19, 20]. The observation that levels of telomerase (the enzyme responsible for telomere DNA replication) are elevated in cancer cells led to concerted attempts to target quadruplex DNA within telomeres as one new avenue of cancer chemotherapy [19, 20]. This is but one of many possible examples of how unique structures of functional importance may present new therapeutic targets.

Attempts to study the structural selectivity of small molecules have been hampered by the lack of a convenient but rigorous assay for structure-selective binding. Investigators typically used laborious binding or thermal denaturation studies to compare ligand interactions with the novel structure to interactions with standard duplex DNA. Such a limited comparison using only two structural forms represents a poor measure of selectivity since the ligand could potentially selectively recognize other conformations that were not studied. To address this problem, we invented a novel competition dialysis assay to evaluate more rigorously structure-selective ligand binding [32]. This chapter will review the competition dialysis assay and its application to a number of nucleic acid recognition problems. New approaches for the global analysis of competition dialysis data will be described. The competition dialysis method promises to enhance studies of structural-selective ligand interactions in the same way that chemical and enzymatic footprinting methods revolutionized studies of sequence-selective binding.

2
Description of the Competition Dialysis Method

2.1
Historical Roots

The competition dialysis assay is based on the fundamental thermodynamic principle of equilibrium dialysis [33]. The approach is easily summarized: A macromolecule solution is placed inside a semi-permeable dialysis membrane with pore sizes that allow small molecules to pass through, but which retain large macromolecules. The dialysis unit is then suspended in a solution containing small ligand molecules. At equilibrium, the chemical potential of the

free ligand must be equal inside and outside the dialysis unit, and any excess ligand on the macromolecule side of the membrane may be attributed to binding to the macromolecule. Equilibrium dialysis has been widely used to measure the binding of small molecules or ions to macromolecules to provide the primary data for the determination of accurate binding constants [33, 34].

Crothers and Mueller first introduced an important variant of the equilibrium dialysis experiment to evaluate the base specificity of drug-DNA interactions, the competition dialysis procedure [35]. The procedure utilized a three-part dialysis chamber, in which two dialysis membranes separated the central chamber from two outside chambers. Two natural, genomic DNA samples of identical concentration but of differing base composition were placed in each of the two outer chambers. A ligand solution was placed in the central chamber and allowed to equilibrate among the three chambers. If the ligand bound selectively to AT or GC base pairs, more ligand would accumulate in the chamber containing the DNA whose base composition contained the higher percentage of the preferred base. Rather detailed quantitative inferences concerning the nature of the preferred binding site are possible by this method using simple probability concepts [35, 36]. This version of the competition dialysis method unfortunately saw little widespread application, and was supplanted by footprinting methods that provide a higher resolution glimpse into sequence-selective binding. Competition dialysis was, however, adapted to provide a tool for probing structural selective ligand binding to nucleic acids. Becker and Dervan used the approach to show selective binding of bis(methidium)spermine to a DNA:RNA hybrid [37]. Chaires and coworkers used the method as a probe for ligand binding to left- and right-handed DNA [22, 28]. More recently, the method was used to probe ligand binding to duplex and triplex DNA [39].

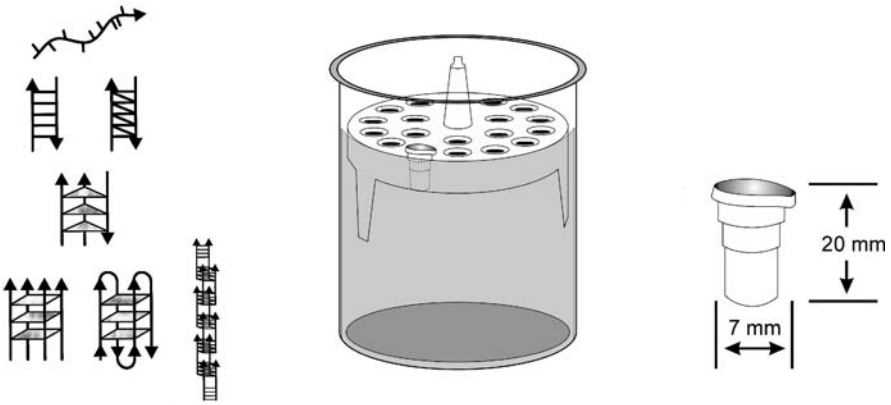


Fig. 1 Schematic of the competition dialysis experiment. A schematic representation of the structures used is shown on the *left*. The *center* shows the simple experimental setup, a beaker containing the test ligand solution and the holder for the dialysis units. A schematic of the microdialysis unit is shown on the *right*

The competition dialysis assay described here [32] is a simple and straightforward extension of Crothers' original method. Instead of a three-chambered dialysis apparatus, disposable dialysis units are used to contain a wide variety of nucleic acid structures at identical concentrations. These units are simply placed into a beaker containing ligand solution, as illustrated in Fig. 1. At equilibrium, the free ligand concentration is identical throughout the system, and preferential binding by a particular structure leads to a greater accumulation of total ligand within that dialysis unit. Structural selectivity can be easily measured by measuring total ligand concentration within each dialysis unit.

2.2 Array of Structures Used in the Competition Dialysis Assay

The first challenge to be met in the competition dialysis assay is to find a suitable buffer in which the structures of interest are all stable. In the first generation assay [32], 13 structures were found to be stable in a simple phosphate buffer containing 200 mM NaCl. The assay was expanded to include 19 structures in the second-generation assay [40], using the same buffer conditions. Table 1 lists the samples used, and Fig. 1 shows the samples in schematic form. The first generation assay used samples 1 through 13. The second-generation assay was expanded to include samples 14–19.

Table 1 Nucleic acid conformation and samples used in competition dialysis experiments

	Conformation	Poly- or oligonucleotide	λ (nm)	ε (M^{-1} cm^{-1})
1	Single-strand purine	Poly dA	257	8,600
2	Single-strand pyrimidine	Poly dT	264	8,520
3	Duplex DNA	*C. perfringens* (31% GC)	260	12,476
4	Duplex DNA	Calf thymus (42% GC)	260	12,824
5	Duplex DNA	*M. lysodeikticus* (72% GC)	260	13,846
6	Duplex DNA	Poly dA: poly dT	260	12,000
7	Duplex DNA	Poly (dAdT)	262	13,200
8	Duplex DNA	Poly (dGdC)	254	16,800
9	DNA-RNA hybrid	Poly rA: poly dT	260	12,460
10	Duplex RNA	Poly A: poly U	260	14,280
11	Z DNA	Br-poly (dGdC)	254	16,060
12	Triplex DNA	Poly dA: $(poly\ dT)_2$	260	17,200
13	Quadruplex DNA 1	$(5'T_2G_{20}T_2)_4$	260	39,267
14	Single strand purine	Poly rA	258	9,800
15	Single-strand pyrimidine	Poly rU	260	9,350
16	Triplex RNA	PolyA:$(polyU)_2$	260	17,840
17	Quadruplex DNA 2	$5'AG_3TTAG_3TTAG_3TTAG_3$	260	73,000
18	Quadruplex DNA 3	$(5'G_{10}T_4G_{10})_4$	260	39,400
19	i-Motif	Poly dC	274	7,400

λ Wavelength, ε molar extinction coefficient.

The samples listed in Table 1 provide an array that includes a wide variety of nucleic acid structures. Single-stranded DNA and RNA are represented. Duplex DNA is represented by both natural DNA samples and synthetic polydeoxynucleotides of defined sequence. These samples cover a range of base composition and simple dinucleotide repeat sequences. Duplex RNA is represented, as is a DNA:RNA hybrid structure. Left-handed Z-DNA represents an extreme secondary structural variant. Multistranded triplex, quadruplex and i-motif structures are represented by several samples.

It should be emphasized that the array of structures listed in Table 1 are but a point of departure. The competition dialysis method is completely general, and arrays of structures of particular interest can be designed as desired. The only limitations are that the structures be large enough to be retained by the dialysis tubing chosen for use, and that they are verified to be stable under the ionic conditions of the experiment. Essential quality control experiments to characterize nucleic acid samples were described in detail [41, 42] and include UV absorbance and circular dichroism spectroscopy, and thermal denaturation studies.

2.3 Experimental Procedures

Once stock solutions of the structural array of interest are prepared, the competition dialysis experiment is simple. Figure 1 shows a schematic of the experiment. Nucleic acid structures are dispensed into individual disposable microdialysis units. Typically, 0.5 mL of each sample is used. All nucleic acids are at identical concentration, typically 75 μM in terms of the monomeric unit of the sample. "Monomeric unit" means nucleotides for single-strands, base pairs for duplexes, base triplets for triplexes, and base quartets for quadruplex and i-motif structures. The monomeric unit was chosen as a concentration standard to negate differences in lengths among the samples. One caution remains: The use of the monomeric unit as a concentration standard does not normalize the concentration of ends. Shorter samples would have a higher effective concentration of ends, a potential problem if a ligand should preferentially bind at such sites.

The sample array is then placed into a beaker containing a solution of the test ligand, typically at 1 μM concentration. The system is then allowed to equilibrate to ensure that dialysis equilibrium is attained, typically 24 h. Nucleic acid samples are then individually removed from their dialysis units, and the detergent sodium dodecyl sulfate (SDS) is added to a final concentration of 1% (w/v). Addition of SDS dissociates bound ligand, ensuring that there are no complexities from differences in the optical properties of free and bound ligand. Bound ligand concentrations are then determined using absorbance measurements or, if necessary or desirable, fluorescence measurements. Any analytical method could be used for concentration determinations, as dictated by the properties of the ligand. Data are then plotted as a simple bar graph, with the amount bound to each of the structures in the array.

The above is a simple overview of the experimental procedure for the competition dialysis assay, and is not intended to be a detailed protocol. Step-by-step descriptions of sample preparation and characterization and the protocol for the competition dialysis experiment have appeared elsewhere [41, 42], and should be consulted by those wishing to implement the method.

2.4 Examples of Experimental Results

Figure 2 shows examples of results obtained by the competition dialysis method using the first-generation array of structures. Examples of compounds with low, moderate, and high selectivity are shown. MMQ1 (Fig. 2A) is a dibenzophenanthroline derivative that effectively inhibits telomerase [43]. The competition dialysis assay reveals it to be rather promiscuous, with comparatively low selectivity. Adriamycin (Fig. 2B) is an anthracycline antibiotic widely used in cancer chemotherapy [44]. The competition dialysis assay reveals a strong preference for binding to right-handed, B-form duplex structures, and a clear preference for GC-rich duplexes. There is no binding to single-stranded forms,

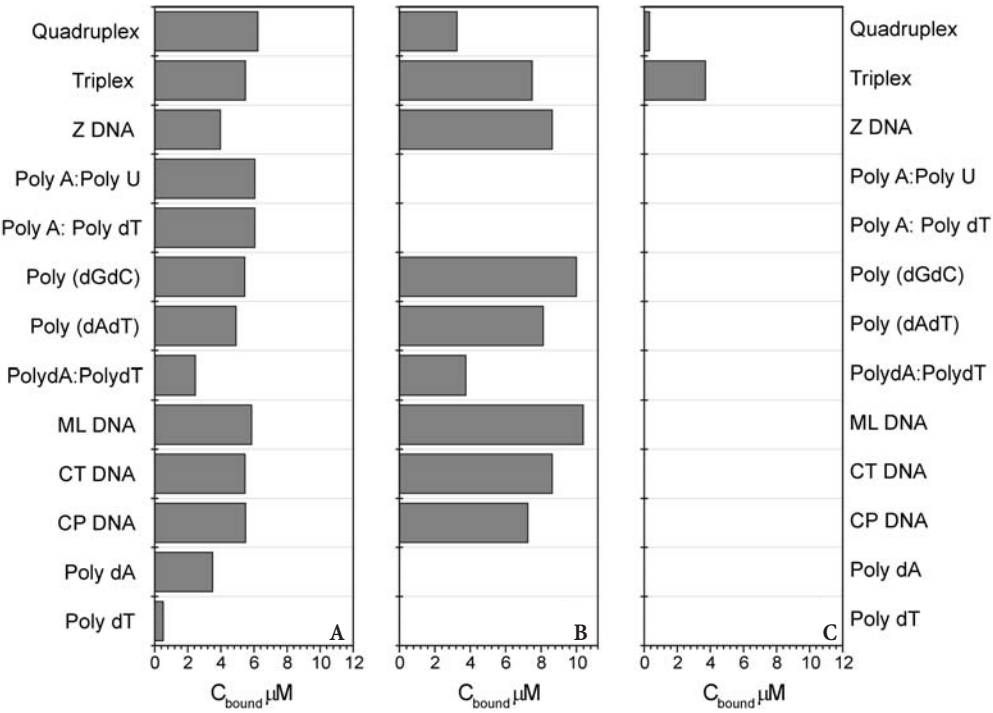


Fig. 2A–C Examples of competition dialysis data. The amount bound to each of the 13 different structures is shown. **A** MMQ1, a dibenzophenanthroline [43]. **B** Adriamycin, an anthracycline antibiotic. **C** DODC, a cyanine dye [45]

RNA, or to a DNA:RNA hybrid. Adriamycin binds to multistranded triplex and quadruplex structures. Its apparent binding to left-handed Z DNA is illusory, and results from the allosteric conversion of left-handed DNA to an intercalated, right-handed form [32]. DODC (3,3′ diethyloxacarbocyanine; Fig. 2C) was selected by the DOCK algorithm to be a quadruplex-selective binding agent [45]. The competition dialysis assay confirms that DODC is a highly selective binding agent, but shows unambiguously that it binds selectively to a triplex structure over any other form [40]. The data of Fig. 2 are intended to show the range of structural selectivity that can be visualized by the competition dialysis data. A more detailed discussion of the quantitative interpretation of such data will follow.

One should note in Fig. 2 the differences in the amounts bound among the three compounds. Each panel is scaled the same. The amounts bound are directly proportional to the relative binding affinities of each compound for the various nucleic acid structures. Thus, while DODC is highly selective, its binding affinity is comparatively low. Adriamycin, in contrast, has comparatively high affinity for the various duplex forms to which it binds.

Using the first generation competition dialysis assay that used 13 nucleic acid structures and sequences, 126 different compounds were studied. These compounds were obtained from commercial sources or were synthesized and provided by research laboratories interested in the design of structural-selec-

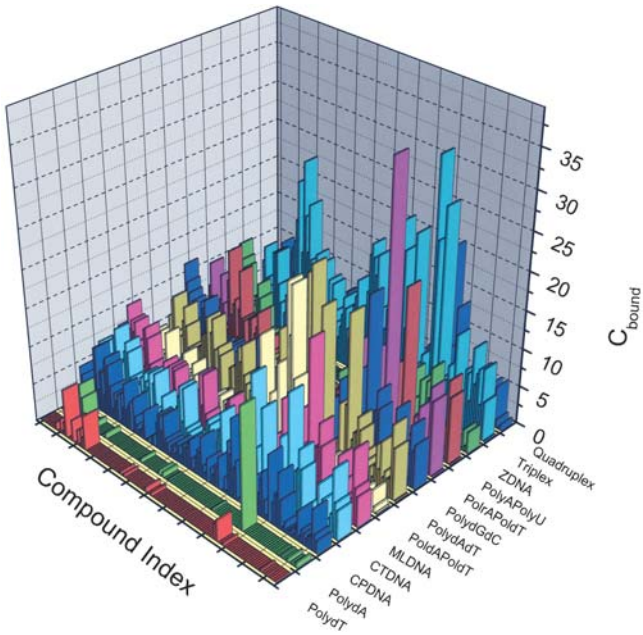


Fig. 3 Summary of competition dialysis results from the first generation assay, with 126 compounds binding to 13 nucleic acid structures and sequences

tive binding agents. (Space limitations prevent illustration of the structures of all of these compounds. The structures may be found as supplementary materials to published work [46]). Figure 3 shows the results obtained for all of these compounds. Figure 3 represents a kind of structural-selectivity "skyline", and contains a wealth of information. The data in Fig. 3 summarize results from 13×126=1,638 binding interactions. How to extract information from such a database is a non-trivial bioinformatics problem. Initial attempts to develop tools for the interpretation and analysis of the data shown in Fig. 3 will be described in a later section.

2.5 Published Applications of Competition Dialysis

The competition dialysis assay was introduced in 1999 in a publication [32] that presented data using 15 compounds and 13 nucleic acid structures and sequences. The 15 compounds were carefully selected to represent standard, well-characterized DNA binding agents. Among these were intercalators, known groove-binders, putative triplex-selective binders, and putative quadruplex-selective binders. Berenil, chromomycin and 1-pyrenemethylamine were also studied, as compounds whose binding properties were less well characterized. The competition dialysis method was fully described in this initial publication, and details of the biophysical properties of the nucleic acid samples used in the array were described in an extensive supporting information section. Detailed, step-by-step protocols for the competition dialysis method, including sample preparation, were subsequently published [41, 42]. Subsequent use of the assay revealed several novel types of structural selectivity [40, 46–53]. A few highlights follow.

The carbocyanine dye 3,3′-diethyloxadicarbocyanine (DODC) was selected by a structure-based drug discovery program as a quadruplex-selective binding agent [45]. The competition dialysis assay revealed, however, that DODC in fact binds preferentially to a triplex structure [40]. This finding emphasizes that it is imperative to test ligands against the widest possible array of structures lest false positives emerge. Further, the need for integration of computational and experimental approaches as a routine part of the drug discovery process is readily apparent following that experience.

The competition dialysis assay was used to discover a novel class of compounds that selectively recognize a DNA:RNA hybrid structure, a structure of profound biological importance [46]. Secondary assays revealed that these compounds were effective inhibitors of RNase H, an enzyme that is of great interest as a therapeutic target. Competition dialysis thus is useful in lead discovery.

Competition dialysis also proved to be useful for the quantitative study of compounds that bind to multistranded structures. A variety of compounds with selectivity for triplexes [47, 49] or quadruplexes [48, 51, 53] were studied by the method. One study is of particular note [49]. A series of triplex-selective

naphthylquinolines were studied by competition dialysis. The results were used to construct a quantitative structure affinity relationship (QSAR) that revealed some of the molecular determinants of triplex binding free energy. The combined approach proved to be a powerful new tool for the discovery of design principles to guide the synthesis of new compounds with enhanced structural selectivity.

Several laboratories have adopted and adapted the competition dialysis assay for their particular interests. The Mergny laboratory created a more extensive array of triplex and quadruplex structures for use in their pursuit of small molecules that might selectively recognize clinically important four-stranded structures within the telomere [54–60]. The Arya laboratory has made extensive use of competition dialysis in their studies of aminoglycoside-nucleic acid interactions [61, 62]. Portugal and coworkers used competition dialysis to infer an unusual cytosine–cytosine site-selectivity for cryptolepine, an antimalarial intercalator [63]. Competition dialysis was used in the discovery of a polycyclic acridine with selectivity for G-quadruplexes [64]. These studies reveal the versatility of competition dialysis, and its applicability to a number of nucleic acid recognition problems.

3
Global Analysis of the Data from the First Generation Assay

The first generation competition dialysis assay used 13 structures and sequences (Table 1). A total of 126 compounds from a variety of chemical classes was studied. Results from these studies are compiled in Fig. 3. Each row of the three-dimensional plot shown in Fig. 3 corresponds to data obtained for one compound, as, for example, shown in Fig. 2. The columns in Fig. 3 show binding of all 126 compounds to one particular structure. Figure 3 contains a wealth of information, but extracting that information requires new analytical tools. Some of those tools will be described here.

3.1
Trends from Tukey Box Plots

The global trends in ligand binding to the different structures are of fundamental interest. For the 126 compounds studied, what is the average amount bound to each structure? What are the differences in binding to each structure? Do ligands bind equally well to all structures in the array? Tukey box plots are an ideal statistical tool for addressing these questions. A Tukey box plot [65] is a graphical tool that concisely summarizes the properties of distributions. In a box plot, the horizontal line within the box marks the 50th percentile, while the square symbol marks the arithmetic mean. The top edge of the box marks the 75th percentile, the lower edge the 25th percentile. The vertical lines extending from the box mark the 90th (top) and 10th (bottom) percentiles.

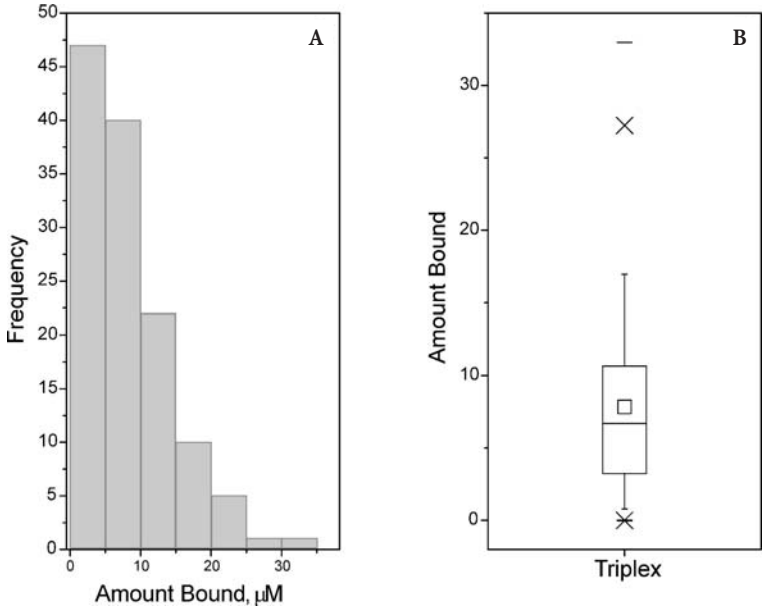


Fig. 4A,B Illustration of the Tukey box plot [65]. *Panel A* shows the distribution of the amount bound to a DNA triplex for the 126 compounds studied in the first-generation competition dialysis data. *Panel B* shows the representation of that distribution as a Tukey box plot

Crosses mark the 1st and 99th percentiles, and bars mark the complete range of the data. Figure 4 provides an example. Figure 4A shows the actual distribution of the amount bound for the 126 compounds to the triplex poly dA-(poly dT)$_2$. Figure 4B shows the Tukey box plot of the same data.

Figure 5A shows box plots for all of the structures used in the first generation assay. A few generalizations are immediately evident. The compounds studied bind poorly, if at all, to single-stranded DNA forms, to left-handed Z DNA, and to RNA. In contrast, triplex DNA has the highest average amount bound, but also has the widest distribution of binding. In part this results from the fact that many of the 126 compounds studied were solicited from laboratories actively engaged in the design and synthesis of compounds targeted toward multistranded triplex and quadruplex forms. Little can be said about the data for the remaining structures. The average amounts bound and the widths of the distributions are similar for all remaining cases. The power of the box plot representation is that it concisely defines the global, average binding properties for the compound and structural arrays. Individual compounds that show statistically significant deviations from these average binding properties are those that display structural selectivity.

Box plots reveal their utility for discerning trends when subsets of the 126 compounds are examined. Figure 5 shows, in addition to the data for all compounds, data for selected groove binders (Fig. 5B) and intercalators (Fig. 5C).

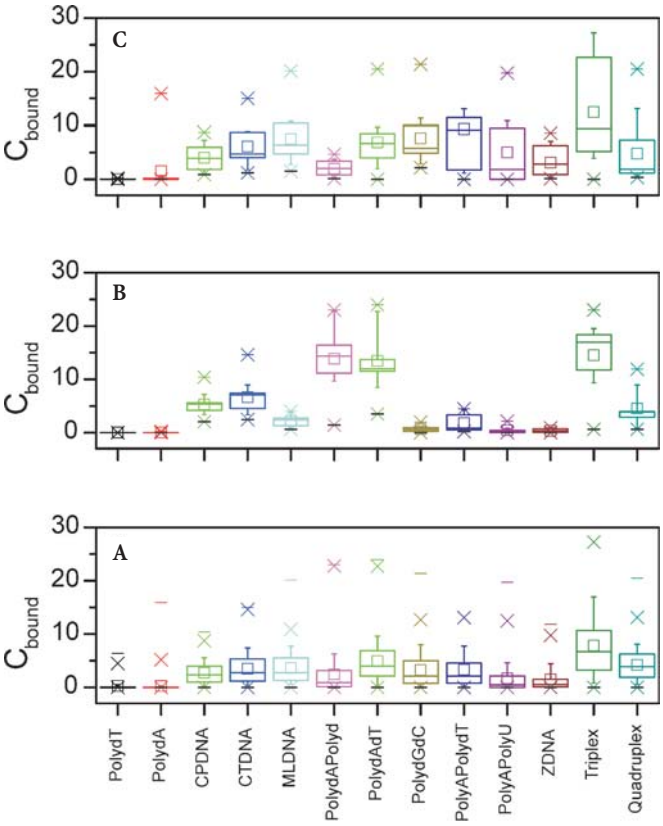


Fig. 5A–C Box plots for all the first-generation binding data (**A**), selected groove-binders (**B**), and selected intercalators (**C**)

Data for a subset of proven groove binders were selected and box plots constructed (Fig. 5B). The groove binders include: berenil, distamycin, netropsin, DAPI, DB075, DB351, DB181, and methyl green. (The DB compounds are diphenyl furans from the Wilson-Boykin laboratories at Georgia State University.) Data for intercalators (Fig. 5C) include Actinomycin D, ethidium bromide, propidium iodide, proflavin, 9-aminoacridine, quinacrine, daunorubicin, doxorubicin, coralyne, thiazole orange, and ellipticine. The patterns observed for groove binders and intercalators clearly differ from one another, and from that obtained for all compounds.

Groove binders (Fig. 5B) show a marked preference for AT-containing polynucleotides. Little binding is observed for all other structures, apart from the AT-rich natural DNAs from *C. perfringens* (32% GC) and calf thymus (42% GC). There is apparently strong binding to the triplex poly dA-$(\text{poly dT})_2$, but we previously showed that such binding is illusory, and results from displacement of the third strand with tight binding to the remaining duplex [32].

Intercalators (Fig. 5C) bind more promiscuously than do groove binders, but nonetheless show a distinctive pattern compared to the entire 126 compound database. Intercalators show little binding to single-stranded forms, with the notable exception of one outlier that is coralyne. (The unusual binding of coralyne to poly dA, discovered by competition dialysis [32], turned out to be a fascinating study in its own right. Hud and coworkers [66, 67], stimulated by our competition dialysis results, found that coralyne can facilitate duplex disproportionation into triplex and single-stranded forms.) Intercalators show a trend for preferential binding to GC-rich DNA and polynucleotides. Intercalators bind poorly to poly dA-poly dT, but well to the triplex poly dA-$(\text{poly dT})_2$. In these cases, in contrast to groove binders, the triplex was shown to remain intact, and intercalators are truly binding to the triple helix.

The global patterns of binding clearly differ for intercalators and groove binders. In each case, the pattern of binding is fully consistent with the known properties of the binding mode. These results suggest that competition dialysis studies on new compounds with unknown binding properties might offer preliminary (but tentative) insights into their binding mode. Competition dialysis would certainly never displace the more reliable hydrodynamic or unwinding assays [68] for the determination of binding mode, but might offer less laborious means of obtaining a clue about the type of binding until more reliable data are available.

3.2 New Metrics: the Specificity Sum (SS) and SS/C_{max}

The data shown in Fig. 3 contain a wealth of information about ligand–nucleic interactions. In order to identify those compounds with high binding selectivity, new metrics are needed. One simple metric we developed for this purpose is the specificity sum, SS. To calculate SS for a given compound, data are first normalized with respect to the maximal amount bound (C_{max}) to any of the structures in the assay. For example, the data in Fig. 2 may be normalized to yield the plots shown in Fig. 6. Data were normalized using C_{max} values of 6.24 for MMQ1, 10.38 for Adriamycin, and 3.71 for DODC. These correspond to the amounts bound to quadruplex DNA for MMQ1, *M. lysodeikticus* DNA for Adriamycin, and triplex DNA for DODC. Once data are normalized, SS is simply the sum of the normalized amounts bound to each nucleic acid species *i* in the assay:

$$\mathrm{SS} = \sum_i \frac{\mathrm{C}_{b,i}}{\mathrm{C}_{\max}} \tag{1}$$

where $C_{b,i}$ is the amount bound and C_{max} is the maximum amount bound to any species. The index *i* ranges from 1 to 13 in the current version of the assay, corresponding to the 13 different nucleic acid structures and sequences used.

SS can thus range from 1 to 13, with a value of 1 indicative of absolute selectivity where the compound binds to only one structure. A value of SS=13,

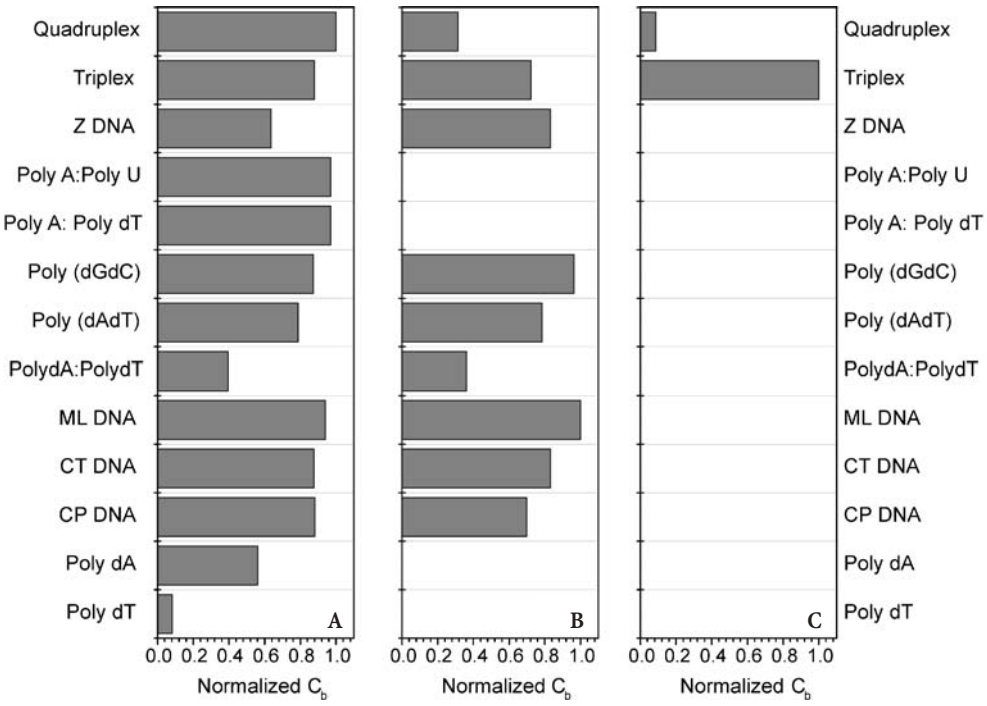


Fig. 6 Normalized competition dialysis data. The data from Fig. 2 were normalized as described in the text to obtain these plots, from which the specificity sum may be calculated

in contrast, would indicate equal binding to all structures in the assay, and a complete lack of selectivity. For the three compounds shown in Fig. 6, SS values of 9.85, 6.51, and 1.08 were calculated for MMQ1, Adriamycin, and DODC, respectively. DODC clearly binds with the greatest selectivity, and MMQ1 the least.

One limitation of the SS metric is that it does not contain information about compound binding affinity. This limitation may be circumvented by calculating the ratio C_{max}/SS where $C_{max}$is the maximal amount bound as defined above. This ratio embodies both affinity and selectivity. C_{max} directly measures compound affinity. If SS=1, the maximal value of C_{max}/SS will be obtained, whereas if SS=13 (no selectivity), the minimal value of the ratio will result. For MMQ1, Adriamycin, and DODC, the values of C_{max}/SS are calculated to be 0.63, 1.59, and 3.43, respectively.

SS and C_{max}/SS values for all 126 compounds studied in the first generation assay are shown in Fig. 7. A wide range of values is seen for both SS and C_{max}/SS, reflecting a considerable diversity in selectivity and affinity. A few compounds approach an SS value of 1, indicative of binding to only a single structure. Figure 8 shows the distributions of SS and C_{max}/SS values. SS values are fairly normally distributed, and range from 1.0 to 9.85. The mean SS value is 4.5±2.0.

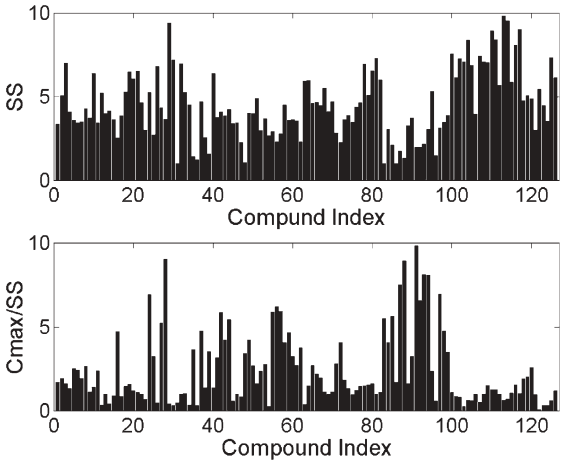


Fig. 7 Specificity sums SS (*top*) and the ratios C_{max}/SS (*bottom*) for all compounds studied

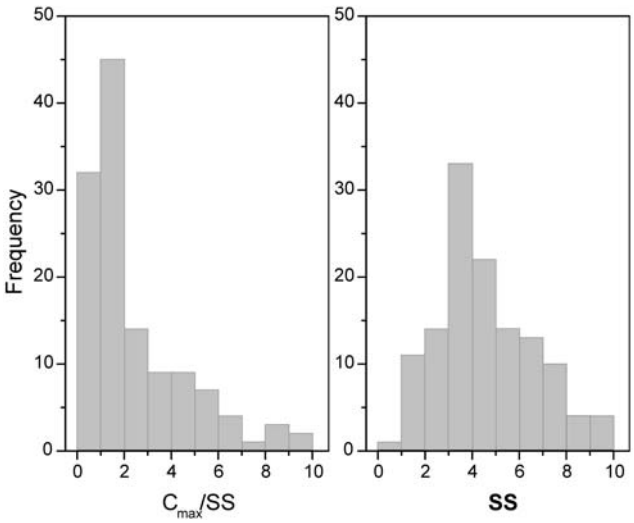


Fig. 8 Distributions of SS and C_{max}/SS values

The distribution of C_{max}/SS values is highly skewed. C_{max}/SS values range from 0.06 to 9.8, with a mean of 2.4±2.2. The significance of SS and C_{max}/SS lies in the fact that their values allow compounds with high selectivity and affinity to be identified quantitatively and without bias.

By using the SS metric, the four compounds with the greatest selectivity from the 126 studied were identified. These compounds are shown in Fig. 9. All compounds have SS values at or near 1.0, indicating that they essentially bind to a single structure. The porphyrin NMM [69] binds solely to the parallel-

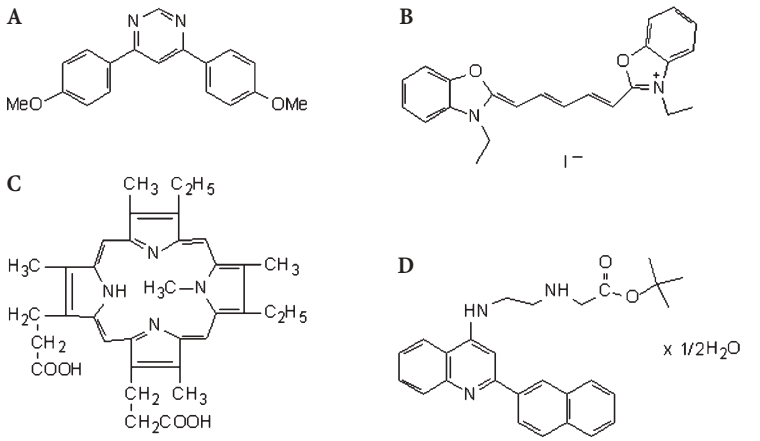


Fig. 9A–D The most selective compounds as judged by SS values near 1. **A** PM008, a biarylpyrimidine [53]. **B** DODC, 3,3′-diethyloxadicarbocyanine [40, 45]. **C** NMM, *N*-methyl mesoporphyrin IX [32, 69]. **D** MHQ14, a naphthylquinoline [49]

stranded G quadruplex $(5'T_2G_{20}T_2)_4$. All of the other compounds bind selectively to the triplex poly dA-(poly dT)$_2$. While all of the compounds in Fig. 9 bind with high selectivity, their binding affinity is generally low, with C_{max} values ranging from 0.5 to 5.6 μM.

Figure 10 shows the top four compounds with combined selectivity and high affinity as judged by the C_{max}/SS metric. For these compounds, SS values range from 1.3 to 3.6, and C_{max}/SS ratios range from 8.1 to 9.8. All of these compounds

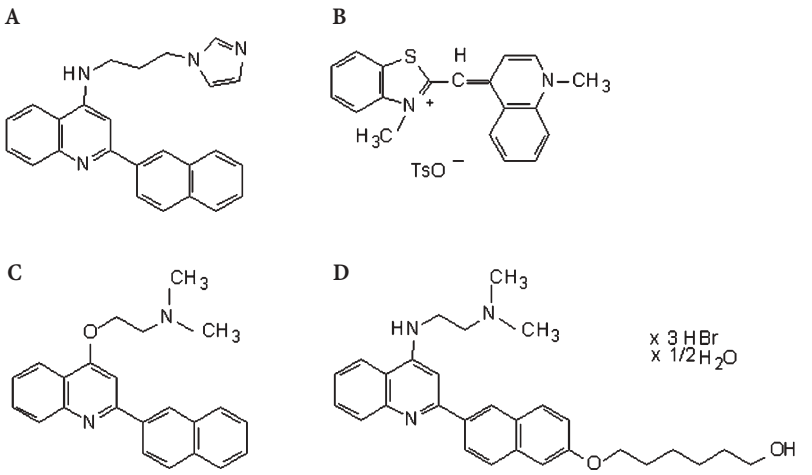


Fig. 10A–D Compounds with the highest combined selectivity and affinity as judged by values of the ratio C_{max}/SS**.** Compounds **A**, **C**, and **D** are naphthylquinolines from the Strekowski laboratory [49]. Compound **B** is thiazole orange

bind selectively to the triplex poly dA-(poly dT)$_2$. Three of the four compounds are naphthylquinolines synthesized by the Strekowski laboratory at Georgia State University as part of a rational design effort to target triplex DNA. These efforts succeeded, and the remarkably selective binding properties of a series of naphthylquinolines (including those in Fig. 10) were fully described in a recent publication [49].

The simple SS and C_{max}/SS metrics are of great use for extracting compounds of interest from the database gathered using the first generation assay. These metrics are of general utility, but their values will clearly differ for different versions of the assay. Addition of more structures will increase the range of SS values. This limits the use of these metrics to comparisons within the same assay. It is inappropriate to compare SS and C_{max}/SS values for different generations of the competition dialysis assay that contain different numbers of structures.

4 Quantitative Analysis of Competition Dialysis Data: Validation

Competition dialysis data may be used to calculate apparent binding constants and binding free energies [42, 49]. Apparent binding constants for each structure or sequence, K_{app}, may be calculated from competition dialysis data such as shown in Figs. 2 and 3 [42, 49]. The simple relationship is $K_{app}=C_b/(C_f)(S_{total}-C_b)$, where C_b is the amount of ligand bound, C_f is the free ligand concentration and S_{total} is the total nucleic acid concentration. By virtue of the experimental design used in the competition dialysis experiment, C_f=1 μM and S_{total}=75 μM (expressed in terms of the monomeric unit of the nucleic acid, i.e., nucleotides, base pairs, triplets, or tetrads). Binding free energies may be calculated by using the standard relationship $\Delta G_{app}=-RT \ln K_{app}$.

Binding constants calculated in this manner are of great utility, but should be used with caution since they are derived from a single set of reactant concentrations. They essentially would correspond to a single point on a more complete binding isotherm. The primary advantage is that apparent binding constants may be quickly obtained. On several occasions [32, 40, 46, 47, 49, 52], we have validated apparent binding constants obtained by competition dialysis by more laborious spectrophotometric titration studies in which complete binding isotherms were obtained and rigorously fitted to appropriate models using nonlinear least squares fitting procedures [70–72]. Figure 11 shows a comparison of the more rigorously determined binding constants with apparent binding constants obtained quickly by competition dialysis. The data of Fig. 11 serve to validate apparent binding constants obtained by competition dialysis. For the 10 cases in which complete binding isotherms were determined to provide more rigorous binding constants, these were highly correlated with the apparent binding constants. The correlation coefficient was found to be 0.85, which is significant at the level of P=0.001.

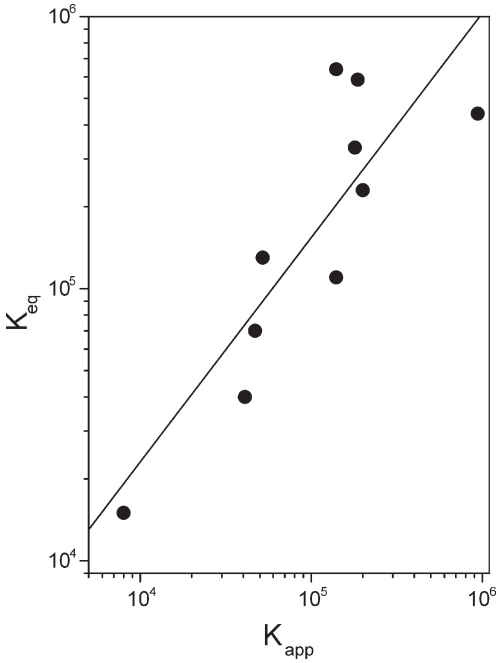


Fig. 11 Validation of apparent binding constants determined by competition dialysis. Apparent binding constants (K_{app}) are plotted against binding constants (K_{eq}) obtained by spectrophotometric or spectrofluorometric titrations. The line is a linear least squares fit, yielding a linear correlation coefficient of 0.853, with P=0.0017

Apparent binding constants obtained by competition dialysis are thus valid and useful for rapid, quantitative screening. For serious, rigorous insights into the thermodynamic basis of compound selectivity, however, more laborious spectrophotometric [70–72] and calorimetric [73, 74] titration procedures remain essential.

5 Summary and Outlook

Competition dialysis has proved to be a powerful new tool for the study of the molecular recognition of nucleic acids. The assay is simple to implement, is thermodynamically sound, and provides quantitative data that can unambiguously identify structure-selective binding interactions. The targeting of unusual nucleic acid secondary or multistranded structures is at the frontier of drug discovery efforts. Competition dialysis promises to greatly enhance efforts to identify small molecules capable of binding selectively to particular structures that are involved in the control of gene expression.

The outlook for the continued development and refinement of the competition dialysis assay is bright. The assay is versatile, and could be developed along a number of lines. As new nucleic structures of functional importance are identified, it will be easy to add them to expanded versions of the competition dialysis assay. The assay can be expanded to accommodate any desired number of structures in the nucleic acid array. Of perhaps more importance and urgency, the competition dialysis assay is fully amenable to improvements that might make it a truly high-throughput screening assay. Transporting the assay to a 96-well plate format, where it might be used in combination with robotic systems, should be possible. Such efforts are under way in this laboratory.

Acknowledgements Dr. Jinsong Ren worked diligently to develop and refine the competition dialysis assay, and her efforts are gratefully acknowledged. Research in the Chaires laboratory is supported by grant CA35635 from the National Cancer Institute. Jonathan B. Chaires holds the James Graham Brown Endowed Chair in Biophysics, and he is grateful to the James Graham Brown Foundation for their support.

References

1. Drews J (2000) Science 287:1960
2. Hurley LH, Boyd FL (1988) Trends Pharmacol Sci 9:402
3. Hurley LH (1989) J Med Chem 32:2027
4. Hurley LH (2002) Nat Rev Cancer 2:188
5. Thurston DE (1999) Br J Cancer 80 Suppl 1:65
6. Dervan PB (1986) Science 232:464
7. Fox KR, Waring MJ (1984) Nucleic Acids Res 12:9271
8. Lane MJ, Dabrowiak JC, Vournakis JN (1983) Proc Natl Acad Sci USA 80:3260
9. Van Dyke MW, Hertzberg RP, Dervan PB (1982) Proc Natl Acad Sci USA 79:5470
10. Dervan PB, Burli RW (1999) Curr Opin Chem Biol 3:688
11. Dervan PB (2001) Bioorg Med Chem 9:2215
12. Gottesfeld JM, Neely L, Trauger JW, Baird EE, Dervan PB (1997) Nature 387:202
13. Sinden RR (1994) DNA Structure and function. Academic, San Diego
14. Neidle S (1999) Oxford handbook of nucleic acid structure. Oxford University Press, New York
15. van Holde K, Zlatanova J (1994) Bioessays 16:59
16. Rich A (1993) Gene 135:99
17. Belmont P, Constant J-F, Demeunynck M (2001) Chem Soc Rev 30:70
18. Mergny JL, Duval-Valentin G, Nguyen CH, Perrouault L, Faucon B, Rougee M, Montenay-Garestier T, Bisagni E, Helene C (1992) Science 256:1681
19. Mergny JL, Helene C (1998) Nature Med 4:1366
20. Jenkins TC (2000) Curr Med Chem 7:99
21. Chaires JB (1983) Nucleic Acids Res 11:8485
22. Chaires JB (1985) Biochemistry 24:7479
23. Chaires JB (1986) J Biol Chem 261:8899
24. Chaires JB (1986) Biochemistry 25:8436
25. Walker GT, Stone MP, Krugh TR (1985) Biochemistry 24:7462
26. Walker GT, Stone MP, Krugh TR (1985) Biochemistry 24:7471

27. Walker GT (1986) The interactions of drugs with left-handed (Z) and right-handed (B) DNA. PhD thesis, University of Rochester
28. Pohl FM, Jovin TM, Baehr W, Holbrook JJ (1972) Proc Natl Acad Sci USA 69:3805
29. Crothers DM, Fried M (1983) Cold Spring Harb Symp Quant Biol 47:263
30. Schurr JM, Delrow JJ, Fujimoto BS, Benight AS (1997) Biopolymers 44:283
31. Qu X, Trent JO, Fokt I, Priebe W, Chaires JB (2000) Proc Natl Acad Sci USA 97:12032
32. Ren J, Chaires JB (1999) Biochemistry 38:16067
33. van Holde KE, Johnson WC, Ho PS (1998) Principles of physical biochemistry. Prentice Hall, Upper Saddle River, NJ
34. Klotz IM (1997) Ligand receptor energetics. Wiley, New York
35. Muller W, Crothers DM (1975) Eur J Biochem 54:267
36. Chaires JB (1992) In: Hurley LH (ed) Advances in DNA sequence specific agents, vol 1. JAI Press, Greenwich, CT, p 3
37. Becker MM, Dervan PB (1979) J Am Chem Soc 101:3664
38. Satyanarayana S, Dabrowiak JC, Chaires JB (1993) Biochemistry 32:2573
39. Haq I, Ladbury JE, Chowdhry BZ, Jenkins TC (1996) J Am Chem Soc 118:10693
40. Ren J, Chaires JB (2000) J Am Chem Soc 122:424
41. Chaires JB (2002) In: Beaucage SL, Bergstrom DE, Glick GD, Jones RA (eds) Current protocols in nucleic acid chemistry, vol 1. Wiley, New York, p 831
42. Ren J, Chaires JB (2001) Methods Enzymol 340:99
43. Mergny JL, Lacroix L, Teulade-Fichou MP, Hounsou C, Guittat L, Hoarau M, Arimondo PB, Vigneron JP, Lehn JM, Riou JF, Garestier T, Helene C (2001) Proc Natl Acad Sci USA 98:3062
44. Arcamone F (1981) Doxorubicin anticancer antibiotics. Academic, New York
45. Chen Q, Kuntz ID, Shafer RH (1996) Proc Natl Acad Sci USA 93:2635
46. Ren J, Qu X, Dattagupta N, Chaires JB (2001) J Am Chem Soc 123:6742
47. Ren J, Bailly C, Chaires JB (2000) FEBS Lett 470:355
48. Carrasco C, Rosu F, Gabelica V, Houssier C, De Pauw E, Garbay-Jaureguiberry C, Roques B, Wilson WD, Chaires JB, Waring MJ, Bailly C (2002) Chembiochem 3:1235
49. Chaires JB, Ren J, Henary M, Zegrocka O, Bishop GR, Strekowski L (2003) J Am Chem Soc 125:7272
50. Murr MM, Harting MT, Guelev V, Ren J, Chaires JB, Iverson BL (2001) Bioorg Med Chem 9:1141
51. Alberti P, Ren J, Teulade-Fichou MP, Guittat L, Riou JF, Chaires J, Helene C, Vigneron JP, Lehn JM, Mergny JL (2001) J Biomol Struct Dyn 19:505
52. Reynisson J, Schuster GB, Howerton SB, Williams LD, Barnett RN, Cleveland CL, Landman U, Harrit N, Chaires JB (2003) J Am Chem Soc 125:2072
53. Murphy PM, Phillips VA, Jennings SA, Garbett NC, Chaires JB, Jenkins TC, Wheelhouse RT (2003) Chem Commun (Camb) 10:1160
54. Alberti P, Ren J, Teulade-Fichou MP, Guittat L, Riou JF, Chaires J, Helene C, Vigneron JP, Lehn JM, Mergny JL (2001) J Biomol Struct Dyn 19:505
55. Alberti P, Schmitt P, Nguyen CH, Rivalle C, Hoarau M, Grierson DS, Mergny JL (2002) Bioorg Med Chem Lett 12:1071
56. Teulade-Fichou MP, Carrasco C, Guittat L, Bailly C, Alberti P, Mergny JL, David A, Lehn JM, Wilson WD (2003) J Am Chem Soc 125:4732
57. Teulade-Fichou MP, Hounsou C, Guittat L, Mergny JL, Alberti P, Carrasco C, Bailly C, Lehn JM, Wilson WD (2003) Nucleosides Nucleotides Nucleic Acids 22:1483
58. Rosu F, De Pauw E, Guittat L, Alberti P, Lacroix L, Mailliet P, Riou JF, Mergny JL (2003) Biochemistry 42:10361
59. Guittat L, Alberti P, Rosu F, Van Miert S, Thetiot E, Pieters L, Gabelica V, De Pauw E, Ottaviani A, Riou JF, Mergny JL (2003) Biochimie 85:535

60. Alberti P, Hoarau M, Guittat L, Takasugi M, Arimondo PB, Lacroix L, Mills M, Teulade-Fichou M-P, Vigneron J-P, Lehn J-P, Maillet P, Mergny J-L (2003) In: Demeunynck M, Bailly C, Wilson WD (eds) Small molecule DNA and RNA binders:from small molecules to drugs, vol 1. Wiley-VCH, Darmstadt, p 315
61. Arya DP, Xue L, Tennant P (2003) J Amn Chem Soc 125:8070
62. Arya DP (2003) J Am Chem Soc 125:10148
63. Lisgarten JN, Coll M, Portugal J, Wright CW, Aymami J (2002) Nat Struct Biol 9:57
64. Heald RA, Modi C, Cookson JC, Hutchinson I, Laughton CA, Gowan SM, Kelland LR, Stevens MF (2002) J Med Chem 45:590
65. Cleveland WS (1985) The elements of graphing data. Wadsworth Advanced Books and Software, Monterey, CA
66. Jain SS, Polak M, Hud NV (2003) Nucleic Acids Res 31:4608
67. Polak M, Hud NV (2002) Nucleic Acids Res 30:983
68. Suh D, Chaires JB (1995) Bioorg Med Chem 3:723
69. Arthanari H, Basu S, Kawano TL, Bolton PH (1998) Nucleic Acids Res 26:3724
70. Chaires JB (2003) In: Demeunynck M, Bailly C, Wilson WD (eds) DNA and RNA Binders, vol 2. Wiley-VCH, Weinheim, p 461
71. Chaires JB (2001) In: Chaires JB, Waring MJ (eds) Drug-nucleic acid interactions. Methods in Enzymology, vol 340. Academic, San Diego, p 3
72. Qu X, Chaires JB (2000) In: Johnson ML, Brand L (eds) Methods in Enzymology, vol 321. Academic, San Diego, p 353
73. Haq I, Jenkins TC, Chowdhry BZ, Ren J, Chaires JB (2000) Methods in Enzymology, vol 323. Academic, San Diego, pp 373
74. Ren J, Jenkins TC, Chaires JB (2000) Biochemistry 39: 8439

Top Curr Chem (2005) 253: 55–76
DOI 10.1007/b100442

Cyanine Dye–DNA Interactions: Intercalation, Groove Binding, and Aggregation

Bruce A. Armitage (✉)

Department of Chemistry, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213–3890, USA
army@andrew.cmu.edu

Abstract The cyanine dyes are among the oldest classes of synthetic compounds but continue to find applications in a variety of fields. In many cases, the dyes act as fluorescent labels for biomolecules, and can interact with the biomolecule either through covalent or noncovalent bonding. One particularly important application involves dyes that bind to double helical DNA by intercalation and exhibit large fluorescence enhancements upon binding. This chapter describes recent investigations of the noncovalent binding modes by which cyanine dyes recognize DNA, with special emphasis placed on relationships between the dye structure and DNA binding mode. In addition to simple binding of the dye as a monomer, several dyes form well-defined helical aggregates using DNA as a template. While these dyes are typically achiral, the aggregates exhibit induced chirality due to the right-handed helical structure of the underlying DNA template. Spectroscopic methods for characterizing these supramolecular assemblies as well as the monomeric complexes are described.

Keywords Cyanine dyes · DNA · Intercalators · Minor groove binders · Aggregates

Abbreviations

CD	Circular dichroism
DNA	Deoxyribonucleic acid
LD	Linear dichroism
PNA	Peptide nucleic acid
UV-vis	Ultraviolet-visible

1 Introduction

The history of cyanine dyes spans nearly 150 years, beginning with the first reported synthesis of a blue solid by Williams in 1856 [1]. They have found widespread use as sensitizers for color photography, as fluorescent labels for biomolecules such as proteins and nucleic acids, and as environmentally sensitive probes for reporting on local properties such as viscosity or polarity [2]. Synthetic methods have been developed that allow preparation of cyanines that have absorption and fluorescence spectra extending from 400 to 1000 nm [3], solubility in a wide range of organic and aqueous solvents, and reactive groups that allow covalent linkage to other molecules [4].

1.1 Cyanine Dye Structural Features

Cyanines have the general structure shown in Fig. 1. The conjugated system is varied in increments of two carbons, based on the length of the polymethine bridge separating the two nitrogen atoms. Because this bridge has an odd number of carbons, the positive charge is delocalized by resonance over both nitrogens. Formally, the terms "cyanine", "carbocyanine", "dicarbocyanine",

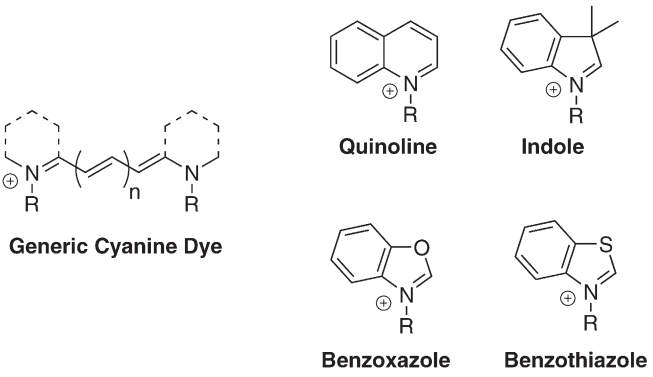


Fig. 1 Generic structure of cyanine dyes and common heterocyclic components. *R* refers to substituents, which are usually alkyl groups

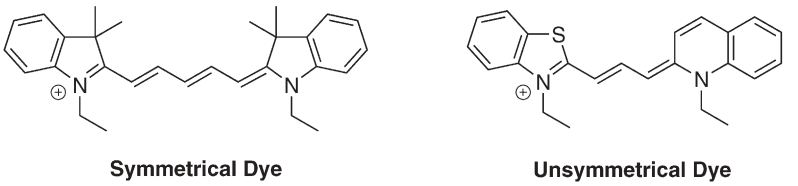


Fig. 2 Examples of symmetrical (*left*) and unsymmetrical (*right*) cyanine dyes

and "tricarbocyanine" refer to dyes having 1 ($n = 0$), 3 ($n = 1$), 5 ($n = 2$), or 7 ($n = 3$) methine groups in the bridge. In practice, the nitrogens are typically part of heterocyclic aromatic groups and are *N*-alkylated. The most common heterocycles are indole, quinoline, benzoxazole, and benzothiazole (Fig. 1). Substituents placed at various positions around these heterocycles are used to improve water solubility, influence aggregation, and vary the electronic transitions for the chromophores.

A further classification of dyes relates to the symmetry of the chromophore. Specifically, symmetrical dyes are composed of identical heterocycles linked at the same position, while unsymmetrical dyes consist either of two different heterocycles or two identical heterocycles linked at different positions. Examples of symmetrical and unsymmetrical cyanine dyes are shown in Fig. 2.

Many of the properties and applications of cyanine dyes were recently reviewed by Behera and coworkers [2]. This chapter will focus on one aspect of cyanine dye research, namely, their noncovalent interactions with DNA. A brief overview of the interactions between small molecules and DNA will be helpful for the remainder of the chapter.

1.2 DNA Binding Modes

Small molecules can interact with DNA in several modes [5, 6]. Electrostatic attraction to the anionic phosphodiester groups along the DNA backbone is possible for cationic molecules, although this is generally weak under physiological conditions. Binding within the major groove of the double helix is rare for small molecules, although more common for proteins. By far, the two most common binding modes are intercalation into the base pair stack at the core of the double helix, and insertion into the minor groove. Intercalation is typically observed for cationic molecules having planar aromatic rings. The positive charge need not be part of the ring system, but rather could be on a substituent. This binding mode requires two adjacent base pairs to separate from one another to create a binding pocket for the ligand [7].

Minor groove binders, on the other hand, usually have at least limited flexibility since this allows the molecule to adjust its structure to follow the groove as it twists around the central axis of the helix [8]. Binding in the minor groove requires substantially less distortion of the DNA compared with intercalative

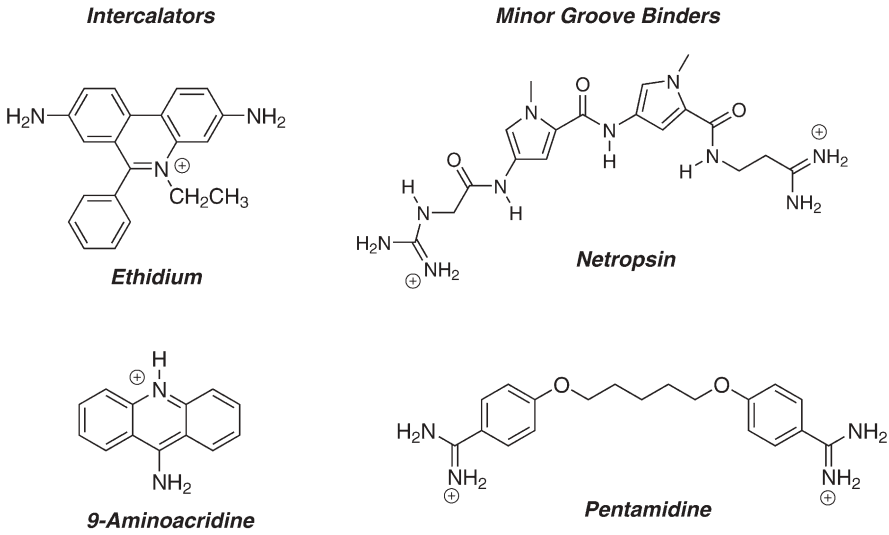


Fig. 3 Examples of DNA intercalators (*left*) and minor groove binders (*right*)

binding. Examples of well-known DNA intercalators and minor groove binders are shown in Fig. 3.

Several methods are commonly used to provide insight into the binding modes of small molecules with DNA [9]. Multidimensional NMR spectroscopy and x-ray crystallography can provide atomic-level detail of DNA–small molecule complexes. Even one-dimensional NMR can be useful since intercalation will often cause substantial changes in the chemical shifts of the imino protons involved in hydrogen bonding of the base pairs above and below the intercalation site. Likewise, binding in the minor groove can alter the chemical shifts of those protons present in the minor groove, specifically those on the deoxyribose sugar. While assignment of the protons in the 1D spectrum often requires multidimensional experiments, there are numerous DNA sequences already reported in the literature with peak assignments that can be used to study small molecule binding [10–13].

Both optical spectroscopy and hydrodynamic experiments provide indirect evidence to support or refute suspected binding modes. In the case of optical spectroscopy, UV-vis, fluorescence, circular dichroism (CD), and linear dichroism (LD) are all used. Binding to DNA will often cause a change in the absorption maximum and/or peak extinction coefficient. While this is insufficient to determine a binding mode, equilibrium binding constants can be determined based on the concentration dependence of any observed shifts. Fluorescent small molecules not only can exhibit changes in wavelength or quantum yield upon binding, but are often able to act as energy acceptors from the DNA bases. In these experiments, the DNA is excited with UV light and one looks for fluorescence from the ligand. If there is reasonably good overlap between the

emission spectrum of the bases and the absorption spectrum of the ligand, resonance energy transfer can occur [14]. Due to an orientation factor, the efficiency of this process is much higher for intercalators than for groove binders.

The two dichroism techniques can also be quite useful. Circular dichroism measures the differential absorption of right- and left-handed circularly polarized light. Small molecule ligands for DNA often possess achiral chromophores and thus do not exhibit any CD in solution. However, when bound to the chiral DNA, CD is induced in the ligand [15, 16]. This stands in contrast to UV-vis spectra, where both the bound and unbound ligand give detectable signals, usually in similar regions of the spectrum. The actual sign and magnitude of the induced CD signal is complicated and depends on the binding mode, DNA sequence, and orientation of the transition dipole of the ligand. However, intercalators will often exhibit lower intensity CD spectra compared with groove binders and this is most likely due to the fact that a groove binder contacts a larger part of the helix, ca. 4–6 base pairs, and twists to follow the groove. In contrast, a simple intercalator only contacts two base pairs and in most cases exhibits little or no twist. CD is also useful for detecting electronic coupling between multiple ligands that are bound in close proximity as this usually causes a splitting of the CD signal into positive and negative components [17].

Linear dichroism can be very useful for determining DNA binding modes [18]. LD measures the differential absorption of light polarized parallel and perpendicular to the macroscopic orientation axis. It relies on the use of long DNA polymers (hundreds of base pairs) that can be oriented in a flow cell. Any specifically bound ligands will also be oriented in this fashion and therefore will exhibit LD. The sign of the LD signal is directly related to the orientation of the ligand's transition dipole moment with respect to the direction of flow. Intercalators will be approximately perpendicular to the flow, while groove binders will be closer to parallel. These usually give LD signals of opposite sign (although the polarization of the transition dipole moment must be known in order to properly interpret the results). Drawbacks to this method include the need for (i) a specialized flow cell and (ii) a long DNA polymer that can be sufficiently oriented to give a measurable signal. The latter makes it difficult to make minor variations in sequence and study the effect on the binding mode.

In addition to the optical spectroscopic methods, two other methods for studying DNA binding modes are worth describing. Both techniques rely on the fact that DNA lengthens in order to create a binding site for an intercalator whereas no such perturbation is required to accommodate a minor groove binder. This lengthening can be observed directly using atomic force microscopy (AFM) in which single DNA molecules are imaged and their contour lengths measured with and without the ligand [19]. Alternatively, viscometry provides indirect measurements of helix lengthening since the viscosity of the solution has been observed to increase as a result of intercalation, but not minor groove

binding [9]. The latter method requires high concentrations of DNA and intercalator as well as long DNA polymers in order to detect the increases in viscosity. Long DNA molecules are also required for AFM in order to accurately measure the contour length, but the concentration used is limited only by the binding constant of the ligand.

Even when NMR or X-ray crystallography provide a high-resolution structure of a DNA-bound ligand, a number of the other techniques described above are used to further characterize the binding event. One reason for this is that DNA-binding ligands will often exhibit sequence-dependent binding modes. This can usually be detected easily and quickly using the lower resolution methods since the throughput is much higher (and cheaper) than with NMR or X-ray crystallography.

1.3 Cyanine Dye-DNA Interactions: Background

Cyanine dyes possess properties that are characteristic of both intercalators and minor groove binders, so it should not be surprising that both binding modes are commonly observed for this class of cationic compounds. The planar heterocycles on either end of the dye favor intercalation while the semiflexible polymethine bridge can permit the twisting required to follow the curve of the minor groove. Subtle variations in either the structure of the dye or sequence of the DNA can cause a change from one binding mode to the other. This not only offers great opportunities to synthetic and physical organic chemists to study structure-function relationships for DNA-binding cyanines, but can also impact the performance of dyes used in various applications.

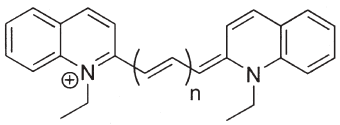
The interactions of cyanine dyes with DNA have focused mainly on applications in the life sciences and biotechnology. However, recent work has shown that cyanines can also assemble into interesting supramolecular aggregates by using DNA as a template. These findings, which extend the well-known propensity of cyanines to aggregate in aqueous solution as well as on a variety of synthetic and biological surfaces [20], are taking cyanine dye research into the realm of materials science and nanotechnology. Work in this area will be summarized at the end of this chapter.

The remainder of the chapter is organized according to binding modes: intercalation, minor groove binding, and aggregation. Each section is further subdivided according to the structure of the dye: symmetrical or unsymmetrical. While DNA recognition by the symmetrical dyes is a topic of considerable interest, it is the unsymmetrical dyes that have gained the most attention, due primarily to their large increases in fluorescence when bound to DNA.

2
Intercalators

2.1
Symmetrical Dyes

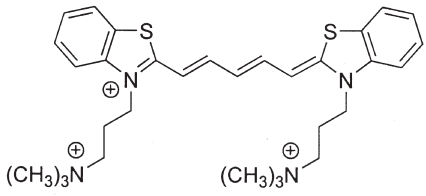
The first example of intercalation by a cyanine dye was reported by Nordén and Tjerneld, who carefully studied the interaction of **$DiQC_2$(1)** (also known as pseudoisocyanine, or PIC) with DNA isolated from calf thymus and therefore of random sequence [21]. A high affinity intercalation complex was formed, based on linear dichroism measurements. Interestingly, weaker complexes were formed at low ionic strength and high dye:DNA ratios. These complexes, which were identified as nonintercalated dimers and aggregates, will be discussed further in Sect. 4.1. The ability to recognize DNA via multiple binding modes is hardly unique to **$DiQC_2$(1)**, as will become evident later in this chapter.



$DiQC_2$(1): n = 0

$DiQC_2$(3): n = 1

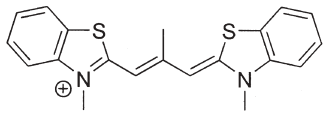
More recently, the symmetrical dicarbocyanine trication **$DiSC_{3+}$(5)** has been studied in terms of its binding to four different DNA sequences: the alternating duplexes $[Poly(dA\text{-}dT)]_2$ and $[Poly(dI\text{-}dC)]_2$, the nonalternating duplex Poly(dA)-Poly(dT), and the triplex Poly(dT)·Poly(dA)-Poly(dT) [22]. Viscometry experiments indicated that the dye intercalated into the triplex DNA as well as into the two alternating duplexes, but only at low dye:DNA ratios for the latter. When higher ratios were used for the alternating duplexes, no further lengthening of the helix was observed but an exciton-coupled induced circular dichroism band was noted, diagnostic for assembly of helical dye aggregates, as discussed in greater detail in Sect. 4.1. No helix lengthening was observed for



$DiSC_{3+}$(5)

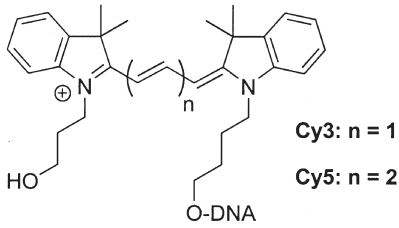
the nonalternating duplex but a strong induced CD band was detected, suggesting that minor groove binding occurred with this sequence.

Yarmoluk and coworkers also reported DNA binding studies on the polymethine-bridge substituted dye **Cyan 2** [23]. This dye exhibited a negative LD signal in the presence of mixed sequence DNA as well as alternating A-T and G-C copolymer duplexes, indicating intercalative binding of the dye to DNA.



Cyan 2

The symmetrical cyanines are encountered most often as fluorescent labels for DNA rather than as noncovalent binding agents. For example, **Cy3** and **Cy5** are among the most commonly used fluorophores and constitute a useful donor-acceptor pair for Förster resonance energy transfer (FRET) studies. The dimethylindole heterocycles work against DNA binding either by intercalation or groove binding due to the steric bulk introduced by the geminal methyl groups. However, an interesting NMR study by Lilley and coworkers showed that Cy3 covalently attached to the 5′-end of a DNA strand forms a pi-stacked complex with the terminal base pair of a DNA duplex formed in the presence of the complementary strand [24]. This type of "end-capping" would presumably be possible for the longer Cy5 dye as well.



2.2 Unsymmetrical Dyes

A number of intercalating cyanine dyes having unsymmetrical structures are shown in Fig. 4. Thiazole orange (**TO**) and oxazole yellow (**YO**) are unsymmetrical cyanine dyes based on benzothiazole, benzoxazole and quinoline heterocycles. **TO** was originally found to bind to double stranded (ds) DNA with a stoichiometry characteristic of intercalation (one dye for two base pairs) [25]. More recently, viscometry experiments demonstrated lengthening of DNA in the presence of **TO**, consistent with an intercalative binding mode [26].

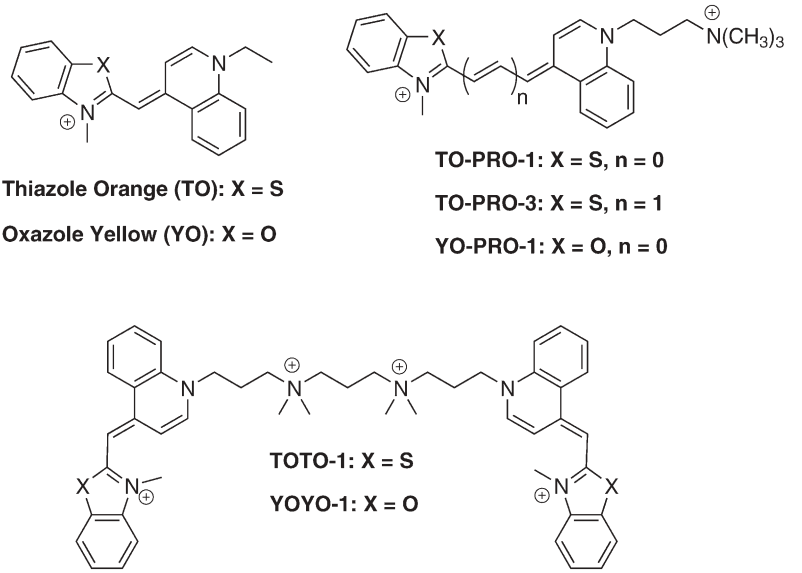


Fig. 4 Unsymmetrical cyanine dyes that bind to DNA by intercalation

Linear dichroism measurements on a dicationic analogue of **YO** (**YO-PRO-1**) were also consistent with an intercalative binding mode [27].

Nevertheless, the unsymmetrical cyanine dyes often exhibit mixed binding modes, as evidenced by various spectroscopic or hydrodynamic methods [27–29]. For the simple monomethine dyes these nonintercalative binding modes usually appear at higher dye:DNA ratios. In addition, extension of the bridge linking the two aromatic groups in the dye can lead to even more complex binding behavior, based on studies by Petty and coworkers using **TO-PRO-3**, which is analogous to **TO-PRO-1** except that a trimethine bridge connects the benzothiazole and quinoline systems (Fig. 4) [29]. As the bridge length increases, cyanine dyes become more structurally similar to classical minor groove binders and this probably contributes to the complexity of DNA binding by **TO-PRO-3**.

The structure of the intercalation complex for the unsymmetrical dyes is of interest. Neither high-resolution NMR nor X-ray crystallography has been used to investigate binding of **TO**, **YO**, or their dicationic analogues with DNA. (NMR structures for **YOYO-1** [30] and **TOTO-1** [13] bisintercalated into DNA have been published, but the geometries of the intercalated chromophores are undoubtedly influenced by the linker and its contacts within the minor groove of the DNA.) Recognizing that the positive charge in the unsymmetrical dyes will not be delocalized equally between the two rings, Yarmoluk and coworkers synthesized a variety of **TO** analogues and assessed the impact of various substituents on dye binding based on fluorescence enhancements [31]. Both steric and electronic effects were considered and led to the proposal of a "half-inter-

calation" model for binding in which the ring that bears less of the positive charge is intercalated while the ring that bears more of the charge projects into the minor groove.

While compounds having unfused aromatic systems such as that of **TO** have interesting properties as DNA intercalators, this is not the reason that such intense development of **TO** and its derivatives and analogues has occurred over the past 15 years. Rather, it is the fluorescence properties of the unsymmetrical cyanines that are so appealing. In aqueous solution, the dyes exhibit very low fluorescence. This is attributed to rapid nonradiative decay of the excited state dye through torsional motion in the methine bridge joining the two rings [27], as is the case for symmetrical cyanine dyes [32]. However, this conformational mobility is restricted when the dye is intercalated between DNA base pairs, meaning fluorescence enhancements of greater than 1,000-fold can be observed. Detecting small amounts of DNA, particularly in agarose or polyacrylamide gels, is of great importance in a variety of molecular biological applications, so achieving low background fluorescence in the absence of DNA but strong fluorescence in the presence of DNA was a significant advance [33, 34]. The fluorescence properties of thiazole orange and oxazole yellow have led to their incorporation into a number of DNA detection or probing systems, two of which are described here.

2.2.1 Light-up Probes

Detection of specific DNA sequences is important for genetic screening, clinical diagnostics and microarray analyses of gene expression. There are two steps that must occur: hybridization of a complementary probe strand to the DNA target followed by signaling that hybridization occurred. Although other methods are gaining in popularity, fluorescence is currently the most common form of detection. In many cases, either the DNA target or probe carries a fluorescence label. When hybridization occurs, a change in fluorescence wavelength, intensity, or polarization can be measured. A well known example of this strategy involves molecular beacon probes, in which the probe folds into a stable hairpin secondary structure, bringing together a fluorophore and quencher. Hybridization to a complementary strand requires opening of the hairpin, thereby separating the quencher from the fluorophore and enhancing fluorescence [35].

The sensitivity of **TO** fluorescence based on whether it is free in solution or bound to dsDNA suggested its use as a reporter for hybridization. This led to the development of "light up probes", where the cyanine is covalently attached to the end of the probe strand [36]. In the original example, the probe strand was composed of peptide nucleic acid (PNA, [37, 38]), which uses the same nucleobases as DNA, but has a polyamide backbone instead of a polyphosphodiester. In the absence of complementary DNA, the **TO** has low fluorescence due to the lack of a binding site. However, hybridization of the probe to a target strand,

particularly where the target is relatively long, allows the dye to intercalate between adjacent bases on the DNA target and leads to a large increase in fluorescence. A concern in these experiments is that **TO** has some affinity for the single stranded PNA, most likely due to a combination of partial stacking with adjacent nucleobases and the hydrophobic effect, and this can lead to significant background fluorescence [39]. Nevertheless, light-up probes have been commercialized making them available to laboratories that do not have synthetic chemistry capabilities.

2.2.2 A Nucleosome Structure Probe

A second application of thiazole orange as a probe for nucleosome assembly and dynamics was recently reported by Woodbury and coworkers [40]. The nucleosome consists of chromosomal DNA wrapped around cationic histone proteins. While this structure helps protect the genetic material from damage, the process by which the sequence is read during transcription is of great interest as it requires weakening of the assocation between the DNA and the protein. To study the protein-DNA interaction, a reactive form of **TO** was covalently linked to a unique cysteine residue in one of the histone proteins. Reconstitution of the protein with DNA resulted in a substantial increase in fluorescence for the **TO** label. Whether this was due to intercalation into the DNA or binding into one of the grooves was not determined. More importantly, binding of the protein and DNA most likely restricted the conformational mobility of the dye, causing the increase in fluorescence. The authors envision applications in real-time analysis of nucleosome assembly and remodeling, including single-molecule spectroscopy experiments.

These two applications illustrate the utility of covalently attached cyanine dyes as fluorescent reporter groups. There is also interest in using fluorescent dyes as noncovalent reporters, particularly in staining of gels to detect DNA. The fact that the dyes exhibit intense fluorescence only when bound to the DNA is advantageous because it permits detection of small amounts of DNA due to the low background fluorescence. However, the relatively low affinities of these dyes significantly decrease the detection sensitivity. In order to address this point, dimeric analogues of **TO** and **YO** (see **TOTO-1** and **YOYO-1** in Fig. 4, for example) were synthesized and found to bind DNA with very high affinity [34]. High resolution NMR spectroscopy demonstrated that both dyes are bis-intercalators [13, 30] although other experiments indicate that nonintercalative binding can also occur [27, 28]. These dyes maintain the high fluorescence enhancements exhibited by the monomeric analogues but can be used for detecting single DNA molecules [41].

Use of these dimeric dyes to detect DNA is complicated by the fact that they can bind inhomogeneously, where multiple dyes bind to one section of DNA (or one DNA molecule) while other regions (or DNA molecules) possess no bound dyes. The origin of this effect is not fully understood, although the slow

off-rate for these dyes dissociating from DNA is probably involved. Elevated temperatures [42] or high ionic strength [26] promote equilibration of the dye among all binding sites and yields a homogeneous distribution of dye–DNA complexes. The ability to disperse dye molecules throughout DNA rather than allowing them to cluster is important, based on work by Tuite and coworkers that showed how nonintercalated dyes undergo fluorescence quenching much more efficiently than intercalated dyes after prolonged irradiation [43].

3
Minor Groove Binders

3.1
Symmetrical Dyes

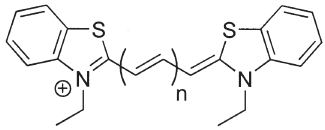
Most minor groove binding small molecules possess four structural features: (i) a positive charge; (ii) curvature; (iii) flexibility and (iv) hydrogen bond donor and acceptor groups to interact with complementary groups on the floor of the groove. (Note that the bottom of the groove corresponds to the edges of the DNA base pairs.) An example of a canonical minor groove binder is netropsin (Fig. 3). Meanwhile, the tricationic dye **DiSC$_{3+}$(5)** has the first three of these characteristics, but lacks the ability to form strong hydrogen bonds with the DNA. Nevertheless, spectroscopic and viscometric experiments indicate that the dye binds to the minor groove of DNA at nonalternating A-T sequences [22].

The first example of a cyanine dye-DNA interaction was reported for the quinoline derived carbocyanine dye **DiQC$_2$(3)** (also known as pinacyanol) by Kodama and coworkers in 1966 [44]. Linear dichroism measurements on calf thymus DNA indicated that the dye was bound within one of the grooves. More recent work by Mikheikin and coworkers provided stronger evidence in favor of minor groove binding by **DiQC$_2$(3)** [45]. Whereas the original study involved use of a random sequence DNA, the later work determined the linear dichroism spectra for regular polynucleotide sequences. A positive spectrum was recorded in the presence of Poly(dA)-Poly(dT), consistent with a minor groove bound dye. However, in Poly(dG)-Poly(dC), a negative spectrum was observed. The authors assigned this to intercalated dye within the duplex DNA.

The results for **DiQC$_2$(3)** were compared with three other carbocyanine dyes derived from indole, benzothiazole, and benzoxazole heterocycles. Of these, only the benzothiazole dye gave a positive LD spectrum with CT DNA, indicating that it too bound in the minor groove. The benzoxazole dye exhibits a reasonably strong negative LD suggestive of intercalation, while the indole dye gave a very weak negative LD. This indicates that the affinity of the dye for DNA is perhaps lower than the other dyes under the experimental conditions but that the dye that is bound does so within the minor groove.

Interestingly, at higher dye:DNA ratios, **DiQC$_2$(3)** retained a negative induced LD signal, but formed a 2:1 complex with DNA, indicating formation of

a minor groove bound dimer. Independent work by He and coworkers also indicated nonintercalative binding of this dye [46]. This is similar to the behavior of the natural product distamycin [11, 47] as well as a variety of synthetic compounds that recognize the minor groove sequence specifically [48, 49]. Earlier studies with dicarbocyanine dyes also indicated that dimerization within the minor groove is a significant DNA binding mode for symmetrical cyanines [50]. Several features of this binding mode were elucidated for the benzothiazole dye **$DiSC_2(5)$**. First, dimerization is cooperative, meaning that binding of one dye within the minor groove facilitates binding of a second dye to form the dimer. This is presumably due to better van der Waals interactions between two dyes stacked against one another than two isolated dyes in contact on either side with the sugar-phosphate backbone of the DNA. The minor groove probably has to widen in order to accommodate two ligands as a dimer, meaning dimerization is cooperative in spite of the energetic penalty incurred in distorting the DNA.



$DiSC_2(3)$: n = 1

$DiSC_2(5)$: n = 2

$DiSC_2(7)$: n = 3

A second feature to arise from these studies is that dimerization is highly selective for alternating A-T or I-C sequences over nonalternating A-T or alternating G-C. Binding in the minor groove as a monomer or intercalation are possible alternatives for the latter sequences. The sequence preference is consistent with a minor groove-binding mode since the exocyclic amino group of guanine projects into the minor groove and hinders binding of small molecules [51] (except for those designed to accept a hydrogen bond from G [49]).

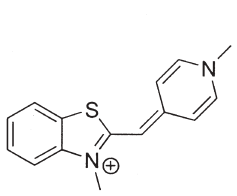
Further experiments revealed that the preference for dimerization in the minor groove decreased in the order quinoline >benzothiazole >benzoxazole >dimethylindole [52], as reported previously for monomeric minor groove binders [45]. The fact that quinoline-derived dyes exhibit the most favorable minor groove binding, whether as a monomer or dimer, indicates that recognition for the cyanine dyes is driven more by the hydrophobic effect and van der Waals contacts than by specific hydrogen bonding to the DNA.

3.2 Unsymmetrical Dyes

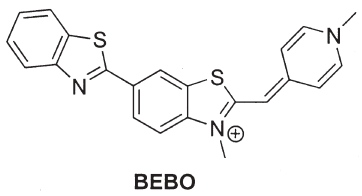
As noted in Sect. 2, binding of cyanine dyes to DNA can occur with significant fluorescence enhancements due to restricted internal rotation when bound to

the nucleic acid. This should be true regardless of whether the dye is intercalated or bound within the minor groove. In fact, Mikheikin and coworkers exhibited significant enhancements for the minor groove binding pinacyanol and benzothiazole carbocyanine dyes [45]. Fluorescence was also increased upon binding of the tricationic **DiSC$_{3+}$(5)** in the minor groove of the nonalternating duplex Poly(dA)-Poly(dT) [22].

In complementary work, Westman and coworkers have reported a series of unsymmetrical dyes that bind in the minor groove of DNA and give rise to enhanced fluorescence. The first compound in this series, **BEBO**, binds in the minor groove at alternating A-T sequences but intercalates into alternating G-C sequences, based on LD measurements [53]. In contrast, the parent compound **BO** intercalates in both sequences. Competition experiments in which both A-T and G-C DNA were present yielded CD spectra for **BEBO** that were similar to those recorded when the dye was added to pure A-T DNA. This indicates that the dye has a strong preference for binding in the minor groove of A-T sequences over intercalating into G-C sequences. The dye fluorescence quantum yield increased by approximately 10- to 20-fold for both the minor groove and intercalative binding modes. However, due to a shift of the absorption spectrum to longer wavelength upon binding to DNA, this quantum yield increase translates into a ca. 200-fold increase in the fluorescence intensity if an excitation wavelength is used where the dye absorbance increases upon binding.

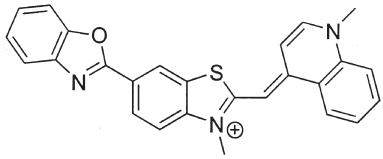


BO



BEBO

A variety of other dyes were synthesized based on the **BEBO** design. Of these, **BOXTO**, which varies in the first (benzoxazole instead of benzothiazole) and last (quinoline instead of pyridine) heterocycles, exhibits the most favorable minor groove binding [54]. The fluorescence quantum yield when bound to calf thymus DNA increases 50-fold and the intensity increases by up to 260-fold. These dyes fluoresce in the 500–600 nm range of the visible spectrum,



BOXTO

similar to ethidium bromide, but exhibit much greater stability during gel electrophoresis with dissociation half times of approximately 2 h [55].

4 DNA-Templated Cyanine Dye Aggregates

4.1 Symmetrical Dyes

Cyanine dyes are well known to aggregate in aqueous solution [20, 56]. Pi stacking of the conjugated systems of the dyes is favored based on both the hydrophobicity and polarizability of the dye. Two limiting types of supramolecular structures are formed and are referred to as "H" and "J" aggregates (Fig. 5). Unsubstituted dyes favor H-aggregation as this provides the greatest number of van der Waals interactions and minimizes exposure to water. However, substituents placed at various positions on the dye structure can promote J-aggregation due to steric and/or electrostatic factors. H- and J-aggregates typically exhibit absorption maxima that are shifted to either shorter or longer wavelength, respectively [57, 58].

As described in Sect. 3.1, the dicarbocyanine dye **$DiSC_2(5)$** was found to dimerize in the minor groove of DNA at alternating A-T and I-C sequences [50]. In aqueous solution, additional dyes stack on the dimer leading to an H-aggregate. However, the walls of the DNA minor groove preclude stacking additional dyes onto the dimer. Instead, the aggregate propagates by an end-to-end assembly mechanism. Propagation of the aggregate is cooperative, meaning that assembly of one dimer facilitates assembly of additional dimers directly

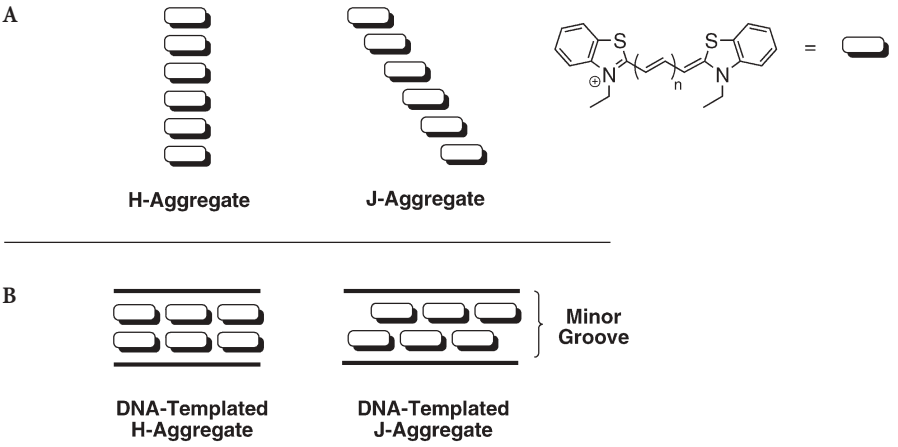


Fig. 5 Illustration of H- and J-aggregation by cyanine dyes in water (**A**) and on DNA templates (**B**).

adjacent to the first. This probably arises from the need to widen the groove in order to accommodate a dimer: the groove should be partially widened adjacent to a bound dimer before returning to its natural width further away, where the full energy penalty would have to be paid again in order to assemble a new dimer.

$DiSC_2(5)$ has a binding site size of approximately 5 base pairs, so placing 10 alternating A-T pairs in succession leads to assembly of two dimers in an end-to-end arrangement. This could be distinguished from isolated dimers due to splitting of the CD band: for a single dimer, a weak positive CD band is observed in the region where the dye absorbs but when two dimers are bound in close proximity, additional electronic coupling leads to a second order splitting of the excited state and, therefore, splitting of the CD band. Figure 6 illustrates this effect for the DNA-bound H-dimer and aggregate. In the H-dimer, transition to the upper state is allowed while direct excitation to the lower state is forbidden [57–59]. This leads to the hypsochromic shift that gives the H-dimer its name. When two dimers interact in an end-to-end fashion, the excited states split again but transition to both of the upper states is now allowed. The circular dichroism of these two states have opposite signs due to the different orientations of the transition dipoles, leading to a Cotton effect. A right-handed relationship was inferred based on the sign of the CD couplet [17] and this is expected for assembly of two dimers within the minor groove, where the chirality of the DNA template is effectively transferred to the dye molecules. On a polymeric DNA such as $[Poly(dA\text{-}dT)]_2$, which is ca. 200 base pairs long

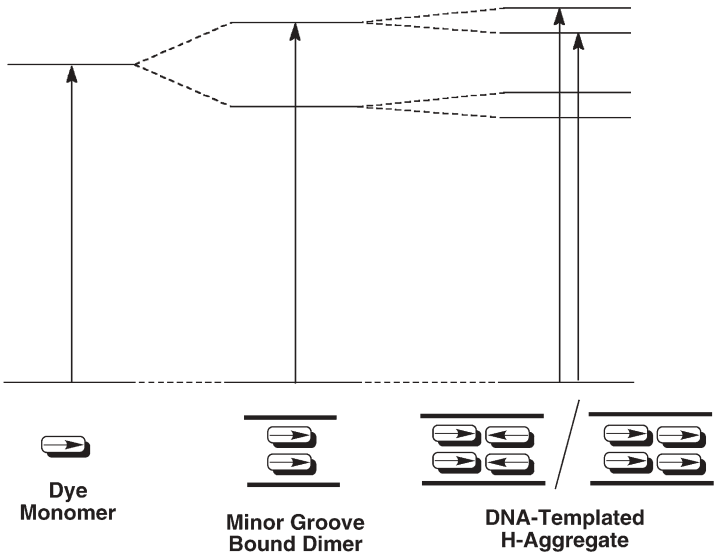


Fig. 6 Electronic coupling and allowed transitions for DNA-templated cyanine dye dimers and H-aggregates. *Arrows* are used to indicate transition dipole moment orientations

on average, a helical aggregate consisting of approximately 40 cyanine dimers assembles using the minor groove as a template. While a high-resolution structure of a DNA-templated aggregate has not been obtained, semiempirical calculations are consistent with the end-to-end assembly of dye dimers [60, 61].

Electronic couplings in these aggregates have been calculated directly from spectroscopic measurements. The face-to-face coupling between two dyes within a single dimer ranged between 2,000 and 3,100 cm^{-1}, while the end-to-end coupling between two dimers adjacent to one another within the groove provides approximately 700–1,000 cm^{-1} [52, 61, 62]. The difference makes sense in terms of the amount of orbital overlap available in face-to-face versus end-to-end interactions in H-aggregates. These values were relatively insensitive to the identity of the heterocycle but longer polymethine bridges gave larger face-to-face couplings [52].

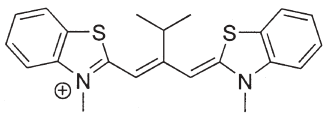
Another interesting feature of these aggregates revealed by the spectroscopy relates to the linewidth for the absorption bands. When cyanines form H-aggregates in solution, the dyes can adopt a range of translational and rotational relationships with one another. This leads to considerable broadening of the absorption band relative to the monomeric dye [56]. In contrast, the DNA-templated H-aggregates exhibit absorption bands that are narrower than the corresponding bands for unbound monomeric dyes [50]. This attests to the relatively rigid and well-defined structure of the dye aggregates templated by the DNA.

In aqueous solution, the tendency of cyanine dyes to aggregate increases with the length of the polymethine bridge, as expected based on the fact that both hydrophobicity and polarizability should increase for the longer dyes. For example, for the series of dyes based on benzothiazole, aggregation increased in the order **$DiSC_2(3)$ <$DiSC_2(5)$ <$DiSC_2(7)$** [56]. However, formation of the extended H-dimer aggregates on DNA increased in the order **$DiSC_2(3)$ <$DiSC_2(7)$ <$DiSC_2(5)$** [50]. Thus, the DNA template favored aggregation by the medium-length pentamethine dye over the longer heptamethine dye. One possible explanation for this discrepancy with the solution results relates to the induced fit nature of the recognition between the dye and DNA. In order to form stable minor-groove bound dimers, the dye must be able to twist in order to match the helicity of the DNA. The longer dye might simply be unable to twist sufficiently to follow the groove, leading to a less stable complex with the DNA. An interesting test of this hypothesis would be to use the "expanded" DNA helix recently reported by Kool and coworkers [63]. This DNA is underwound relative to a B-form helix, meaning that a minor groove binding ligand would not need to twist as much to maintain stable van der Waals contacts along the entire length of the complex without clashing with the wall of the groove. Such an experiment would be complicated for distamycin, which relies on a network of hydrogen bonding contacts to stabilize its complexes with DNA. In contrast, the cyanines presumably rely on van der Waals contacts and the hydrophobic effect to bind in the groove and, therefore, might more unambiguously reflect the changes in the helical parameters for modified DNA templates.

All of the unsubstituted dyes favor H-aggregation. However, J-aggregates are of much greater interest for both fundamental and technological reasons. Dyes used to sensitize photographic film to visible light typically form J-aggregates on silver halide crystals, while theoretical studies indicate that chiral J-aggregates should have useful nonlinear optical properties [64]. The tricationic cyanine dye **DiSC$_{3+}$(5)**, which showed diverse sequence-dependent DNA binding modes [22], also demonstrated the ability to assemble into both H- and J-aggregates on DNA [65]. Both aggregates were observed at higher concentrations of dye than were required for the monocationic analogue **DiSC$_2$(5)**, most likely because of the greater interdye repulsions for the trication. Nevertheless, on [Poly(dI-dC)]$_2$, the dye formed H- and J-aggregates that could be interconverted by heating (H) and cooling (J). J-aggregation by the dye could be a manifestation of the need to minimize repulsions between the *N*-(trimethylammonium)propyl substituents. The lower degree of stacking in the J-arrangement is consistent with the observation that H-aggregates were formed preferentially on shorter templates that supported binding of only one or two dimers. On the polymeric template, the lost stacking interactions within an individual dimer may be at least partially compensated by stacking with the next dimer, in which case only the outer heterocycles of the dyes at the very ends of the aggregate are unstacked.

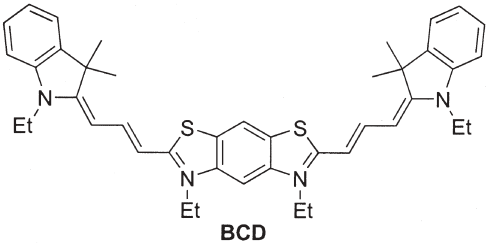
For the J-aggregated **DiSC$_{3+}$(5)**, the end-to-end coupling increased by ca. 50%, consistent with the greater interdimer overlap in structures where there is a greater offset between the chromophores within a single dimer [52, 62]. Peteanu and coworkers investigated the extent of electron delocalization within these J-aggregates and found that the aggregate size is a modest 4 chromophores [62], which appears to be inconsistent with the notion of an extended aggregate. High resolution structural information from either NMR or X-ray diffraction experiments are needed to fully understand these aggregates, but will be difficult to obtain due to the inability to form J-aggregates on the relatively short DNA duplexes normally used for the two methods. Nevertheless, the sharp absorption band, fluorescence, and induced chirality observed for these DNA-templated J-aggregates are intriguing and synthesis of other cyanine dyes that use mechanisms other than electrostatic repulsion to drive J-aggregation might provide an avenue for further development.

One such dye was recently reported by Yarmoluk and coworkers, who nicely demonstrated the use of a steric effect to promote J-aggregation of a cyanine on DNA [66, 67]. The bridge-substituted carbocyanine dye **Cyan β*i*Pr** is less prone to form H-aggregates in aqueous solution than the simpler methyl substituted analogue **Cyan 2**. The fact that **Cyan β*i*Pr** aggregates less in water than the less hydrophobic **Cyan 2** is most likely due to steric inhibition from packing the bulky isopropyl groups together in the closely overlapped structure of an H-aggregate. Interestingly, **Cyan β*i*Pr** exhibits the narrow, red-shifted absorption band that is the hallmark of J-aggregates in the presence of the nonalternating duplex Poly(dA)-Poly(dT) but not in alternating [Poly(dG-dC)]$_2$. While this suggests that the minor groove is involved, experiments with the alternating A-T or I-C duplexes might be more revealing.



Cyan β*i*Pr

One other example of a cyanine dye that assembles a J-aggregate in the presence of DNA was recently reported by Borissevitch and coworkers [68]. The dimeric cyanine **BCD** formed J-aggregates in the presence of calf thymus DNA at high dye:DNA ratios. Note that the rigidity of the dye makes it impossible for an intramolecular dimer to form, so intermolecular interactions must be responsible for the red-shifted absorption band observed in the presence of DNA. The actual structure for the aggregate will not fit the simple models shown in Fig. 6 due to the geminal dimethyl groups on the terminal indole heterocycles. As with the bridge-substituted dye described above, experiments with homogeneous DNA polymers will be interesting in terms of defining the sequence requirements for J-aggregation by **BCD**.



BCD

In each of the cases described above, a relatively high dye:DNA ratio was required to drive J-aggregation. Thus, these dyes are not particularly well suited to aggregate on DNA. In the case of the isopropyl-substituted and bichromophoric dyes, this could simply reflect steric factors, while for the tricationic dye, electrostatic repulsions between dye molecules makes it difficult to form stable aggregates. The challenge for future work along these lines lies in designing more hydrophobic dyes that can readily twist to follow the minor groove but do not easily stack into an overlapped H-dimer, but rather preferentially form offset J-dimers.

4.2 Unsymmetrical Dyes

While early work with the unsymmetrical cyanines indicated that DNA-bound aggregates could assemble at high dye:DNA ratios, very little work has gone into characterizing these structures. This is most likely because, in contrast to symmetrical dyes such as **DiSC$_2$(5)**, formation of the aggregates does not occur cooperatively and does not compete effectively with intercalation. However,

a recent report by Petty and coworkers described formation of DNA-templated aggregates by **TO-PRO-3**, the trimethine analogue of **TO-PRO-1** [29]. This dye can form at least three different complexes with DNA, two of which show significantly blue-shifted absorption spectra. Experiments with short oligomeric duplexes may help to better define the stoichiometry and binding mode for these intriguing dye–DNA complexes.

5 Outlook

The utility of the unsymmetrical cyanine dye intercalators based on thiazole orange and oxazole yellow is well established and analogues are available that fluoresce over much of the visible range. The more recently developed minor groove binding monomeric dyes such as **BEBO** and **BOXTO** provide alternatives to the intercalating dyes and it will be interesting to see what applications these dyes find, perhaps as longer wavelength DNA stains than the commonly used Hoechst or DAPI dyes.

The symmetrical dyes are currently most useful as covalently attached fluorescent labels. It is unlikely that these dyes will compete with the natural and synthetic polyamide compounds that recognize the DNA minor groove sequence specifically due to limited hydrogen bonding potential in the dyes. The observation that certain representatives of this class are capable of forming extended aggregates on DNA is perhaps of greater interest from a fundamental supramolecular chemistry perspective than for applications in life sciences research. The novel optical properties of these well-defined aggregates coupled with considerable control over the length and aggregation number affords the opportunity to carefully study electronically coupled multichromophore systems. Finally, there are a number of other interesting cyanine dyes that clearly interact with DNA, but insufficient data has been collected to assign binding modes [69–71]. Further examination of these compounds and their derivatives should lead to an even better understanding of structure-function relationships in cyanine dye–DNA recognition.

The cyanine dyes, originally developed to color clothing, later used as sensitizers in color photography, and most recently used as fluorescent labels and probes for life sciences and biotechnology research, continue to find new applications. The successes in designing new structures that recognize DNA in predetermined ways (intercalation, minor groove binding) illustrates our growing understanding of small molecule–DNA interactions as well as the impressive versatility of this class of compounds.

References

1. Williams CHG (1856) Trans R Soc Edinburgh 21:377
2. Mishra A, Behera RK, Behera PK, Mishra BK, Behera GB (2000) Chem Rev 100:1973
3. Sturmer DM (1977) In: Weissberger A, Taylor EC (eds) The chemistry of heteroaromatic compounds, vol 30. Wiley, New York
4. Mujumdar RB, Ernst LA, Mujumdar SR, Lewis CJ, Waggoner AS (1993) Bioconjugate Chem 4:105
5. Wilson WD (1996) In: Blackburn GM, Gait MJ (eds) Nucleic acids in chemistry and biology. Oxford University Press, Oxford, p 329
6. Mountzouris JA, Hurley LH (1996) In: Hecht SM (ed) Bioorganic chemistry:nucleic acids. Oxford University Press, New York, p 288
7. Lerman LS (1961) J Mol Biol 3:18
8. Geierstanger BH, Wemmer DE (1995) Annu Rev Biophys Biomol Struct 24:463
9. Suh D, Chaires JB (1995) Bioorg Med Chem 3:723
10. Klevit RE, Wemmer D, Reid BR (1986) Biochemistry 25:3296
11. Pelton JG, Wemmer DE (1989) Biochemistry 27:8088
12. Dwyer TJ, Geierstanger BH, Bathini Y, Lown JW, Wemmer DE (1992) J Am Chem Soc 114:5911
13. Spielmann HP, Wemmer DE, Jacobsen JP (1995) Biochemistry 34:8542
14. LePecq J-B, Paoletti C (1967) J Mol Biol 27:87
15. Lyng R, Rodger A, Nordén B (1991) Biopolymers 31:1709
16. Lyng R, Rodger A, Nordén B (1992) Biopolymers 32:1201
17. Nakanishi K, Berova N, Woody RW (1994) Circular dichroism: principles and applications. VCH, New York
18. Nordén B, Kubista M, Kurucsev T (1992) Quart Rev Biophys 25:51
19. Coury JE, McFail-Isom L, Williams LD, Bottomley LA (1996) Proc Natl Acad Sci USA 93:12283
20. Herz AH (1974) Photogr Sci Eng 18:323
21. Nordén B, Tjerneld F (1977) Biophys Chem 6:31
22. Cao R, Venezia CF, Armitage BA (2001) J Biomol Struct Dynam 18:844
23. Yarmoluk SM, Lukashov SS, Losytskyy MY, Akerman B, Kornyushyna OS (2002) Spectrochim Acta A 58A:3223
24. Norman DG, Grainger RJ, Uhrín D, Lilley DMJ (2000) Biochemistry 39:6317
25. Lee LG, Chen C, Liu LA (1986) Cytometry 7:508
26. Bordelon JA, Feierabend KJ, Siddiqui SA, Wright LL, Petty JT (2002) J Phys Chem B 106:4838
27. Larsson A, Carlsson C, Jonsson M, Albinsson B (1994) J Am Chem Soc 116:8459
28. Larsson A, Carlsson C, Jonsson M (1995) Biopolymers 36:153
29. Sovenyhazy KM, Bordelon JA, Petty JT (2003) Nucleic Acids Res 31:2561
30. Johansen F, Jacobsen JP (1998) J Biomol Struct Dynam 16:205
31. Yarmoluk SM, Lukashov SS, Ogul'chansky TY, Losytskyy MY, Kornyushyna OS (2001) Biopolymers 62:219
32. O'Brien DF, Kelly TM, Costa LF (1974) Photogr Sci Eng 18:76
33. Rye HS, Quesada MA, Peck K, Mathies RA, Glazer AN (1991) Nucleic Acids Res 19:327
34. Rye HS, Yue S, Wemmer DE, Quesada MA, Haugland RP, Mathies RA, Glazer AN (1992) Nucleic Acids Res 20:2803
35. Tyagi S, Kramer FR (1996) Nature Biotechnol 14:303
36. Svanvik N, Westman G, Wang D, Kubista M (2000) Anal Biochem 281:26
37. Nielsen PE, Egholm M, Berg RH, Buchardt O (1991) Science 254:1498

38. Nielsen PE (2002) In: Nielsen PE (ed) Peptide nucleic acids: methods and protocols (Methods in molecular biology). Humana, Towana, NJ, p 3
39. Svanvik N, Nygren J, Westman G, Kubista M (2001) J Am Chem Soc 123:803
40. Babendure J, Liddell PA, Bash R, LoVullo D, Schiefer TK, Williams M, Daniel DC, Thompson M, Taguchi AKW, Lohr D, Woodbury NW (2003) Anal Biochem 317:1
41. Auzanneau I, Barreau C, Salome L (1993) C R Acad Sci III 316:459
42. Carlsson C, Larsson A, Jonsson M, Albinsson B, Nordén B (1994) J Phys Chem 98:10313
43. Kanony C, Åkerman B, Tuite E (2001) J Am Chem Soc 123:7985
44. Kodama M, Tagashira Y, Nagata C (1966) Biochim Biophys Acta 129:638
45. Mikheikin AL, Zhuze AL, Zasadatelev AS (2000) J Biomol Struct Dynam 18:59
46. Wu H-L, Li W-Y, Miao K, He X-W, Yu C-F, Huang X-T (2002) Chin J Chem 20:462
47. Pelton JG, Wemmer DE (1989) Proc Natl Acad Sci USA 86:5723
48. White S, Szewczyk JW, Turner JM, Baird EE, Dervan PB (1998) Nature 391:468
49. Wemmer DE (2001) Biopolymers 52:197
50. Seifert JL, Connor RE, Kushon SA, Wang M, Armitage BA (1999) J Am Chem Soc 121:2987
51. Sehlstedt U, Kim SK, Nordén B (1993) J Am Chem Soc 115:12258
52. Garoff R, Litzinger EA, Connor RE, Fishman I, Armitage BA (2002) Langmuir :6330
53. Karlsson HJ, Lincoln P, Westman G (2003) Bioorg Med Chem 11:1035
54. Karlsson HJ, Eriksson M, Perzon E, Åkerman B, Lincoln P, Westman G (2003) Nucleic Acids Res 31:6227
55. Eriksson M, Karlsson HJ, Westman G, Åkerman B (2003) Nucleic Acids Res 31:6235
56. West W, Pearce S (1965) J Phys Chem 69:1894
57. Kasha M (1964) Physical processes in radiation biology. Academic, New York
58. Kasha M, Rawls HR, Ashraf El-Bayoumi M (1965) Molecular spectroscopy. Proc. VIII European congress molecular spectroscopy. Butterworths, London, p 371
59. Davydov AS (1971) Theory of molecular excitons. Plenum, New York
60. Yaron D, Armitage BA, Raheem I, Kushon S, Seifert JL (2000) Nonlinear Opt 26:257
61. Chowdhury A, Yu L, Raheem I, Peteanu L, Liu LA, Yaron DJ (2003) J Phys Chem A 107:3351
62. Chowdhury A, Wachsmann-Hogiu S, Bangal PR, Raheem I, Peteanu LA (2001) J Phys Chem B 105:12196
63. Liu H, Gao J, Lynch SR, Saito YD, Maynard L, Kool ET (2003) Science 302:868
64. Kobayashi T (1996) J-Aggregates. World Scientific, Singapore
65. Wang M, Silva GL, Armitage BA (2000) J Am Chem Soc 122:9977
66. Ogul'chansky TY, Losytskyy MY, Kovalska VB, Lukashov SS, Yashchuk VM, Yarmoluk SM (2001) Spectrochim Acta A 57:2705
67. Losytskyy MY, Yaschuk VM, Yarmoluk SM (2002) Mol Cryst Liq Cryst 385:27
68. Schaberle FA, Kuz'min VA, Borissevitch IE (2003) Biochim Biophys Acta 1621:183
69. Davidson YY, Gunn BM, Soper SA (1996) Appl Spec 50:211
70. Valyukh IV, Kovalska VB, Slominskii YL, Yarmoluk SM (2002) J Fluorescence 12:105
71. Kovalska VB, Losytskyy MY, Yarmoluk SM (2004) Spectrochim Acta A 60:129

Top Curr Chem (2005) 253: 77–88
DOI 10.1007/b100443

Between Objectivity and Whim: Nucleic Acid Structural Biology

Loren Dean Williams (✉)

School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332–0400, USA
loren.williams@chemistry.gatech.edu

Abstract Here we explore the subjectivity intrinsic to macromolecular crystallography, focusing on the hydration/counter-ion region of nucleic acids. Water molecules, monovalent and divalent cations, and polyamines compete for similar or adjacent sites. Many of these species give identical electron distributions (electron density maps). Such scattering iso-types allow one to construct different models that give similar fits of model to data. Even models with different electron densities can give similar fits of model to data because various parameters compensate. The geometries of the coordinating ligands of many species (magnesium excluded) are similar to each other, and are effectively identical within the

limitations of macromolecular crystallography. The observed distances are commonly occupancy-weighted averages. In sum, atom-type assignments and occupancies in the hydration/counter region cannot be unambiguously extracted from conventional X-ray diffraction experiments. We give several examples where incorrect atom-type assignments have been revealed by anomalous scattering experiments. Published structures and the database that contains them do not provide realistic representations of subjectivity. Reconsideration of macromolecular model-building protocols may be in order. Anomalous scattering provides information that can allow one to characterize the hydration/counter-ion region with greater accuracy than was previously possible.

Keywords Hydration · Water · Sodium · Potassium · Ammonium · Magnesium · Monovalent · Divalent · Cation · DNA · DNA–drug complexes · RNA · Uncertainty · Error

1 Introduction

The subjectivity inherent in macromolecular crystallography is not always apparent to end-users, or even practitioners of crystallography. Here we wish to focus on the dark side, and to discuss and illustrate errors and uncertainty. The focus is the hydration/counter-ion milieu of nucleic acids – including DNA and RNA oligonucleotides, DNA–drug complexes and DNA–protein complexes, and ribosomes. For nucleic acids and their complexes the hydration/counter-ion region is especially non-homogeneous because the high anionic charge induces association of many types of cations. Water molecules, monovalent and divalent cations, and polyamines compete for binding. Therefore many of the issues discussed here are relevant to nucleic acids, but not necessarily to proteins. The goal is to stimulate new approaches, and to promote cautious and reasoned interpretation of structural information. We believe reconsideration of model-building, reporting, and archiving protocols is in order.

The method scrutinized here is the assignment of water molecules and ions to isolated sum and difference electron density peaks ($2|F_o|-|F_c|$ and $|F_o|-|F_c|$ Fourier peaks) during the latter stages of refinement. The protocol is 'when in doubt, make it a water', with occupancy 1.00. If the thermal factor and geometry remain reasonable after refinement, then the water atom-type assignment and occupancy of 1.00 are considered to be validated. As explained below, advanced experimental approaches (anomalous scattering) demonstrate that these methods are not reliable. In the absence of anomalous scattering data, alternate atom-type assignments are possible, giving many models with reasonable thermal factors and geometry that fit the data equally well. Selecting one model, with a given composition of the hydration/counter-ion region, over another model with a different composition is essentially a subjective enterprise, and is not grounded in the experiment. In the concluding section we give some examples from our own work illustrating how the conventional approach can go very wrong.

2
Estimates of Errors

The macromolecular crystallographic method provides less information on errors, and less information to assess the fit of model to data, than is the norm for physical-chemical or biophysical fitting processes. To help the reader appreciate the distinctiveness of macromolecular crystallography, here we compare it with isothermal titration calorimetry (ITC) and small molecule crystallography.

3
Isothermal Titration Calorimetry

In this method one collects data (heats of injection) and conceives a model with parameters such as an equilibrium constant, a stoichiometric coefficient, and an enthalpy of binding. The parameters and the model are used to obtain calculated heats of injection. Differences between observed heats and the calculated heats are minimized by adjusting the parameters. During and after the fitting process one obtains a measure of global fit (χ^2), along with estimates of error of each parameter. The relationship between global fit and estimates of parameter error is not direct. The measure of global fit might be very good, even while estimate of error in one or more of the parameters is large, for example, if data on one side of the binding curve were absent.

4
Small Molecule Crystallography

In this method one collects data (observed structure factor amplitudes, $|F_o(hkl)|$) and establishes a preliminary model. The parameters of the model are the x, y, z coordinates of various atom-types, their thermal factors (which are generally anisotropic) and their occupancies. The parameters are used to obtain calculated structure factor amplitudes ($|F_c(hkl)|$). Differences between $|F_o(hkl)|$ and $|F_c(hkl)|$ are minimized by adjusting the parameters, and often by adding or subtracting atoms from the model. As with ITC, correctness of the model is indicated by global measures of fit (the R-factor) in addition to estimates of error in individual parameters.

5
Macromolecular Crystallography

As in the small molecule method, in the macromolecular method differences between $|F_o(hkl)|$ and $|F_c(hkl)|$ are minimized by adjusting the parameters of the model. Both methods yield global measures of fit (R-factor, R-free). How-

ever, realistic estimates of uncertainties of individual parameters (the x, y, z coordinates of each atom, the thermal factors, and the occupancies) are not obtained from the macromolecular method. The origins of this deficiency are beyond the scope of this discussion, but are related to very large numbers of parameters and very large numbers of data. Macromolecular structures (entries in the Nucleic Acid Database, NDB [1]) do not specify uncertainties of individual x, y, z coordinates, their thermal factors, or occupancies because those estimates of error are not available.

The protocol, after a reasonable model of the macromolecule along tight binding ligands is established, is to model the hydration/counter-ion region by adding water molecules and/or ions to corresponding sum and difference peaks of electron density (examples of sum electron peaks are shown in Fig. 1). However, the electron density in the hydration/counter-ion region of nucleic acid crystals is intrinsically uninformative and ambiguous. Ambiguity arises (i) from mixed and partial occupancies, (ii) from scattering iso-types, and (iii) from parameter-compensation. Each of these effects is explained and discussed below.

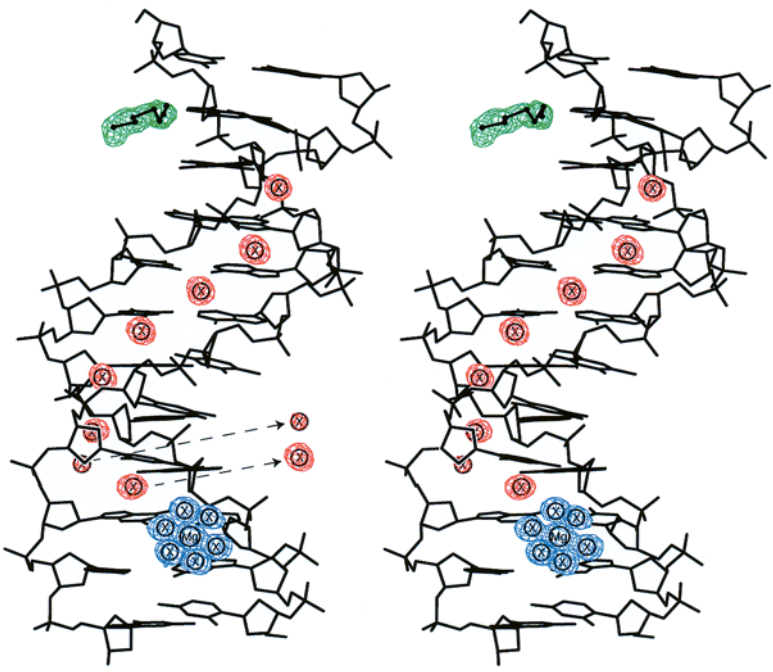


Fig. 1 Stereoview of sum ($2|F_o|-|F_c|$) electron density from a DNA duplex CGCGAATTCGCG at 1.4 Å resolution contoured at 1.0 σ (NDB entry bdl084). Only the electron density surrounding the hydrated magnesium ion (*blue net*), the partial spermine molecule (*green net*), and the minor groove solvent sites (*red net*) that interact with DNA bases is shown. For this structure, solvent sites were fitted to water molecules. The sum density from two minor groove solvent sites has been abstracted on the *left side*, to illustrate the lack of correspondence between the radius of a peak of electron density and the radius of the atom type

6 Mixed and Partial Occupancies

Crystallographic electron density and models are not restricted to formally correct chemical entities. The observed electron density is an ensemble average. Partially-occupied atoms and hybrid atom-types are 'observed'. This phenomenon is particularly acute in the hydration/counter-ion region of nucleic acids. Many species including water molecules, monovalent and divalent cations, and polyamines compete for similar or adjacent sites. It is possible for an electron density peak to arise from 40% one species, 30% another species, and 30% nothing.

7 Scattering Iso-Types

A large number of chemically distinct models give identical electron distributions (electron densities). This degeneracy arises because many of the species in the hydration/counter-ion region are *scattering iso-types* (our term). Scattering iso-types are defined as chemically distinct species (molecules or ions) with indistinguishable electron distributions (and X-ray scattering properties). An ammonium ion and a water molecule are scattering iso-types. The number of electrons in each is the same (10 electrons). The shape of the electron density is the same (spherical). A list of scattering iso-types is given in Table 1. Since partial and mixed occupancies are possible, the list of scattering iso-types is not restricted to formally correct chemical entities. The effective number of electrons of a potassium ion, when 55% occupied, is the same as that of a water molecule. In Table 1, occupancies that give equivalent numbers of elec-

Table 1 Scattering Iso-types: full and partial occupancy

Species	Occupancy	Radius (Å)	Number of electrons
H_2O	1.0	1.40	10
Na^+	1.0	0.95	10
Na_4^+	1.0	1.40	10
Mg^{2+}	1.0	0.65	10
$K_{0.6}^{+}$ [a]	0.6	1.33	10
$Rb_{0.4}^{+}$ [b]	0.28	1.48	10
$Tl_{0.11}^{+}$ [c]	0.11	1.49	10
$Ca_{0.4}^{+}$ [d]	0.4	1.69	10

[a] 60% occupied potassium ion.
[b] 28% occupied rubidium ion.
[c] 11% occupied thallium ion.
[d] 40% occupied calcium ion.

Table 2 Scattering iso-types: mixed occupancy H_2O/Na^+

Occupancy H_2O	Occupancy Na^+	Effective radius (Å)	Effective number of electrons
1.0	0.0	1.40	10
0.9	0.1	1.36	10
0.8	0.2	1.31	10
0.7	0.3	1.27	10
0.6	0.4	1.22	10
0.5	0.5	1.18	10
0.4	0.6	1.13	10
0.3	0.7	1.09	10
0.2	0.8	1.04	10
0.1	0.9	1.00	10
0.0	1.0	0.95	10

Table 3 Scattering iso-types: mixed occupancy H_2O/K^+

Occupancy H_2O	Occupancy K^+	Effective radius (Å)	Effective number of electrons
1.0	0.00	1.40	10
0.9	0.05	1.40	10
0.8	0.10	1.39	10
0.7	0.15	1.39	10
0.6	0.20	1.38	10
0.5	0.30	1.37	10
0.4	0.35	1.37	10
0.3	0.40	1.36	10
0.2	0.45	1.35	10
0.1	0.50	1.34	10
0.0	0.55	1.33	10

trons are indicated for common species. The list of possible scattering iso-types can be infinitely expanded by continuously varying the partial occupancies as illustrated for H_2O/Na^+ hybrids in Table 2 and H_2O/K^+ hybrids in Table 3.

One might naively assume that differences in shape, arising from differences in ionic/atomic radii could differentiate various species. The molecular radius of H_2O is greater than the ionic radius of Mg^{2+}, suggesting that the electron density from a Mg^{2+} ion should be of greater amplitude at the peak center and of less breadth than that of a water molecule. However, the effective dispersion of the electrons about the atom center, which is described in the structure factor equation as the rms amplitude of the atomic displacement, varies with

thermal fluctuations, positional disorder, and data quality. The contributions of those effects generally obscure differences in ionic/atomic radii. This effect is illustrated in Fig. 1.

8 Compensating Parameters

The problem extends beyond scattering iso-types and partial occupancies. Models with different electron densities can fit the same data equally well. Changes in one parameter are compensated in the fit by changes in other parameters. For example, decreasing the number of electrons of a K^+ ion (achieved by decreasing the occupancy or by changing the atom type from K^+ to Na^+ or H_2O) is compensated wholly or partially by decreasing the thermal factor. This compensation is related to determinacy (ratio of data to parameters). Therefore, hydration/counter-ion occupancies are generally not refined. Instead occupancies are fixed at a default value of 1.00 (the number of significant figures is absurd). This default value is arbitrary. In many cases switching an occupancy from 1.00 to 0.80 would not affect map quality or refinement statistics. Adjustments in thermal factors would eat up the difference.

9 Coordination Fingerprints

As explained above, the hydration/counter-ion region of nucleic acids is especially difficult to characterize unambiguously. Yet there is some hope provided by coordination fingerprints. The octahedral geometry and the ligand-to-metal distances of a well-ordered Mg^{2+} do allow reliable identification. It is conceivable that Na^+ could be distinguished from H_2O by a coordination fingerprint. The ideal distance from Na^+ to ligand is less than that from H_2O to ligand (Brown [2] has provided useful surveys of coordination geometries). However, in the macromolecular experiment, the experimentalist must interpret subtle differences in geometries because the observed distances are generally occupancy-weighted averages (Tables 2 and 3). How does one interpret nominally short contacts between water molecules? Do they arise from partial Na^+ occupancy or simply from coordinate error? The database contains very few examples of Na^+ ions with unambiguous coordination geometry, many with unreasonable geometry, and none with reasonable estimates of atom-type error. Many if not most of the Na^+ ions in the NDB have coordination geometries consistent with those of water molecules.

10
Revaluation of Published Structures

Considering the limitations imposed by mixed and partial occupancies, scattering iso-types, compensating parameters, and occupancy-weighted geometries it is simply impossible to unambiguously extract the correct identities of the contributing species from the data (in the absence of anomalous scattering data). The hydration/counter-ion region cannot be reliably characterized by standard macromolecular x-ray methods. We believe there is a realistic possibility that many or even most 'water molecules' and ions are mis-assigned and/or incorrectly described in the NDB. In the following sections we provide some examples where the mis-assignment of atom-types has been experimentally revealed.

11
Mis-assigned Peaks

11.1
Z-DNA

A very high resolution structure (1.0 Å resolution) of the spermine form of Z-DNA [3] was determined by Martin Egli, Loren Williams, and Qi Gao, as postdoctoral researchers, in the laboratory of Alexander Rich. Water molecules were assigned to sum and difference electron density peaks, including a series of regularly spaced peaks in the minor groove, which formed a 'spine of hydration'. The water molecules were well-behaved, with reasonable geometry and interactions, realistic thermal factors, good electron density, etc. The final published structure (NDB entry) contains water molecules within the 'spine of hydration', which are assigned atom types of O (hydrogen atoms are implicit).

After that structure had been completed and published, evolution of cryocrystallographic techniques allowed the same group to collect the first ever data set from a flash-frozen crystal of DNA [4], consisting of the same spermine-form Z-DNA. Lowering the temperature dampens molecular motions, increasing the information content of the diffraction data. The data obtained from a flash-frozen crystal indicated that the minor groove contained a spermine molecule. Lowering the temperature converted the isolated peaks in the minor groove that had been assigned to a spine of hydration into a tube of electron density. The spine of hydration morphed into a spermine molecule. At room temperature the methylene groups of the spermine molecule were thermally disordered, and so not visible in the electron density maps. The amino groups of the spermine formed hydrogen bonds to the floor of the minor groove and were ordered at room temperature, and appeared as spheres of electron density. The peaks of electron density in the minor groove had been mis-assigned during the original refinement of the room temperature structure. They were not water

molecules but were the amino groups of a spermine molecule. Our retrospective assessment is that our original conclusion, that peaks of electron density indicated a spine of hydration, was an over-interpretation of the data. However, even the revised structure, with occupancies of 1.00 for the spermine molecule, etc., is also an over-interpretation.

11.2
B-DNA #1

By 1998 there were 68 isomorphous members of the CGCXAATTYGCG (X=G or A, Y=C or T) dodecamer family in the NDB. Collectively those structures contain thousands of water molecules, and no monovalent cations, divalent cations, or polyamines.[1] We proposed an alternative interpretation of the data [5–7] in which the grooves of the B-DNA in those crystals are decorated with various types of cations. That proposal is now accepted, based in part on substitution experiments with scattering cations that give distinctive (anomalous) scattering information [8–10] (also see work by Egli and coworkers [11]). It is now clear that much of the hydration/counter-ion region contains partial and mixed-occupancy water molecules and cations. The results of anomalous experiments indicate that the A-tract minor groove contains monovalent cations. Increasing data quality has revealed that the major groove is associated with hydrated magnesium ions and spermine molecules (Fig. 1). Our retrospective assessment is that the original models, which for example assumed that peaks of electron density indicated a well-ordered spine of hydration [12, 13], were over-interpretations of the data. The predominant monovalent cation in the crystallization solutions was Na^+ (a scattering iso-type with water). Models with Na^+/H_2O partial occupancy hybrids would therefore fit the data equally as well as models with water only. There was never any experimental basis for setting the atom-types as pure water or their occupancies as 1.00.

11.3
B-DNA #2

In 1998 we identified a Mg^{2+} ion in the major groove of CGCGAATTCGCG [7]. The ion was well-behaved in the refinement, with excellent electron density, thermal factors, geometry, etc. The location of the Mg^{2+} was confirmed by other investigators [14, 15]. It is a 'consensus' cation, observed in structures obtained from a variety of crystallization conditions and DNA modification. The major groove Mg^{2+} has been assigned a structural role by several investigators, and is thought to contribute to the famous 'dodecamer bend'.

[1] A spermine molecule identified in an early dodecamer structure has been revised, and is now considered to be a hydrated magnesium ion.

In 2001 we discovered with an anomalous scattering experiment that the major groove Mg^{2+} is not fully occupied [9]. A monovalent signal is subtle in the sum and difference maps but is unambiguous in the anomalous map. The monovalent site is displaced somewhat from the divalent site, although the proximity is such that occupancy of either site by a cation would preclude occupancy of the other (the sum of the occupancies cannot exceed 1.00). Since the Mg^{2+} is only partially occupied, the role of the Mg^{2+} in contributing to the dodecamer bend is unclear.

Our retrospective assessment is that our original major groove Mg^{2+} ion model was an over-interpretation of the data. The good thermal factor/good geometry criterion was not sufficient to ascribe an occupancy of 1.00 to the Mg^{2+}.

11.4 DNA–Drug Complexes

In X-ray structures of several early DNA–anthracycline complexes, well-defined and fully occupied Na^+ ions mediate interactions between the intercalated chromophore and the DNA [16, 17]. In considering the reliability of those Na^+ assignments the following factors require consideration:

1. The Na^+ atom-type assignments are based on geometric considerations. It was assumed that six ligands surrounding an electron density peak is definitively an indicator of Na^+ ion occupancy.
2. After initial refinement, the geometry of the ligands surrounding the Na^+ peak was restrained, to 'ideal' octahedral Na^+ geometry.
3. The definition of ideal Na^+ geometry has evolved over time. Six-coordinate octahedral geometry is no longer considered a reliable indicator of sodium. As noted by Jeffery [18], water molecules engaging in bifurcated hydrogen bonding can be six-coordinate.
4. Structures of some DNA–anthracycline complexes lack localized cations [19, 20]. The Na^+ ions in the original structure have not proved to be fully reproducible.
5. Recent anomalous experiments show that there are additional monovalent cation sites in the DNA–anthracycline complexes [21]. In the anomalous experiments the original Na^+ sites are not as highly occupied as other sites.

Our retrospective assessment is that localized cations and electrostatic forces are indeed important in structure, thermodynamics, and sequence specificity of DNA–ligand complexes. Favorable interactions of adriamycin and cations with the sequence-specific electrostatic landscape of DNA may be universal characteristics of DNA–small molecule interactions and may be useful in sequence-specific ligand design. However, the certitude and the details of the original Na^+ descriptions are in error.

12 Summary

Water molecules, monovalent and divalent cations, and polyamines compete for similar or adjacent sites in the hydration/counter-ion regions of nucleic acids. In general, crystallographic models of hydration/counter-ion regions are biased simplifications. In assessing the reliability of crystallographic models one must consider many factors:

1. With scattering iso-types one can construct many different models with indistinguishable electron density maps.
2. Even models with different electron density maps can give similar fits of model to data.
3. The geometries of the coordinating ligands of common species (Mg^{2+} excluded) in the hydration/counter-ion region of DNA are very similar to each other, and are effectively identical within the limitations of macromolecular crystallography.
4. The good thermal factor/good geometry criterion is not sufficient to ascribe an occupancy of 1.00.
5. Published structures and the database that contains them do not provide realistic representations of uncertainty.
6. The atom type assignments and occupancies in the hydration/counter region are generally subjective.
7. In many cases electron density peaks should be assigned to wild-card atoms (10 electrons, identity unknown).
8. Anomalous scattering experiments provide important information that allows one to characterize the hydration/counter-ion region with greater accuracy than was previously possible.

References

1. Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh S-H, Srinivasan AR, Schneider B (1992) The nucleic acid database. a comprehensive relational database of three-dimensional structures of nucleic acids. Biophys J 63:751–759
2. Brown ID (1988) What factors determine cation coordination numbers. Acta Crystallogr B44:545–553
3. Egli M, Williams LD, Gao Q, Rich A (1991) Structure of the pure-spermine form of Z-DNA (magnesium free) at 1 Å resolution. Biochemistry 30:11388–11402
4. Bancroft D, Williams LD, Rich A, Egli M (1994) The low temperature crystal structure of the pure-spermine forms of Z-DNA reveals binding of a spermine molecule in the minor groove. Biochemistry 33:1073–1086
5. Shui X, Sines C, McFail-Isom L, VanDerveer D, Williams LD (1998) Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. Biochemistry 37:16877–16887
6. McFail-Isom L, Sines C, Williams LD (1999) DNA structure: cations in charge? Curr Opin Struct Biol 9:298–304

7. Shui X, McFail-Isom L, Hu GG, Williams LD (1998) The B-DNA dodecamer at high resolution reveals a spine of water on sodium. Biochemistry 37:8341–8355
8. Woods K, McFail-Isom L, Sines CC, Howerton SB, Stephens RK, Williams LD (2000) Monovalent cations sequester within the a-tract minor groove of [D(CGCGAATTCGCG)]2. J Am Chem Soc 122:1546–1547
9. Howerton SB, Sines CC, VanDerveer D, Williams LD (2001) Locating monovalent cations in the grooves of B-DNA. Biochemistry 40:10023–10031
10. Tereshko V, Minasov G, Egli M (1999) A "hydrat-ion" spine in a B-DNA minor groove. J Am Chem Soc 121:3590–3595
11. Tereshko V, Wilds CJ, Minasov G, Prakash TP, Maier MA, Howard A, Wawrzak Z, Manoharan M, Egli M (2001) Detection of alkali metal ions in DNA crystals using state-of-the-art X-ray diffraction experiments. Nucleic Acids Res 29:1208–1215
12. Drew HR, Dickerson RE (1981) Structure of a B-DNA dodecamer. Iii. Geometry of hydration. J Mol Biol 151:535–556
13. Kopka ML, Fratini AV, Drew HR, Dickerson RE (1983) Ordered water structure around a B-DNA dodecamer. A quantitative study. J Mol Biol 163:129–146
14. Chiu TK, Kaczor-Grzeskowiak M, Dickerson RE (1999) Absence of minor groove monovalent cations in the crosslinked dodecamer CGCGAATTCGCG. J Mol Biol 292:589–608
15. Minasov G, Tereshko V, Egli M (1999) Atomic-resolution crystal structures of B-DNA reveal specific influences of divalent metal ions on conformation and packing. J Mol Biol 291:83–99
16. Wang AH, Ughetto G, Quigley GJ and Rich A (1987) Interactions between an anthracycline antibiotic and DNA: molecular structure of daunomycin complexed to D(Cpgpt-papcpg) at 1.2 Å resolution. Biochemistry 26:1152–1163
17. Frederick CA, Williams LD, Ughetto G, van der Marel GA, van Boom JH, Rich A, Wang AH-J (1990) Structural comparison of anti-cancer drug-DNA complexes: adriamycin and daunomycin. Biochemistry 29:2538–2549.
18. Jeffrey GA (1997) An introduction to hydrogen bonding. Oxford University Press, New York
19. Williams LD, Egli M, Ughetto G, van der Marel GA, van Boom JH, Quigley GJ, Wang AH-J, Rich A, Frederick CA (1990) Structure of 11-deoxydaunomycin bound to DNA containing a phosphorothioate. J Mol Biol 215:313–320
20. Lipscomb LA, Peek ME, Zhou FX, Bertrand JA, VanDerveer D, Williams LD (1994) Water ring structure at DNA interfaces: hydration and dynamics of DNA–anthracycline complexes. Biochemistry 33:3649–3659
21. Howerton SB, Nagpal A, Williams LD (2003) Surprising roles for electrostatic interactions in DNA–ligand complexes. Biopolymers 69:87–99

Top Curr Chem (2005) 253: 89–108
DOI 10.1007/b100444

Topoisomerase Inhibitors of Marine Origin and Their Potential Use as Anticancer Agents

Nathalie Dias[1] · Hervé Vezin[2] · Amélie Lansiaux[1] · Christian Bailly[1,3] (✉)

[1] INSERM U-524 et Laboratoire de Pharmacologie Antitumorale du Centre Oscar Lambret, IRCL, Lille 59045, France
[2] Laboratoire de Chimie Organique Physique, CNRS UMR8009, USTL Bât C3, 59655 Villeneuve d'Ascq, France
[3] Present address: Institut de Recherche Pierre Fabre, Centre de Recherche en Oncologie Expérimentale, 3 rue des satellites, 31432 Toulouse, France
christian.bailly@pierre-fabre.com

Abstract Cancer chemotherapy relies to a great extent on the use of drugs like doxorubicin, etoposide, and camptothecin derivatives, which are targeted at topoisomerases. These enzymes that resolve DNA topological constraints in cells are the primary targets for a large diversity of natural products extracted from plants, microorganisms or, as reviewed here, from marine organisms. Marine products capable of stabilizing topoisomerase II-DNA covalent complexes include the makaluvamines, pyridoacridines, and xestoquinones. In the pyridoacridine series, comprising over a hundred secondary metabolites extracted from ascidians and sponges, the leading products are ascididemin and neoamphimedine, which both exhibit potent cytotoxic activities against tumor cells, in part due to their capacity to intercalate into DNA and to interfere with topoisomerase II. A few compounds of marine origin targeting topoisomerase I have also been discovered such as wakayin, kalihinol F, and the cyclic peptide sansalvamide A, which inhibits the type I topoisomerase from the *Molluscum contagiosum* virus. But the leading topoisomerase I poison is the marine alkaloid lamellarin D, recently characterized as a potent pro-apoptotic cytotoxic agent. The capacity of lamellarins D and M to stabilize topoisomerase I-DNA covalent complexes provides a novel chemical strategy to design antitumor agents targeting topoisomerase I. Altogether, this review illuminates the vast, under-exploited chemical diversity typical of sea products and underlines the potential benefit of marine-derived topoisomerase inhibitors as antitumor agents.

Keywords Marine products · Topoisomerases · Anticancer drugs · Pyridoacridine · Lamellarin

Abbreviations

CPT	Camptothecin
Lam-D	Lamellarin D
PA	Pyridoacridines
PQ	Pyrroloquinoline

1 Introduction

The DNA double helix model can be regarded as an iconographic association of little spheres connected by twisted rigid rods. This three-dimensional view is useful as a means of understanding the nature and global configuration of the genetic material but it has the disadvantage of masking the complexity of the DNA system, which is in fact a dynamic factory by itself, with multiple, intricate conformations and functions. DNA is a tremendously deep reservoir of vital information; it carries our individual genetic signature. But if we now know the (nearly) complete sequence of the genome, we still have a very limited, not to say simplified, view of its functions, expression, coordination, transmission, etc. Therefore, to better understand DNA, one needs a variety of tools to manipulate it. Our current box of tools for manipulating DNA includes various classes of proteins such as restriction enzymes, transcription factors, helicases, and topoisomerases, which can modulate the structure and/or the dynamics of the genetic information. Over the past two decades, topoisomerases have emerged as an important class of targets for anticancer agents. Numerous small molecules have been designed to try to control the activity of these enzymes with the idea of selectively killing cancer cells, among other purposes.

Topoisomerases are enzymes essential for normal cell proliferation [1]. They are required to maintain the integrity of the DNA helix in replication, transcription, and chromosome condensation in mitosis [2]. Three classes of topoisomerase enzymes are found in human cells. Topoisomerase I introduces breaks into one strand of DNA and mediates DNA relaxation in a process that does not require any cofactor. In contrast, topoisomerase II (α or β) cuts the two strands of the DNA double helix and changes the topology of DNA by catalyzing the passage of one double-stranded segment of DNA through a double strand break produced in a second DNA segment before its resealing [3]. Human topoisomerase III, identified more recently [4, 5], cleaves single-strand DNA and binds covalently to the 5′-end of the cleaved DNA [6]. There are yet no specific inhibitors for topoisomerase III (α or β) whereas there are many for the other two enzymes. Both topoisomerases I and II constitute critical cellular loci for a number of clinically important antitumour agents [7–12]. For example, the two widely used anticancer drugs etoposide and daunomycin act as topoisomerase II poisons whereas topotecan specifically interferes with topoisomerase I functions. The mechanism by which these agents kill cancer cells involves the stabilization of an intermediate topoisomerase-DNA covalent binary complex

in which one (topoisomerase I) or the two (topoisomerase II) DNA strands have undergone strand scission and are covalently attached to the enzyme. By binding to this complex and preventing DNA religation, the topoisomerase inhibitor (or poison) precludes DNA replication and transcription, and thereby leads to the death of cells attempting to undergo these processes. Many topoisomerase inhibitors have been chemically synthesized or isolated from microorganisms or plants [8, 10, 13–16]. Some of them have also been extracted from marine organisms. It is this latter class of marine-derived natural products active against topoisomerases I and/or II which is reviewed here. Our objective is to highlight the chemical diversity of these compounds and to discuss the possible relationships between topoisomerase inhibition and anticancer activity within this class of marine products. For convenience, we start first with a survey of the marine-derived topoisomerase II inhibitors and then we will present the most recent and promising compounds interfering with topoisomerase I.

2 Marine Topoisomerase II Inhibitors

Marine products inhibiting topoisomerase II have been studied principally in two chemical series: the pyrroloquinolines (PQ) and the pyridoacridines (PA) [17]. The leading compounds of the PQ family are the makaluvamines (Fig. 1), principally (but not exclusively) isolated from the sponge *Zyzzya fuliginosa* [18–24]. This marine sponge, found in the Indopacific Ocean and in Papua New Guinea, has provided several collections of PQ compounds, including the batzellines [25]. More than 15 different makaluvamines (named from A to P) have been isolated but not all of them interfere with topoisomerase II activity. Makaluvamine A (Fig. 1) potently inhibits topoisomerase II-mediated decatenation of kinetoplast DNA whereas makaluvamine B, which only differs by the

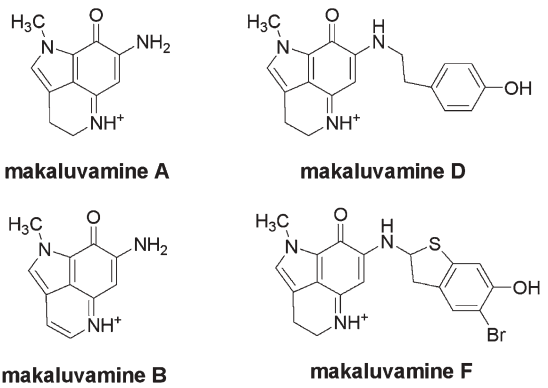


Fig. 1 Structure of selected makaluvamines

presence of a double bond in the quinoline (Q) ring, is inactive. The substitution of the exocyclic amino ring on the Q ring can profoundly modify the cytotoxicity and anti-topoisomerase II activity of the drug. For example, makaluvamine F, with a benzothiophene ring, is equally effective compared to makaluvamine A and its cytotoxic potential is reinforced. On the other hand, makaluvamine D, with an ethylphenol group, is considerably less potent both in terms of cytotoxicity and topoisomerase II poisoning [26, 27]. A similar pyrrolo[4,3,2-*de*]quinoline core (pyrroloiminoquinolone) is identified in a variety of sponge alkaloids such as the tsitsikammines, epinardins, isobatzellines, and discorhabdins but their activity against topoisomerase II has not been reported.

The second main group of marine topoisomerase II inhibitors is that of the PA compounds, which primarily refer to ascidian metabolites typified by ascididemin, first isolated from the Okinawan tunicate *Didemnum* sp. [28], and meridine, from the ascidian *Amphicarpa meridiana* [29]. Over the past 10 years, the list of cytotoxic PA derivatives, obtained by total synthesis or from a marine source, has greatly expanded to include many ascididemin-related topoisomerase II inhibitors. This is the case, for example, for the amphimedines and the shermilamines. Recently, 100 PA derivatives with cytotoxic activities have been described [30] and a very interesting PA family tree has been created to analyse the biodiversity in this rich family of marine compounds [31]. The structure of PA alkaloids (Fig. 2) presents a large structural variety, from tetracyclic to octacyclic compounds [32] as well as sulfur-containing molecules, such as isoplimanine and lissoclinidine both isolated from the New Zealand ascidian *Lissoclinum notti* [33], varamine A from the Fijian ascidian *L. vareau* [34], and dercitamide and sagitol from the sponge *Oceanapia sagittaria* [35]. PAs are frequent in sponges (phylum Porifera) and ascidians (phylum Urochordata) but they are also found in Cnidaria and Mollusca [32].

Ascididemin was initially isolated from the Okinawan tunicate *Didemnum* species [28] but it was also extracted from an Indonesian red *Didemnum* sp. [36] and from the mediterranean ascidian *Cystodytes dellechiajei* collected near the Balearic islands (Spain) [37]. This marine organism (Fig. 3) has also provided cystodamine, the amino counterpart of meridine [38]. Interestingly, the same ascidian *Cystodytes dellechiajei* collected in Brazil yielded two other related PA alkaloids, sebastianines A and B (Fig. 2). These two PA derivatives, with a fused pyrrole or pyrrolidine ring, display p53-dependent cytotoxic activities in the micromolar range [39]. A Fijian *Cystodytes* sp. ascidian was also characterized and it yielded a variety of PA alkaloids, shermilamine B and C, cystodytin A and J, eilatin, most of which showed topoisomerase II inhibitory and DNA intercalation properties [40]. More recently, ascididemin was also isolated from a Singaporean ascidian, along with the related PA, kuanoniamines A, D, E and F, subarine [41]. The kuanoniamines are widely distributed among ascidians and sponges [42]. The *Cystodytes* species has yielded many PA zoochrome alkaloids, such as arnoamines A and B, isolated from an ascidian collected near Arno Atoll (Republic of the Marshall Islands) [43, 44].

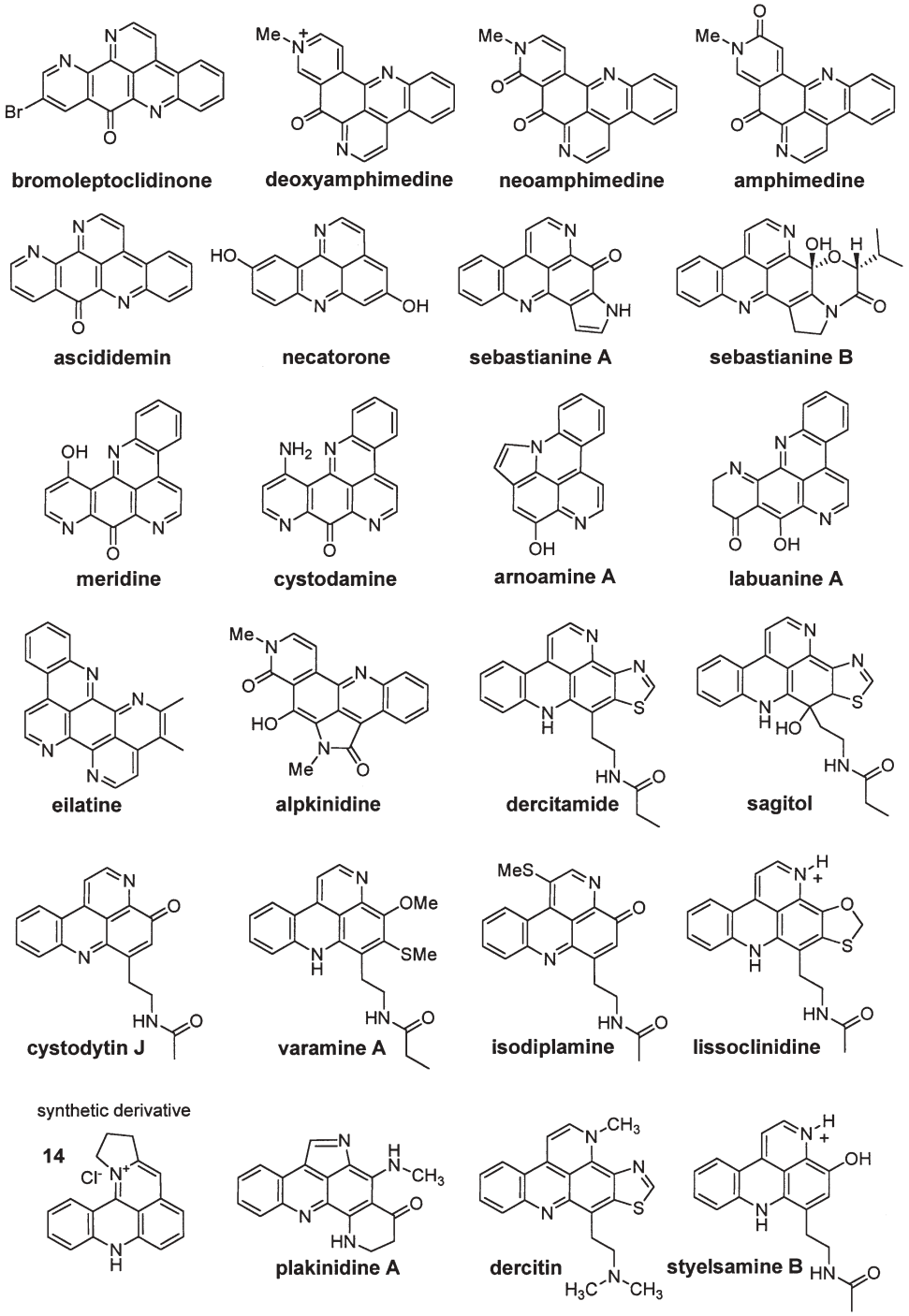


Fig. 2 Structure of various marine pyridoacridine alkaloids

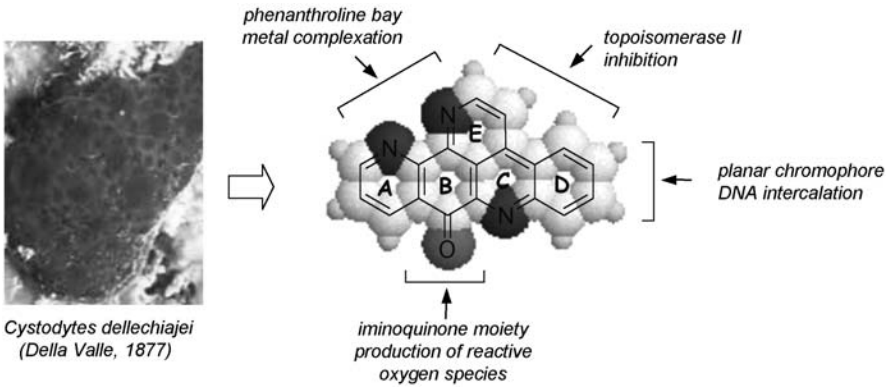


Fig. 3 Specimen of *Cystodytes dellechiajei* collected in Spain (Taxonomy: Phylum, Chordata; Class, Ascidiacea; Order, Enterogona; Suborder, Aplousobranchia; Family, Polycitoridae) (from http://www.guiamarina.com/ascidiacea/) and structural model of ascididemin with its pharmacophoric domains

One of the interesting bio-analogues of ascididemin is the compound named labuanine A, recently isolated from the Indonesian marine sponge *Biemna fortis*, which exhibits neuritogenic activity and accumulation of neuroblastoma. Neuro 2A cells are blocked in the G2/M phase of the cell cycle, possibly related to topoisomerase II inhibition. This compound, along with an amino derivative also bearing the phenanthroline and iminoquinone moieties typical of ascididemin, represents an interesting lead compound for neuronal differentiation [45].

Ascididemin exhibits potent cytotoxic activities in vitro against various types of tumor cell lines, including multidrug-resistant cells [37] but it shows moderate in vivo antitumor activities. It also displays antiparasitic activities, in particular against *Plasmodium falciparum* and trypanosomes [46]. It is a typical DNA intercalating agent interfering with activities of both topoisomerase I and topoisomerase II, but the effect is not very pronounced [37]. A dose-dependent stimulation of topoisomerase II-mediated DNA cleavage was observed with ascididemin and the sites of DNA cleavage were identical to those obtained with the reference drug etoposide [47]. The N8 nitrogen in the A-ring and the pyridine E-ring are both essential for topoisomerase II inhibition (Fig. 3). Deletion of that pyridine ring abolishes cytotoxicity [48]. But the extent of inhibition of topoisomerase II is considerably weaker with ascididemin compared to etoposide. Moreover, ascididemin was found to be almost equally toxic to HL60 human leukemia cells sensitive or resistant (HL60/MX2 cells) to the topoisomerase II inhibitor mitoxantrone. It is therefore unlikely that topoisomerases constitute the primary molecular targets of ascididemin and derivatives [47]. The DNA damage produced by ascididemin may not be primarily driven by topoisomerases but may well result from the production of metal-dependent reactive oxygen species in the presence of a reducing agent.

Indeed, Matsumoto and coworkers have elegantly demonstrated that the ascididemin class of PA marine alkaloids is capable of reductive DNA cleavage [49]. Reduction of the iminoquinone moiety generates oxygen free radicals responsible for DNA damage (single strands breaks) and ultimately cytotoxicity. The cytotoxicity of ascididemin is further exacerbated in cell lines defective for DNA repair, such as the Chinese hamster ovary cell lines EM9 and xrs-6, respectively defective for the proteins XRCC1 and Ku80, both implicated in DNA repair. These two cell lines are significantly more sensitive to ascididemin than the corresponding repair competent AA8 cell line [50]. In a very recent study, the mechanism of cell death induced by ascididemin was investigated in detail using Jurkat leukemia T cells and a link between its DNA-damaging activity and drug-induced apoptosis via induction of mitochondrial dysfunction was proposed [51]. One of the early reports on acridine marine alkaloids indicated that the sponge metabolite dercitin exhibits potent cytotoxic and antitumor activities associated with DNA intercalation and inhibition of DNA polymerase I but with minimal effect on topoisomerase activity [52]. The conclusions of this early study parallel those drawn more recently with ascididemin.

If the cytotoxicity of ascididemin does not principally rely on topoisomerase II inhibition, other PA alkaloids do exploit topoisomerase II to express their cytotoxic action. This is the case for the related compound neoamphimedine, which is particularly attractive with respect to its unusual activity on topoisomerase II. This pentacyclic PA derivative, isolated from the sponge *Xestospongia* sp. collected from Surigao (Philippines), has been characterized as a novel type of topoisomerase II inhibitor. Unlike the analogue amphimedine [53], it slightly stimulates topoisomerase II-dependent cleavage of DNA but in addition neoamphimedine has the rare capacity to promote catenation of high molecular weight DNA by topoisomerase II, in a histone-like manner [54]. This atypical activity was investigated further in a recent study [55] which demonstrated that the catenation of DNA induced by neoamphimedine correlates with drug-induced DNA aggregation. This compound displays enhanced cytotoxicity in yeasts over-expressing topoisomerase II and it shows marked cytotoxicity to a range of mammalian tumor cell lines, whereas the isomer amphimedine, inactive against topoisomerase II, is non toxic. Moreover, neoamphimedine has revealed in vivo activities against tumor xenografts. It was found to be as effective as etoposide against mice bearing human epidermoid-nasopharyngeal KB tumors and as effective as 9-aminocamptothecin in mice bearing human colon HCT-116 tumors. The anticancer potential of neoamphimedine is potentially attributable to its action at the topoisomerase II level [55]. Its atypical effect on DNA catenation could be exploited to design novel types of topoisomerase II poisons. Drug-enhanced catenation of DNA may represent a new unexploited tactic for blocking topoisomerase II and thereby tackling cancer cells. The third compound of this group is deoxyamphimedine (Fig. 2), also isolated from *Xestonspongia* sp. This compound is also highly cytotoxic but unlike the other two analogues, it does not require topoiso-

merase II to exert its action. Induction of reactive oxygen species-mediated DNA damage by deoxyamphimedine should be responsible for its cytotoxic action [56].

The expanding family of naturally occurring PA alkaloids also includes 2-bromoleptoclinidinone (the brominated analogue of ascididemin [57]), styelsamine B (the oxidized analogue of cystodytin J [58]) and meridine, the structure of which has been exploited to design novel anticancer agents. Meridine was first isolated from the tunicate *Amphicarpa meridiana* [29] but it was also found in the marine sponge *Corticium* sp. It displays cytotoxic and antifungal activities [59, 60]. Meridine, like other PA, is the product of a biosynthetic pathway common to marine organisms in different phyla. Meridine itself is quite cytotoxic with IC_{50} values (the drug concentration inhibiting cell growth by 50%) ranging from 10 μM to 10 nM depending on the histopathological cell type, but its cytotoxic potential can be further enhanced upon suitable chemical modification. Indeed, a few highly potent meridine analogues with IC_{50} values in the low nanomolar range have been synthesized, such as the dehydroxy analogue of meridine which shows sub-nanomolar activities in certain cell types [61]. Diverse other molecular constructs analogous to the PA have been reported, for example, pyrroloacridine derivatives such as the plakinidines [36, 62].

The chemical literature on PA is abundant [63–70]. Numerous synthetic derivatives of ascididemin and meridine have been synthesized and evaluated as anticancer agents. Structure–activity relationships have been established [71, 72] and a few promising anticancer drug candidates have been identified. One of them is the pentacyclic indolizino[7,6,5,-*kl*]acridinium salt **14** (Fig. 2), a GC-selective DNA intercalator much more active against topoisomerase II than amsacrine (the reference compound in the anilinoacridine series). This compound is highly cytotoxic to various tumor cell lines, including cells resistant to etoposide and, importantly, it is not a substrate for P-glycoprotein-mediated drug efflux. It accumulates in the nucleus of tumor lung and breast cells where it acts to trigger apoptosis [73]. Many other synthetic derivatives have been designed in the ascididemin, amphimedine, and meridine series [74–77]. The newly synthesized analogues frequently show marked cytotoxic activities, but significant antitumor efficacy in in vivo models has not yet been reported.

A few other marine products targeting topoisomerase II have been described. This is the case for the epipolysulfanyl-dioxopiperazine compound designated as leptosin M (Fig. 4), extracted from a strain *Leptosphaeria* sp., a microorganism inhabiting the marine alga *Sargassum tortile*. This complex dimeric molecule is cytotoxic and marginally active against topoisomerase II (IC_{50} of 59 μM) and ineffective against topoisomerase I [78]. Of considerably higher interest are the xestoquinone-type compounds such as adociaquinone B and xestoquinone itself characterized as effective topoisomerase II poisons [79]. However, these pentacyclic marine quinines, isolated from tropical sponges of the genus *Xestospongia*, display various pharmacological activities. Xesto-

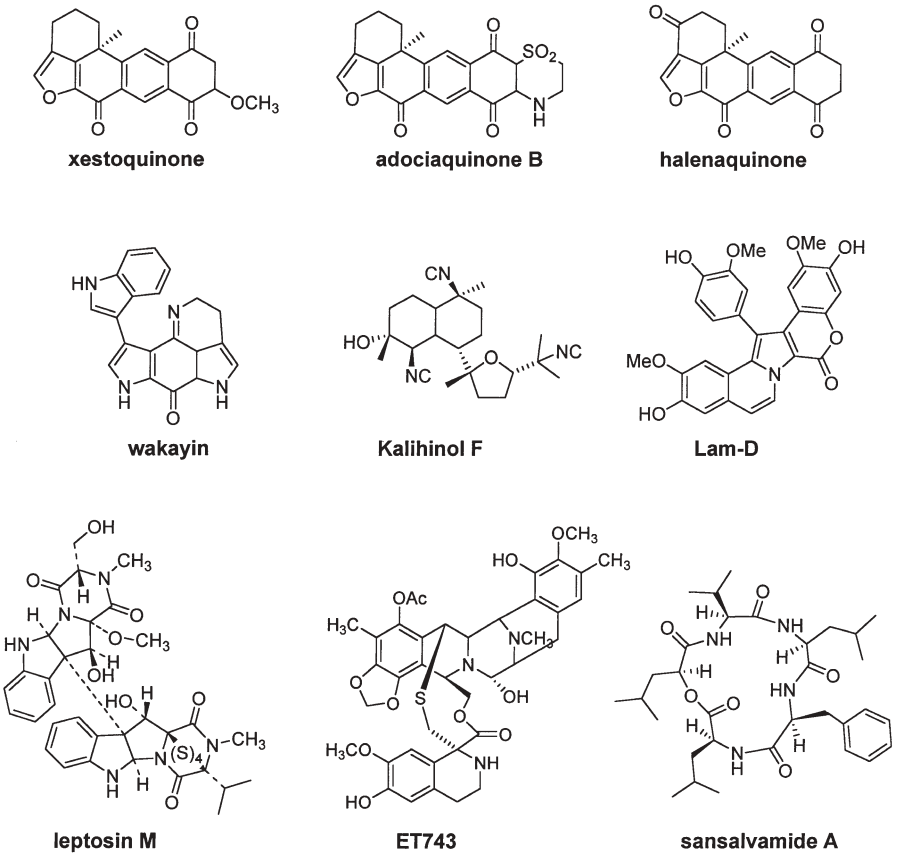


Fig. 4 Other topoisomerase inhibitors of marine origin

quinone inhibits topoisomerase II but in addition it causes calcium release from sarcoplasmic reticulum in skeletal muscle, inhibiting its ATPase activity [80–83]. The related compound halenaquinone, from the Okinawan sea sponge *Xestospongia exigua*, has been described as a pro-apoptotic phosphatidylinositol-3 (PI-3) kinase inhibitor [84] but also as a topoisomerase I inhibitor [85]. Finally, inhibition of the catalytic activity of both topoisomerases I and II, without stabilisation of enzyme-DNA covalent complexes, has been reported recently with a D-galactan sulphated polysaccharide GA3P produced by the marine microalga Dinoflagellate *Gymnodinium* sp. A3 [86]. This extracellular polysaccharide behaves like dextran sulfate and probably interacts with topoisomerases. It displays in vitro activity against a range of cancer cell lines [86] but it is hard to believe that this cytotoxic activity is related to topoisomerase inhibition because it seems unlikely that the extracellular polysaccharide can enjoy direct access to the nuclear enzymes as mentioned in a previous paper [87], which was subsequently retracted [88].

3
Marine Topoisomerase I Inhibitors

The history of the topoisomerase I inhibitors is not as old as that of the topoisomerase II poisons but these compounds arguably show better promise as novel anticancer agents. The field of the topoisomerase I poisons is largely dominated by a single natural product from the plant kingdom, camptothecin (CPT). The natural history and mechanism of action of this compound has been extensively described in many comprehensive reviews [89–96]. The key point of its mechanism of action to bear in mind is the unprecedented capacity of CPT to stabilize topoisomerase I-DNA complexes in the form of a covalent intermediate whereby the enzyme is attached to one strand of the DNA through a phosphotyrosyl linkage. This activity, which leads to the promotion of DNA single strand breaks, is shared with a limited number of natural products, in particular certain glycosyl indolocarbazoles related to the antibiotic rebeccamycin [97] and several series of synthetic molecules including protoberberines [98, 99], terbenzimidazoles [100–102], dibenzonaphthyridinones [103–105] and indenoisoquinolines [106–112]. There are also a few marine products known to inhibit topoisomerase I. They are reviewed here.

One of the first CPT-like topoisomerase I poisons isolated from a marine organism was a bispyrroloiminoquinone metabolite called wakayin (Fig. 4), isolated from the ascidian *Clavelina* collected at Wakaya island (part of the Fiji archipelago) [113]. The structure of this compound is somewhat reminiscent of the aforementioned makaluvamines targeting topoisomerase II, but not topoisomerase I. The indole ring linked to the tetracyclic bispyrroleiminoquinone core must play a significant role in the capacity of wakayin to inhibit topoisomerase I. Replacement of the indole substituent with a phenyl group corresponds to the compound called tsitsikammamine A, isolated from a latruncalid sponge and this molecule showed no topoisomerase I inhibitory activity [17]. This type of architecture may be compared with that of lamellarine D (see below), which has also a large planar chromophore substituted with a bulky phenoxy group. Wakayin is considerably (ten- to 100-fold) less potent than CPT at stimulating DNA cleavage by topoisomerase I but its sequence selectivity profile is identical. Both CPT and wakayin induce topoisomerase I-mediated DNA cleavage principally at sites containing a T residue on the 3′ side of the break [113]. However, the covalent DNA-topoisomerase I complexes stabilized by wakayin are much less stable than those detected after exposure to CPT. In addition, unlike CPT, wakayin is considered as a DNA intercalating agent (it induces unwinding of supercoiled DNA) as it is the case with the makaluvamines. Intercalation of the planar drug chromophore between base pairs is probably required for topoisomerase I inhibition [114].

Inhibition of topoisomerase I has also been reported with the alkaloid ecteinascidin 743 (ET-743 in Fig. 4, trabectidin, Yondelis) extracted from the mangrove Caribbean tunicate *Ecteinascidia turbinata* [115]. However, it is now considered that the poisoning activity cannot play a role in the mechanism of

action of ET-743 because it is only observed at very high concentrations, in the range of 10 µM, i.e., 500- to 1000-fold higher than optimal 50% cell growth inhibitory concentration values. Moreover, the deletion of the *top1* gene in yeast has no effect on ET-743 cytotoxicity [116] and its cytotoxic action toward HCT-116 colon carcinoma cells is also independent of topoisomerase I expression [117].

Recently, inhibition of topoisomerase I was reported with kalihinol F (Fig. 4) extracted from the sponge *Acanthella* collected off the coast of Cape Sada (Japan). This diterpene with three isonitrile groups showed no effect on topoisomerase II but was found to inhibit topoisomerase I, at least at high concentrations [118]. A much more convincing set of data was reported with the marine fungal product sansalvamide A (Fig. 4), a cyclic depsipeptide produced by *Fusarium* species [119]. This compound was found to potently inhibit DNA relaxation mediated by the topoisomerase encoded by the poxvirus *Molluscum contagiosum* (MCV), which is a type-I enzyme. This pathogenic virus causes opportunistic infections in immunocompromised patients and the lesions (molluscum bodies) induced by the virus in AIDS patients are essentially

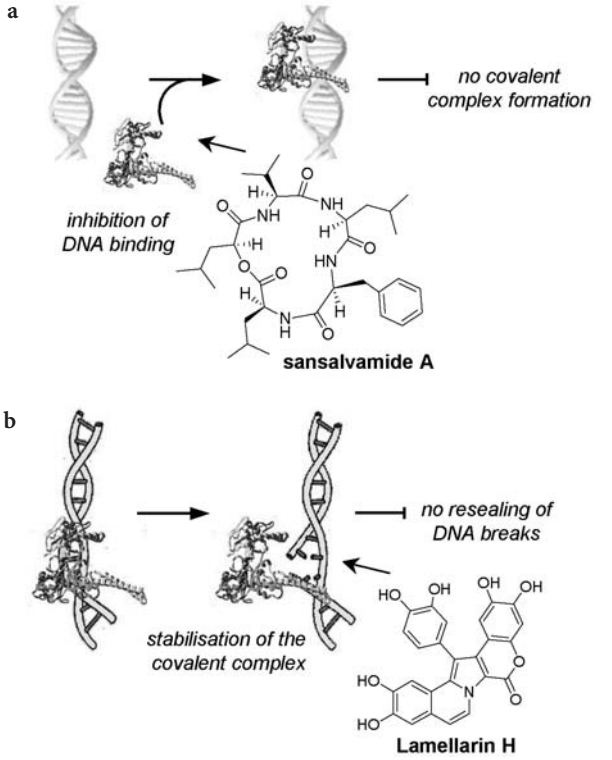
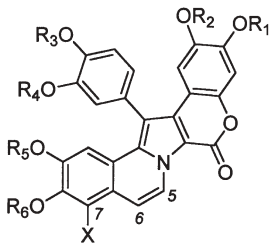


Fig. 5a,b Different mechanisms of inhibition of **a** MCV topoisomerase by sansalvamide and **b** topoisomerase I by lamellarin H

untreatable [120]. Interestingly, this cyclic peptide is not a classical poison stabilizing topoisomerase-DNA covalent complexes but rather inhibits the binding of the enzyme to DNA (possibly via direct interaction with the enzyme), thereby precluding covalent complex formation (Fig. 5a). However, it does not prevent DNA resealing in a preformed covalent complex [119]. Inhibition of MCV topoisomerase has been also reported [121] with another marine product with a totally different structure, lamellarin H, purified (together with lamellarin E, F, and G) from a deep-purple didemnid colonial ascidian, *Didemnum chartaceum*, found in the Indian Ocean [122]. In this case, the drug behaves as a typical poison, stabilizing topoisomerase-DNA covalent complexes so as to prevent the religation of DNA single strand breaks, as illustrated in Fig. 5b. The lamellarins represent a promising family of anticancer agents. The class includes over 30 members denominated by letters from A to Z, including sulfated members as well as the recently discovered compounds lamellarin α and β (see [123] for a comprehensive review). We have recently characterized lamellarin D (Lam-D) as a potent inhibitor of human topoisomerase I [124] but inhibition can certainly occur with all lamellarins possessing a 5,6-double bond in the B ring, such as lamellarins B, D, H, M, N, W, and X (Fig. 6). The capacity of Lam-D to stabilize DNA-(3′-phosphotyrosyl)-enzyme covalent intermediates is comparable to that of the prototypic topoisomerase I poison camptothecin. Both compounds were found to convert supercoiled plasmid DNA efficiently into nicked DNA in a dose-dependent manner. However, their sequence selec-



lamellarin	R_1	R_2	R_3	R_4	R_5	R_6	X
B	H	CH_3	H	CH_3	CH_3	CH_3	OCH_3
D	H	CH_3	H	CH_3	CH_3	H	H
H	H	H	H	H	H	H	H
M	H	CH_3	H	CH_3	CH_3	CH_3	OH
N	H	CH_3	CH_3	H	CH_3	H	H
W	H	CH_3	CH_3	H	CH_3	CH_3	OCH_3
X	H	CH_3	CH_3	H	CH_3	CH_3	H

Fig. 6 Structure of selected marine lamellarines

tivity is slightly different. As shown in Fig. 7, within a DNA fragment of 117-bp, Lam-D and Lam-M stimulate DNA cleavage by topoisomerase I at three main sites, T$^{\downarrow}$G26, T$^{\downarrow}$G48 and C$^{\downarrow}$G73. The first two sites are shared in common with camptothecin whereas site C$^{\downarrow}$G73 is specific to the two lamellarins, as no cleavage occurs with camptothecin at this site even at high concentrations. Site T$^{\downarrow}$G81 corresponds to a strong site for camptothecin but to a weak site for the lamellarins (Fig. 7). These differences in terms of sequence selectivity may translate at the cellular level into different activation of the molecular circuits leading to cell death. Both Lam-D and Lam-M are highly toxic to tumor cells, with a marked selectivity for prostatic cells [124]. Cells resistant to camptothecin are cross-resistant to lamellarins. For example, murine leukaemia P388CPT5 cells expressing a mutated topoisomerase I enzyme are 11-fold less sensitive to Lam-D than the parental P388 cells having an intact enzyme. Recent data attest that topoisomerase I contributes to the cytotoxicity of lamellarin D [125] but there are also indications that this enzyme is certainly not the unique target for this marine alkaloid which is a potent pro-apoptotic agent insensitive to P glycoprotein-mediated drug efflux [126]. Recently, in the course of a study aimed at characterizing the molecular events downstream of topoisomerase I inhibition, we observed that the drug had a direct effect on mitochondria, which may be considered as a second site of action for this compound. Whatever may be the exact mechanism of action of the lamellarins, the data accumulated thus far are sufficiently encouraging to support the development of this family of marine compounds as potential anticancer drugs. The design of lamellarin-amino acid conjugates may furnish a suitable option to reach this objective [125].

The structure of Lam-D is quite different from that of CPT (Fig. 8). They both contain a lactone function but in lamellarin it is an aromatic lactone, which is highly stable. We compared the optimized geometries and atomic charges of Lam-D and CPT obtained from ab initio calculations and superimposed the two structures (Fig. 8). For the pair Lam-D/CPT, the correlation coefficient is 1.27 Å which reflects a very low degree of similarity. The two molecules have very few features in common. Searching for other known topoisomerase I inhibitors [10, 127–129], we could not find a drug having a structure similar to that of Lam-D. One of the partially related non-camptothecin inhibitors of topoisomerase I is the aporphine alkaloid dicentrinone (Fig. 8), extracted from the plant *Ocotea leucoxylon* [130]. In this case, the steric and electrostatic alignment of the two drugs Lam-D and dicentrinone revealed a better (but still modest) degree of similarity with a correlation coefficient of 0.93 Å. The 6*H*-[1]benzopyrano[4′,3′:4,5]pyrrolo[2,1-*a*]isoquinolin-one pentacyclic chromophore of Lam-D represents a novel pharmacophore for topoisomerase I targeting. The planarity of the chromophore appears essential because the 5,6-dehydro analogue of lamellarin D, designated Lam-501, had no effect on topoisomerase I [124]. In all cases, we found that compounds lacking the 5–6 double bond in the B-ring fail to inhibit the enzyme [125]. The planarity of the chromophore is doubtless required to permit stacking interaction with the DNA base pairs. Lam-D is a low affinity DNA intercalating agent [124].

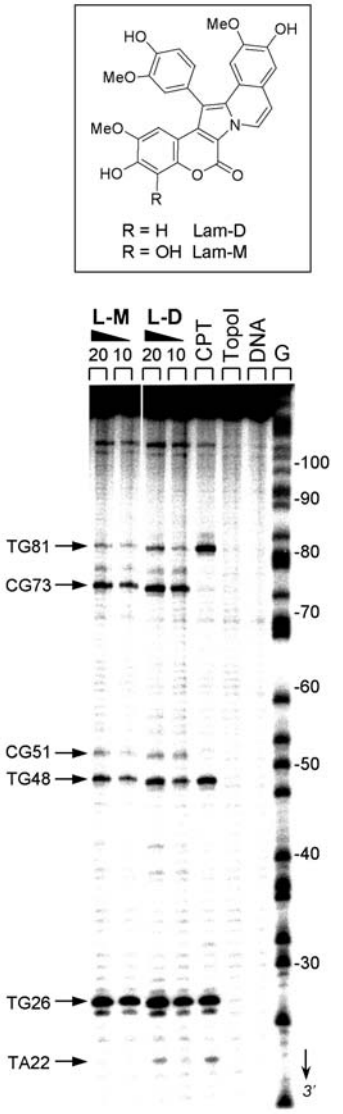


Fig. 7 Cleavage of a 117-bp DNA fragment by human topoisomerase I in the presence of different lamellarin derivatives. The 3′-end labeled fragment (DNA) was incubated in the absence (lane TopoI) or presence of the marine alkaloid at the indicated μM concentration. Camptothecin (CPT) was used at 20 μM. Numbers at the side of the gels show the nucleotide positions, determined with reference to the guanine tracks labeled G. The nucleotide positions and sequences at the cleavage sites are indicated

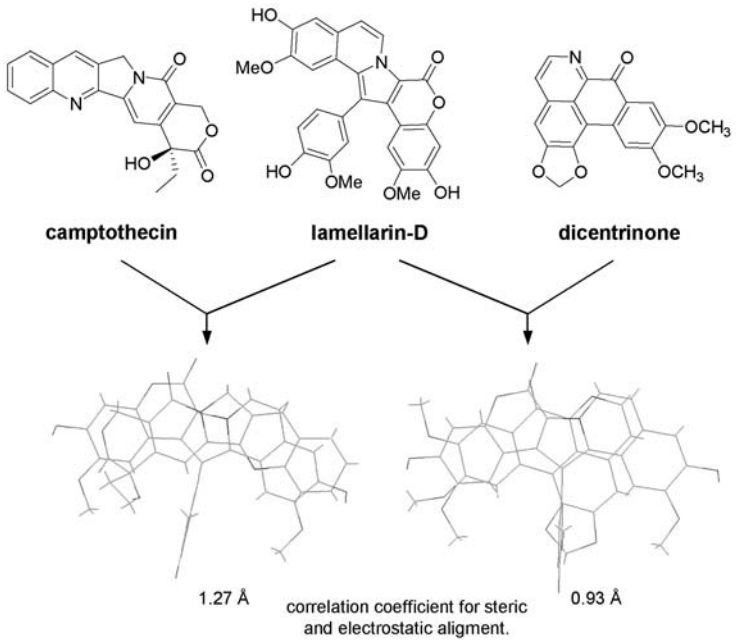


Fig. 8 Structure of camptothecin, lamellarin D and dicentrinone, and superimposition of their structures. Carbon and hydrogen atoms are shown in *green*, nitrogens in *blues* and oxygens in *red*

4 Conclusion

Unsurprisingly, marine organisms have proved to be a rich source of topoisomerase inhibitors. The number of marine products interfering with the functions of topoisomerases I or II has increased significantly over the past 5 years and there is now a quite large chemical diversity, from cyclic peptides to condensed heterocyclic chromophores. In the topoisomerase II family, pyridoacridines provide a large set of tools to manipulate the activity of the enzyme, either stimulating cleavage of DNA through stabilization of covalent topoisomerase II-DNA complexes (e.g., ascididemin) or promoting the catenation of DNA, as is the case with neoamphimedine. The chemical diversity remains quite modest in the topoisomerase I series but the recent disclosure of the mechanism of action of lamellarin D stimulates the search for other marine products capable of promoting DNA cleavage by topoisomerase I. Marine organisms provide a bountiful source of cytotoxic cyclic peptides. It is likely that sansalvamide A is not an isolated case and other compounds of this type will be characterized as topoisomerase I interacting agents.

Chemotherapy regimens using topoisomerase inhibitors have proved successful for the treatment of various cancers. Etoposide and doxorubicin, both

potently inhibiting topoisomerase II, are among the most commonly used anticancer drugs but nevertheless their use in oncology is limited by secondary effects, such as cardiac toxicity for the anthracyclines or leukemogenesis for the epipodophyllotoxins. There is a strong demand for new, better targeted antitumor agents having fewer unwanted side effects and improved efficacy, in particular for the treatment of chemo- and radio-refractory or metastatic disease. One way to achieve this goal is to modify the parent drug structures to try to optimize their anticancer properties while decreasing their unwanted effects. Along this line, novel epipodophyllotoxins (e.g., tafluposide [131, 132]) and novel anthracyclines are constantly being designed and tested. Another possibility would be to seek drugs with a completely different chemical architecture but with a similar pharmacological profile in terms of target binding. Novel chemotypes capable of inhibiting topoisomerase I (lamellarins) or II (neoamphimedine) might turn out to be efficacious against particular cancer cells and hopefully might lack the general toxicities typical of conventional anticancer drugs currently employed in clinical oncology.

Acknowledgements Research in the authors' laboratory was supported by the Ligue Nationale contre le Cancer (Comité du Nord) and the Institut de Recherches sur le Cancer de Lille (IRCL). We thank the Ligue Nationale contre le Cancer for a postdoctoral fellowship to N.D.

References

1. Wang JC (1996) Annu Rev Biochem 65:635
2. Champoux JJ (2001) Annu Rev Biochem 70:369
3. Roca J, Berger JM, Harrison SC, Wang JC (1996) Proc Natl Acad Sci USA 93:4057
4. Hanai R, Caron PR, Wang JC (1996) Proc Natl Acad Sci USA 93:3653
5. Wang Y, Lyu YL, Wang JC (2002) Proc Natl Acad Sci USA 99:12114
6. Goulaouic H, Roulon T, Flamand O, Grondard L, Lavelle F, Riou JF (1999) Nucleic Acids Res 27:2443
7. Chen AY, Liu LF (1994) Annu Rev Pharmacol Toxicol 34:191
8. Pommier Y (1998) Biochimie 80:255
9. Gatto B, Capranico G, Palumbo M (1999) Curr Pharm Des 5:195
10. Bailly C (2000) Current Med Chem 7:39
11. Nitiss JL (2002) Curr Opin Investig Drugs 3:1512
12. Gatto B, Leo E (2003) Curr Med Chem Anti-Canc Agents 3:173
13. Wang HK, Morris-Natschke SL, Lee KH (1997) Med Res Rev 17:367
14. Burden DA, Osheroff N (1998) Biochim Biophys Acta 1400:139
15. Long BH, Balasubramanian BN (2000) Exp Opin Ther Patents 10:635
16. Stewart CF (2001) Cancer Chemother Biol Response Modif 19:85
17. Ding Q, Chichak K, Lown JW (1999) Curr Med Chem 6:1
18. Radiski DC, Radisky ES, Barrows LR, Copp BR, Kramer RA, Ireland CM (1993) J Am Chem Soc 115:1632
19. Carney JR, Scheuer PJ, Kelly-Borges M (1993) Tetrahedron 49:8483
20. Izawa T, Nishiyama S, Yamamura S (1994) Tetrahedron Lett 35:917
21. Izawa T, Nishiyama S, Yamamura S (1994) Tetrahedron 50:13593
22. Schmidt, EW, Harper, MK, Faulkner, DJ. (1995) J Nat Prod 58:1861

23. Venables DA, Concepcion GP, Matsumoto SS, Barrows LR, Ireland CM (1997) J Nat Prod 60:408
24. Casapullo A, Cutignano A, Bruno I, Bifulco G, Debitus C, Gomez-Paloma L, Riccio R (2001) J Nat Prod 64:1354
25. Chang LC, Otero-Quintero S, Hooper JN, Bewley CA (2002) J Nat Prod 65:776
26. (1993) Anticancer Drug Des 8:333
27. Matsumoto SS, Haughey HM, Schmehl DM, Venables DA, Ireland CM, Holden JA, Barrows LR (1999) Anticancer Drugs 10:39
28. Kobayashi J, Cheng J, Nakamura Y, Ohizumi Y, Hirata Y, Sasaki T, Otha T, Nozoe S (1988) Tetrahedron Lett 29:1177
29. Schmitz FJ, DeGuzman FS, Hossain MB, van der Helm D (1991) J Org Chem 56:804
30. Delfourne E, Bastide J (2003) Med Res Rev 23:234
31. Skyler D, Heathcock CH. (2002) J Nat Prod 65:1573
32. Molinski TF (1993) Chem Rev 93:1825
33. Appleton DR, Pearce AN, Lambert G, Babcock RC, Copp BR (2002) Tetrahedron 58:9779
34. Molinski TF, Ireland CM (1989) J Org Chem 54:4256
35. Salomon CE, Faulkner DJ (1996) Tetrahedron Lett. 37, 9147
36. Smith CJ, Venables DA, Hopmann C, Salomon CE, Jompa J, Tahir A, Faulkner DJ, Ireland CM (1997) J Nat Prod 60:1048
37. Bonnard I, Bontemps N, Lahmy S, Banaigs B, Combaut G, Francisco C, Colson P, Houssier C, Waring MJ, Bailly C (1995) Anti-Cancer Drug Des 10:333
38. Bontemps N, Bonnard I, Banaigs B, Combaut G, Francisco C (1994) Tetrahedron Lett 35:7023
39. Torres YR, Bugni TS, Berlinck RG, Ireland CM, Magalhaes A, Ferreira AG, Moreira Da Rocha R (2002) J Org Chem 67:5429
40. McDonald LA, Eldredge GS, Barrows LR, Ireland CM (1994) J Med Chem 37:3819
41. Nilar N, Sidebottom PJ, Carte BK, Butler MS (2002) J Nat Prod 65:1198
42. Eder C, Schupp P, Proksch P, Wray V, Steube K, Muller CE, Frobenius W, Herderich M, van Soest RW (1998) J Nat Prod 61:301
43. Plubrukarn A, Davidson BS (1998) J Org Chem 63:1657
44. Delfourne E, Roubin C, Bastide J (2000) J Org Chem 65:5476
45. Aoki S, Wei H, Matsui K, Rachmat R, Kobayashi M (2003) Bioorg Med Chem 11:1969
46. Copp BR, Kayser O, Brun R, Kiderlen AF (2003) Planta Med 69:527
47. Dassonneville L, Wattez N, Baldeyrou B, Mahieu C, Lansiaux A, Banaigs B, Bonnard I, Bailly C (2000) Biochem Pharmacol 60:527
48. Lindsay B, Barrows L, Copp B (1995) Bioorg Med Chem Lett 5:739
49. Matsumoto SS, Sidford, MH, Holden JA, Barrows LR, Copp BR (2000) Tetrahedron Lett 41:1667
50. Matsumoto SS, Biggs J, Copp BR, Holden JA, Barrows LR (2003) Chem Res Toxicol 16:113
51. Dirsch VM, Kirschke SO, Estermeier M, Steffan B, Vollmar AM (2004) Oncogene 23:1586
52. Burres NS, Sazesh S, Gunawardana GP, Clement JJ (1989) Cancer Res 49:5267
53. Schmitz FJ, Agarwal SK, Gunasekera SP, Schmidt PG, Shoolery JN (1983) J Am Chem Soc 105:4835
54. de Guzman FS, Carte B, Troupe N, Faulkner DJ, Harper MK, Concepcion GP, Mangalindan GC, Matsumoto SS, Barrows LR, Ireland CM (1999) J Org Chem 64:1400
55. Marshall KM, Matsumoto SS, Holden JA, Concepcion GP, Tasdemir D, Ireland CM, Barrows LR (2003) Biochem Pharmacol 66:447
56. Tasdemir D, Marshall KM, Mangalindan GC, Concepcion GP, Barrows LR, Harper MK, Ireland CM (2001) J Org Chem 66:3246

57. de Guzman FS, Schmitz FJ (1989) Tetrahedron Lett 30:1069
58. Skyler D, Heathcock CH. (2001) Org Lett 3:4323
59. McCarthy PJ, Pitts TP, Gunawardana GP, Kelly-Borges M, Pomponi SA (1992) J Nat Prod 55:1664
60. Longley RE, McConnell OJ, Essich E, Harmody D (1993) J Nat Prod 56:915
61. Delfourne E, Darro F, Bontemps-Subielos N, Decaestecker C, Bastide J, Frydman A, Kiss R (2001) J Med Chem 44:3275
62. Ford PW, Davidson BS (1997) J Nat Prod 60:1051
63. Gellerman G, Rudi A, Kashman Y (1992) Tetrahedron Lett 33:5577
64. Gellerman G, Rudi A, Kashman Y (1993) Tetrahedron Lett 34:1823
65. Taraporewala IB, Cessac JW, Chanh TC, Delgado AV, Schinazi RF (1992) J Med Chem 35:2744
66. Kitahara Y, Tamura F, Nishimura M, Kubo A (1998) Tetrahedron 54:8421
67. Bontemps N, Delfourne E, Bastide J, Francisco C, Bracher F (1997) Tetrahedron 53:1743
68. Alvarez M, Feliu L, Ajana W, Joule JA (2000) Tetrahedron 56:3703
69. Alvarez M, Feliu L, Ajana W, Joule JA, Fernandez-Puentes JL (2000) Eur J Org Chem 5:849
70. Legentil L, Bastide J, Delfourne E (2003) Tetrahedron Lett 44:2473
71. Debnath B, Gayen S, Bhattacharya S, Samanta S, Jha T (2003) Bioorg Med Chem 11:5493
72. Delfourne E, Kiss R, Le Corre L, Dujols F, Bastide J, Collignon F, Lesur B, Frydman A, Darro F (2003) J Med Chem 46:3536
73. Stanslas J, Hagan DJ, Ellis MJ, Turner C, Carmichael J, Ward W, Hammonds TR, Stevens MFG (2000) J Med Chem 43:1563
74. Angel de la Fuente J, Jesus Martin M, del Mar Blanco M, Pascual-Alfonso E, Avendano C, Carlos Menendez J (2001) Bioorg Med Chem 9:1807
75. Brahic C, Darro F, Belloir M, Bastide J, Kiss R, Delfourne E (2002) Bioorg Med Chem 10:2845
76. Thale Z, Johnson T, Tenney K, Wenzel PJ, Lobkovsky E, Clardy J, Media J, Pietraszkiewicz H, Valeriote FA, Crews P (2002) J Org Chem 67:9384
77. Delfourne E, Kiss R, Le Corre L, Merza J, Bastide J, Frydman A, Darro F (2003) Bioorg Med Chem 11:4351
78. Yamada T, Iwamoto C, Yamagaki N, Yamanouchi T, Minoura K, Yamori T, Uehara Y, Andoh T, Umemura K, Numata A (2002) Tetrahedron 58:479
79. Concepcion GP, Foderaro TA, Eldredge GS, Lobkovsky E, Clardy J, Barrows LR, Ireland CM (1995) J Med Chem 38:4503
80. Kobayashi M, Muroyama A, Nakamura H, Kobayashi J, Ohizumi Y (1991) J Pharmacol Exp Ther 257:90
81. Sakamoto H, Furukawa K, Matsunaga K, Nakamura H, Ohizumi Y (1995) Biochemistry 34:12570
82. Ito M, Hirata Y, Nakamura H, Ohizumi Y (1999) J Pharmacol Exp Ther 291:976
83. Nakamura M, Kakuda T, Oba Y, Ojika M, Nakamura H (2003) Bioorg Med Chem 11:3077
84. Fujiwara H, Matsunaga K, Saito M, Hagiya S, Furukawa K, Nakamura H, Ohizumi Y. (2001) Eur J Pharmacol 413:37
85. Bae M, Tsuji T, Kondo K, Hirase T, Ishibashi M, Shigemori H, Kobayashi J (1993) Biosci Biotech Biochem 57:330
86. Umemura K, Yanase K, Suzuki M, Okutani K, Yamori T, Andoh T (2003) Biochem Pharmacol 66:481
87. Sogawa K, Yamada T, Sumida T, Hamakawa H, Kuwabara H, Matsuda M, Muramatsu Y, Kose H, Matsumoto K, Sasaki Y, Okutani K, Kondo K, Monden Y (2000) Life Sci 66:227
88. Sogawa K, Yamada T, Sumida T, Hamakawa H, Kuwabara H, Matsuda M, Muramatsu Y, Kose H, Matsumoto K, Sasaki Y, Okutani K, Kondo K, Monden Y (2002) Life Sci 71:2575

89. Hecht SM (2000) Ann NY Acad Sci 922:76
90. Kohn KW, Pommier Y (2000) Ann NY Acad Sci 922:11
91. Liu LF, Desai SD, Li TK, Mao Y, Sun M, Sim SP (2000) Ann NY Acad Sci 922:1
92. Zunino F, Dallavalleb S, Laccabuea D, Berettaa G, Merlinib L, Pratesi G (2002) Curr Pharm Des 8:2505
93. Garcia-Carbonero R, Supko JG (2002) Clin Cancer Res 8:641
94. Ulukan H, Swaan PW (2002) Drugs 62:2039
95. Bailly C (2003) Crit Rev Oncol Hematol 45:91
96. Pizzolato JF, Saltz LB (2003) Lancet 361:2235
97. Bailly C (2003) Targeting DNA and topoisomerase I with indolocarbazole antitumor agents. In: Demeunynck M, Bailly C, Wilson WD (eds), Small molecule DNA and RNA binders, vol. 2. Wiley-VCH, p 538
98. Pilch DS, Yu C, Makhey D, LaVoie EJ, Srinivasan AR, Olson WK, Sauers RR, Breslauer KJ, Geacintov NE, Liu LF (1997) Biochemistry 36:12542
99. Li TK, Bathory E, LaVoie EJ, Srinivasan AR, Olson WK, Sauers RR, Liu LF, Pilch DS (2000) Biochemistry 39:7107
100. Xu Z, Li TK, Kim JS, LaVoie E, Breslauer KJ, Liu LF, Pilch DS (1998) Biochemistry 37:3558
101. Jin S, Kim JS, Sim SP, Liu A, Pilch DS, Liu LF, LaVoie EJ (2000) Bioorg Med Chem Lett 10:719
102. Rangarajan M, Kim JS, Jin S, Sim SP, Liu A, Pilch DS, Liu LF, LaVoie EJ (2000) Bioorg Med Chem 8:1371
103. Ruchelman AL, Singh SK, Ray A, Wu X, Yang JM, Li TK, Liu A, Liu LF, LaVoie EJ (2003) Bioorg Med Chem 11:2061
104. Ruchelman AL, Singh SK, Wu X, Ray A, Yang JM, Li TK, Liu A, Liu LF, LaVoie EJ (2002) Bioorg Med Chem Lett 12:3333
105. Li TK, Houghton PJ, Desai SD, Daroui P, Liu AA, Hars ES, Ruchelman AL, LaVoie EJ, Liu LF (2003) Cancer Res 63:8400
106. Kohlhagen G, Paull KD, Cushman M, Nagafuji P, Pommier Y (1998) Mol Pharmacol 54:50
107. Strumberg D, Pommier Y, Paull K, Jayaraman M, Nagafuji P, Cushman M (1999) J Med Chem 42:446
108. Cushman M, Jayaraman M, Vroman JA, Fukunaga AK, Fox BM, Kohlhagen G, Strumberg D, Pommier Y (2000) J Med Chem 43:3688
109. Jayaraman M, Fox BM, Hollingshead M, Kohlhagen G, Pommier Y, Cushman M (2002) J Med Chem 45:242
110. Fox BM, Xiao X, Antony S, Kohlhagen G, Pommier Y, Staker BL, Stewart L, Cushman M (2003) J Med Chem 46:3275
111. Nagarajan M, Xiao X, Antony S, Kohlhagen G, Pommier Y, Cushman M (2003) J Med Chem 46:5712
112. Antony S, Jayaraman M, Laco G, Kohlhagen G, Kohn KW, Cushman M, Pommier Y (2003) Cancer Res 63:7428
113. Kokoshka JM, Capson TL, Holden JA, Ireland CM, Barrows LR (1996) Anti-Cancer Drugs 7:758
114. Bailly C, Dassonneville L, Colson P, Houssier C, Fukasawa K, Nishimura S, Yoshinari T (1999) Cancer Res 59:2853
115. Takebayashi Y, Pourquier P, Yoshida A, Kohlagen G, Pommier Y (1999) Proc Natl Acad Sci USA 96:7196
116. Damia G, Silvestri S, Carrassa L, Filiberti L, Faircloth GT, Liberi G, Foiani M, D'Incalci M (2001) Int J Cancer 92:583
117. Takebayashi Y, Goldwasser F, Urasaki Y, Kohlhagen G, Pommier Y (2001) Clin Cancer Res 7:185

118. Ohta E, Ohta S, Hongo T, Hamaguchi Y, Andoh T, Shioda M, Ikegami S (2003) Biosci Biotechnol Biochem 67:2365
119. Hwang Y, Rowley D, Rhodes D, Gertsch J, Fenical W, Bushman F (1999) Mol Pharmacol 55:1049
120. Schwartz JJ, Myskowski PL (1992) J Am Acad Dermatol 27:583
121. Ridley CP, Reddy MV, Rocha G, Bushman FD, Faulkner DJ (2002) Bioorg Med Chem 10:3285
122. Lindquist N, Fenical W, Van Duyne GD, Clardy J (1988) J Org Chem 53:4570
123. Bailly C (2004) Curr Med Chem – Anti-Cancer Agents 4(4):363
124. Facompré M, Tardy C, Bal-Mayeu C, Colson P, Perez C, Manzanares I, Cuevas C, Bailly C (2003) Cancer Res 63:7392
125. Tardy C, Facompré M, Laine M, Baldeyrou B, Garcia-Gravalos D, Francesch A, Mateo C, Pastor A, Jimenez JA, Manzanares I, Cuevas C, Bailly C (2004) Bioorg Med Chem12(7): 1697
126. Vanhuyse M, Kluza J, Tardy C, Otero G, Cuevas C, Bailly C, Lansiaux A (2004) (submitted)
127. Kirstein MN, Turner PK, Stewart CF (2002) Cancer Chemother Biol Response Modif 20:99
128. Denny WA, Baguley BC (2003) Curr Top Med Chem 3:339
129. Meng LH, Liao ZY, Pommier Y (2003) Curr Top Med Chem 3:305
130. Zhou BN, Johnson RK, Mattern MR, Wang X, Hecht SM, Beck HT, Ortiz A, Kingston DG (2000) J Nat Prod 63:217
131. Perrin D, van Hille B, Barret JM, Kruczynski A, Etievant C, Imbert T, Hill BT (2000) Biochem Pharmacol 59:807
132. Barret JM, Cadou M, Hill BT (2002) Biochem Pharmacol 63:251

Top Curr Chem (2005) 253: 109–148
DOI 10.1007/b100445

DNA Major Groove Binders: Triple Helix-Forming Oligonucleotides, Triple Helix-Specific DNA Ligands and Cleaving Agents

Christophe Escudé (✉) · Jian-Sheng Sun

USM 0503, Muséum National d'Histoire Naturelle "Régulation et dynamique des génomes", UMR 5153 CNRS-MNHN, U 565 INSERM Département "Régulations développement et diversité moléculaire", 43 rue Cuvier, Case postale 26, 75231 Paris cedex 05, France
escude@mnhn.fr, sun@mnhn.fr

Abstract Nucleic acids are polymorphic macromolecules that can adopt a variety of single-, double- and multi-stranded conformations, which in turn may provide important signals for regulating gene expression, and for maintaining genome integrity and stability. The design of tailor-made molecules that recognize specific sequences in the DNA double helix would provide interesting tools to interfere with DNA information processing and to target genome modifications. The ability to specifically manipulate genetic information processing genome-wide offers a variety of applications in experimental biology as well as gene-based biotechnology and therapeutics. It presents an important challenge in biological and biomedical sciences. This chapter will cover the topic of sequence-specific ligands that bind within and recognize the major groove of the DNA double helix. In particular, it will focus on the development of triple helix-forming oligonucleotides as highly sequence-specific DNA major groove binders, the related triple helix structures, examples of triple helix-specific ligands and DNA cleaving agents, as well as their biological relevance, during the last 10–15 years.

Keywords Triple helix-forming oligonucleotide (TFO) · Triple helix · H-DNA · Triple helix-specific ligands · Triple helix-specific DNA cleaving agents

1 Introduction

Nucleic acids play the most crucial role in the development of biodiversity on earth, in particular the deoxyribonucleic acids naturally assembled in a double-helical structure first discovered in 1953 [1]. This seminal discovery paved the way that led to today's understanding and development of life science and medicine. The importance of the DNA double helix stems from its functions of genetic information storage, heredity, and information processing. From the point of view of physical chemistry and biochemistry, nucleic acids are polymorphic biopolymers that can adopt a variety of single-, double- and multi-stranded conformations, which in turn may provide important signals for regulating gene expression and other metabolic processes involving DNA such as replication, recombination and repair. For these reasons, the design of molecules that can recognize specific sequences on the DNA double helix, either through direct readout of sequence or indirect recognition of the shape and the deformability of the DNA double helix (often a combination of both), would provide interesting tools to interfere with DNA information processing at an early stage of gene expression, and to target genomic modifications. The ability to specifically manipulate genetic information processing genome-wide offers

varied applications in experimental biology and in gene-based biotechnology and therapeutics. It presents an important challenge in biological and biomedical sciences.

Double-stranded DNA (dsDNA) with Watson–Crick base pairs is a biopolymer of relatively uniform structure with a highly negatively charged sugar-phosphate backbone and a core of stacked base pairs whose edges are exposed in the major and minor grooves. Each base pair has a sort of chemical 'signature' characterized by its pattern of functional groups exposed in the DNA grooves. It is this chemical surface along with sequence-dependent variation in DNA structure and flexibility that is recognized by proteins through surface complementarity, forming a series of favourable hydrogen bonding, electrostatic and van der Waals interactions between the protein and the base pairs. In addition, all protein–DNA complex structures contain a large number of contacts with the negatively charged phosphates that include salt bridges with positively charged side chains as well as hydrogen bonds with uncharged main-chain or side-chain atoms in the protein.

The ability of a protein to bind selectively to a particular DNA site in the genome is the foundation upon which transcriptional regulatory pathways are built [2]. The DNA-binding proteins that regulate transcription are capable of selecting the correct binding site out of a vast number of potential sites in the genome. The rapid increase in structural information on protein–DNA complexes has uncovered a remarkable structural diversity in DNA-binding domains, while at the same time revealing a number of common binding motifs which can be combined to target specific sites at the genome level. The effective length of a DNA-binding site depends upon the conformation, size and number of the DNA-binding domains ('reading heads'), and whether the protein forms homodimers or oligomeric interactions with other DNA-binding partners. Several DNA-binding domains have been identified so far, thanks to tremendous efforts made in the field of structural biology in the last two decades: helix-turn-helix (HTH), zinc-finger, basic region-leucine zipper (bZIP) and basic region-helix-loop-helix (bHLH), ribbon-helix-helix, multiple strand β sheets in either major or minor groove (see [3] for review). In addition to direct sequence-specific DNA contacts made by reading heads, extrinsic factors such as ligand binding, homo- and hetero-dimeric protein associations, and interactions mediated via other transcription factors can also modulate the DNA-recognition properties of DNA-binding domains. In each case, these extrinsic factors act as molecular switches to facilitate recognition of cognate DNA through a variety of mechanisms in which they can significantly extend the repertoire of DNA sites recognized by a given DNA-binding domain (see [4] for review).

A large number of low-molecular weight DNA-binding molecules with limited sequence specificity have been synthesized over the last 50 years and have demonstrated their utility as therapeutic drugs especially in the field of cancer chemotherapy (see [5] for review). During the past decade, several more sequence-specific ligands have emerged. Their discovery has benefited from the

understanding of molecular interactions afforded by X-ray and NMR structures of a number of ligand–DNA and protein–DNA complexes. Three families of such DNA ligands which exhibit strong binding and high sequence-specificity for the DNA double helix have been the focus of intensive development, as outlined below:

1. Small synthetic molecules based upon hairpin polyamides that mainly recognize short DNA sequences in the minor groove. Pyrrole, hydroxypyrrole, imidazole and other amino acid derivatives are used to generate the polyamide hairpin such that recognition of all four base pairs depends on a code of side-by-side aromatic amino acid pairings in the minor groove (see [6] and Chap. 1 of this volume for review). They have been shown to be capable of interfering efficiently with the processes of gene expression in vivo. Strenuous efforts have been devoted to further expand the repertoire of DNA sequences that can be targeted by hairpin polyamides in order to achieve gene-specific recognition. The combination of polyamides with polypeptides that are engineered on the basis of naturally-occurring DNA minor groove-binding motifs of proteins should help to attain this ambitious goal. The topic is the subject of Chap. 1 in this book.
2. The crucial structural feature of one of the families of DNA-binding proteins, namely the zinc finger proteins, has been used to design poly-finger peptides displaying naturally occurring zinc finger domains as modular building blocks in a polypeptide chain (see [7] for review). The poly-finger peptide units specifically recognize DNA sequence motifs that are base triplets $(XNN)_n$ (where X=G or T; N=A,T,G,C and $n \leq 6$). For applications in cell biology these protein motifs are produced in situ using a DNA vector that contains all the information to synthesize the polypeptide.
3. Sequence-specific recognition of the DNA double helix by oligonucleotide-directed triple helix formation was first contemplated by two seminal studies 15 years ago in which triplex-forming oligonucleotides (TFOs) were described that can bind to the major groove of oligopyrimidine•oligopurine sequences in duplex DNA via hydrogen bonding with purine bases [8, 9]. Since then oligonucleotide-directed triple helix formation has mainly been exploited to down-regulate or up-regulate transcription of genes, to induce directed mutagenesis, to promote homologous recombination, or to direct modification of genomic DNA at selected gene loci (see [10] for review). Triple helix-forming oligonucleotides are consequently powerful gene-specific tools that can be employed for a wide range of applications in biology, biotechnology and therapeutics.

This chapter will cover the development of triple helix-forming oligonucleotides as a category of highly sequence-specific DNA major groove binders, the related triple helix structures, the development of triple helix-specific ligands and DNA-cleaving agents, as well as their biological relevance, most of which has taken place during the last 10–15 years. In particular, it will summarize recent advances in this field while also highlighting promises and

obstacles that remain to be overcome prior to further biotechnological and therapeutic applications.

2 Oligonucleotide-Directed Triple Helix Formation

Triple-helical structures of nucleic acids, DNA as well as RNA, can form in either inter- or intra-molecular fashion. The first triple helix was described in 1957 in polyribonucleotides [11]. During the three decades that followed their initial discovery, triple helices composed of various polynucleotides remained a curiosity among nucleic acids and were only studied for their structures and physico-chemical properties. It was assumed that only polynucleotides could form stable triple-helical structures (synthetic oligonucleotides were not available until the early 1980s). However, in 1987 it was demonstrated that an oligonucleotide can bind sequence-specifically to an oligopyrimidine•oligopurine sequence in the major groove of the DNA double helix [8, 9]. Such oligonucleotides are nowadays referred to as triple helix-forming oligonucleotides (TFOs). Since then, the triple helix has been the focus of intense research. Oligonucleotide-directed triple helix formation spawned the concept of 'antigene strategy', reflecting its promise as a means of making rationally designed DNA ligands that could interfere with gene expression and other DNA metabolic functions, as well as its promise in biotechnological and in genome-based therapeutic applications.

2.1 Structural Features of Triple Helix Formation

Either polynucleotide- or oligonucleotide-directed triple helix formation is based on sequence-specific recognition of oligopyrimidine•oligopurine sequences, present in double-helical nucleic acids, by a third nucleic acid strand. This third strand binds in the major groove and recognizes the oligopurine strand by establishing a pair of hydrogen bonds with the hydrogen bond donor and acceptor groups available on the major groove edge of purine bases. Different base triplets can be formed through either Hoogsteen or reverse Hoogsteen hydrogen bond formation (Fig. 1a). Some bases in the third strand can adopt either the Hoogsteen or reverse Hoogsteen configuration, for instance, thymine/uracil and guanine. The reverse Hoogsteen base triplets imply an upside-down rotation and a translation (in some cases) of the bases in the third strand as compared to their Hoogsteen configuration. The third strand orientation with respect to the oligopurine strand of the target duplex is parallel if the involved base triplets are in the Hoogsteen configuration, whereas it is antiparallel for the base triplets in the reverse Hoogsteen configuration. This opposite strand orientation is due to the adoption of an *anti* glycosidic conformation in all base triplets.

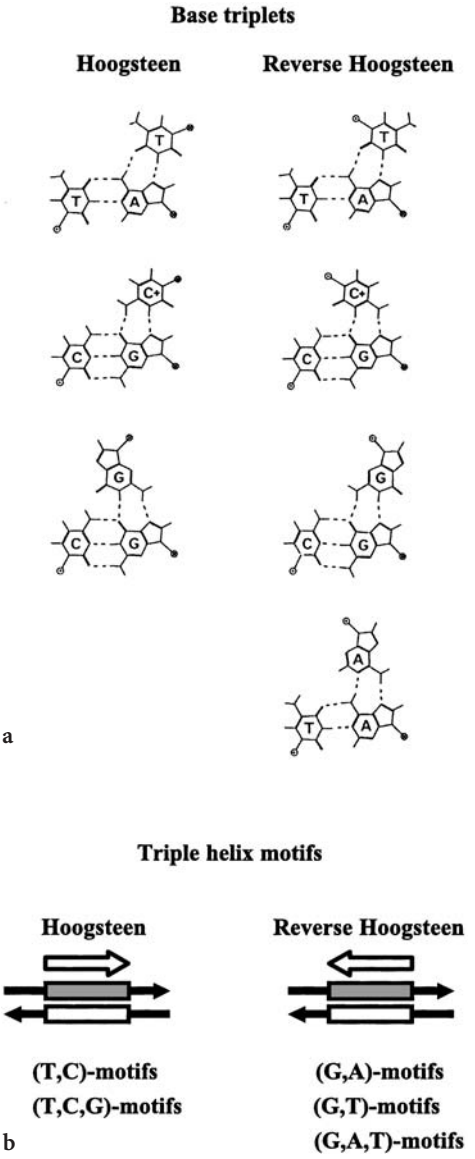


Fig. 1 **a** Base triplets formed by natural bases through either Hoogsteen (*left column*) or reverse Hoogsteen hydrogen bonding (*right column*). **b** Classification of triple-helix motifs according to the involvement of base triplets in either Hoogsteen or reverse Hoogsteen configuration. Oligopyrimidine and oligopurine sequences are shown as *white* and *hatched boxes* respectively, and the third strand oligonucleotide as a *white arrow*. The 5′ to 3′ direction of each strand is shown by the *sense of arrows*. The third strand is parallel to the oligopurine strand in all Hoogsteen motifs, whereas it is antiparallel to the oligopurine strand in all reverse Hoogsteen motifs

In theory, a triple helix can be formed by any combination of base triplets provided they are all in the Hoogsteen or in the reverse Hoogsteen configuration. Such a recognition of oligopyrimidine•oligopurine sequences is a two-letter system. The stability of triple helices depends on the intrinsic stability of base triplets and on whether the triplets are isomorphous or not [12], as well as experimental conditions. Therefore, there are essentially three classes (also called motifs) of triple helices involving natural bases in the third strand (Fig. 1b). They differ in sequence composition and in relative orientation of the phosphodiester backbone of the third strand. In the first class, the third strand binds parallel to the target oligopurine strand, forming T•AxT and C•GxC+ base triplets (the symbols • and x refer to Watson–Crick and Hoogsteen/reverse Hoogsteen base pairing, respectively). The pKa of the imino group of cytosine, which must be protonated, is well below neutral pH, making the formation and stability of a (T,C)-motif triple helix pH-dependent [13]. In the second and third class, a guanine-rich third strand binds to the oligopurine strand of the duplex by reverse Hoogsteen hydrogen bonds, forming either C•GxG and T•AxA base triplets [(G,A)-motif] or C•GxG and T•AxT triplets [(G,T)-motif]. In these classes, the third strand is oriented antiparallel to the target oligopurine strand, but it should be noted that the (G,T)-containing TFO can also bind to the cognate target sequence in parallel orientation, forming Hoogsteen C•GxG and T•AxT triplets, provided there is a very limited number of G-tracts in the target oligopurine strand [14]. In the literature, the (T,C)-motif is often referred to as the pyrimidine-motif, whereas the (G,A)- and (G,T)-motifs are referred to as purine-motifs.

Some hybrid motifs of triple helices are possible and may be of interest. A mixed (T,C,G)-motif can be made of T•AxT, C•GxC+ and C•GxG triplets in the Hoogsteen configuration with parallel third strand orientation with respect to the target oligopurine strand. This (T,C,G)-motif is very useful for replacing strongly pH-dependent contiguous C•GxC+ triplets by pH-independent C•GxG triplets, provided there is a single block of G-tract in the target oligopurine strand [15]. The T•AxT(U), T•AxA and C•GxG triplets in the reverse Hoogsteen configuration can also form a mixed (G,A,T/U)-motif with antiparallel third strand orientation with respect to the target oligopurine strand [16]. It should be kept in mind that all triplex motifs, except the (T,C)-motif, often require a high divalent cation concentration.

It should also be mentioned that single-stranded nucleic acids can be targeted by so called 'clamp'-oligonucleotides or circular oligonucleotides, which form a triple helical structure. They will not be discussed in detail here as they have been reviewed elsewhere [17].

2.2 Thermodynamics, Kinetics and Mechanism of Triple Helix Formation

DNA triple helices offer exciting new perspectives towards oligonucleotide-directed inhibition of gene expression. Purine-motif triple helices appear to be

the most promising for stable binding under physiological conditions compared to those based on the pyrimidine-motif, which forms at relatively low pH. Many thermodynamic studies on triple helix formation have been reported in the literature (see [18] for review). There are, however, very little data available for comparison of the relative stabilities of the different motifs of triple helices under identical conditions. A recent spectroscopic study was made using a designed model system allowing the establishment of competition between purine-rich TFOs (in the reverse Hoogsteen motif) and pyrimidine-rich TFOs (in the Hoogsteen motif) targeting the same Watson–Crick duplex in order to compare directly the thermodynamic parameters of triple helix formation by different motifs with the same target sequence [19]. A low ΔH was found for purine-motif triplexes (–0.1 and –2.5 kcal/mol per base triplet for the (G,A)- or (G,T)-motif, respectively) whose formation is therefore entropy driven, whereas pyrimidine-motif triplex formation is enthalpy driven (–6.1 kcal/mol per base triplet for the (T,C)-motif). This observation explains why purine-motif triplexes are usually associated with a weak hypochromism at 260 nm and may still be formed at high temperatures. The design of this system allowed two TFOs to compete for triplex formation at the same target sequence. The displacement of the purine-rich strand [(G,A)- or (G,T)-containing strand] by a pyrimidine strand could be observed at neutral pH upon lowering the temperature.

Triple helix formation is usually associated with slow kinetics. The association rate constant is about 10^3 mol^{-1} s^{-1}, which is about 2–3 orders of magnitude slower than that of double helix formation, whereas the dissociation rate constant is also slow, making the triple helix a rather long-lived complex once formed [20, 21]. A detailed kinetic study of triple helix formation has recently been carried out by surface plasmon resonance (BIAcore) [22]. Three 15-mer triplex systems, where the base composition was kept constant and the different sequences were designed by permutation of base triplets, were investigated. The rate constants and activation energies were evaluated by comparison with the binding constant and enthalpy calculated from UV-melting profile analysis. The replacement of a T•A base pair by a C•G pair at either the 5′- or the 3′-end of the target sequence and the deletion of the terminal nucleotide at either the 5′- or the 3′-end of the TFO were performed to assess mismatch and deletion effects, as well as to delineate the mechanism of triple helix formation under various conditions of pH, cation concentration and ionic strength. Taken together, all the data consistently showed that the association rate constant is governed by the nature of base triplets (sequence) on the 5′-side of the triple helix (referred to as the 5′-side of the target oligopurine strand), providing evidence that the reaction pathway for triple helix formation in the pyrimidine-motif proceeds from the 5′-end to the 3′-end of the triple helix according to the 5′→3′ directional nucleation-zipping model.

Further experiments carried out on other motifs involving the T•AxT, T•AxA and C•GxG base triplets in the reverse Hoogsteen configuration led to similar conclusions ([23] and Alberti et al., unpublished data). They are consistent with

Fig. 2 Structure of the duplex-triplex junction from a previously published energy-minimized model [22]. The nucleotides involved in the base triplet are coloured in *blue*, *red* and *yellow* for Watson–Crick base pair and TFO, respectively. The Watson–Crick base pair next to the junction is coloured in *green*. The 5′-axis of the oligopurine strand extends out of the plane. The 5′- and 3′-junctions are shown on the *left* and *right*, respectively. Hydrogen atoms are omitted for clarity

previous reports suggesting that the nature of the bases at the 3′-side of a TFO in purine-rich motifs is a key factor affecting triplex stability (the third strand orientation of these triplexes is opposite to that in pyrimidine-motif triplexes) [24]. Therefore, it seems that the 5′→3′ directional nucleation-zipping mechanism is very likely a general feature of all triple helix formation.

A plausible explanation of such a directional nucleation-zipping model comes from a molecular modelling study by conformational energy minimization of the structure of the junction between a double and a triple helix. Figure 2 shows that the first base pair on the 5′-side of a triplex overlaps well with all three bases in the first base triplet at the 5′-junction, whereas the base pair on the 3′-side of the triplex stacks only with the Watson–Crick base pair of the base triplet at the 3′-junction. The base stacking at the 5′-junction between the preceding base pair and the first base triplet was about 0.7 kcal/mole stronger than that between the last base triplet and the next base pair at the 3′-junction. The difference in base stacking interaction is due to the right handedness of the double-helical structure. As a consequence of such structural and energetic features, nucleation at the 5′-end should be favoured since once the very first base triplet is formed, formation of a second base triplet (in the 5′→3′ direction) would not disrupt the stacking between the first base triplet and its 5′-side base pair neighbour, whereas progression in the opposite direction would cause a partial rearrangement of the base pair with respect to the 3′-first base triplet and rearrangement between the second triplet with its 5′-base pair, and so on.

2.3 Improvement of Physical Chemistry and Biochemical Properties of TFOs

The chemical and physical properties of TFOs made of natural nucleotides can be incompatible with biological applications, mainly due to:

1. The requirement for protonation of cytosine in pyrimidine-motif TFOs, especially adjacent cytosines.
2. Self association of purine-rich motif TFOs which can form G4-tetraplex or other structures at the physiological concentration of Na^+/K^+.
3. The negatively charged sugar-phosphate backbone of TFOs requires a relatively high concentration (5–10 mM) of divalent cations such as Mg^{2+}, which is higher than is thought to be available inside cells. The charge repulsion is likely to be responsible, at least in part, for the slow kinetics of triple helix formation.

During the past decade, many efforts have been devoted to modifying the chemical structure of TFOs mainly in order to improve their binding to the cognate target sequences, and also to confer resistance to nucleases in a cellular environment. In general, base modifications can overcome some of the limitations outlined in points 1 and 2, and modifications to the sugar-phosphate backbone provide means for improvement regarding points 2 and 3. The attachment of an intercalating agent, usually at the 5′-side of a pyrimidine TFO or the 3′-side of a purine-rich TFO, can stabilize triple helix formation [25]. The development of triple helix-specific ligands is also an alternative for efficiently stabilizing triplexes (cf. Sect. 5 and 6 of this chapter).

2.3.1 Base Modifications

The use of 5-methylcytosine (Fig. 3a) partially alleviates the pH restriction of TFOs in the pyrimidine-motif, mainly because the methyl group enhances base stacking [26] and/or excludes water molecules from the major groove [27]. A number of other cytosine analogues have been synthesized (see [12, 28] for reviews). One particular cytosine analogue is very promising, 2-aminopyridine (2AP) (Fig. 3b) [29, 30], which has a pKa of 6.86 and can be readily protonated at neutral pH due to the polyelectrolyte property of nucleic acids, which condense cations including protons (thus the local pH is lower than that in bulk buffer). It is particularly useful for substituting adjacent cytosines, which cannot be fully protonated at neutral pH due to proton competition. In addition, the positive charges provided by protonated cytosine analogues make an important contribution to triplex stability.

5-Propynyl-uracil (Fig. 3c) is an analogue of thymine that provides enhanced base stacking when it is used to replace adjacent thymines. In addition, it renders pyrimidine-motif triplex structures less Mg^{2+} sensitive [31]. Recently, 5-aminopropynylargyl-deoxyuridine derivatives bearing one or two guanidinium groups have been synthesized and shown to slightly stabilize the triple helix but the effect was not related to their content [32]. For reasons that are unclear, the incorporation of 5-propynyl-cytosines in pyrimidine TFOs destabilizes triple helices.

Some (G,A)-containing TFOs form very stable triplexes and their stability is highly dependent on the target sequence; high G content seems to be manda-

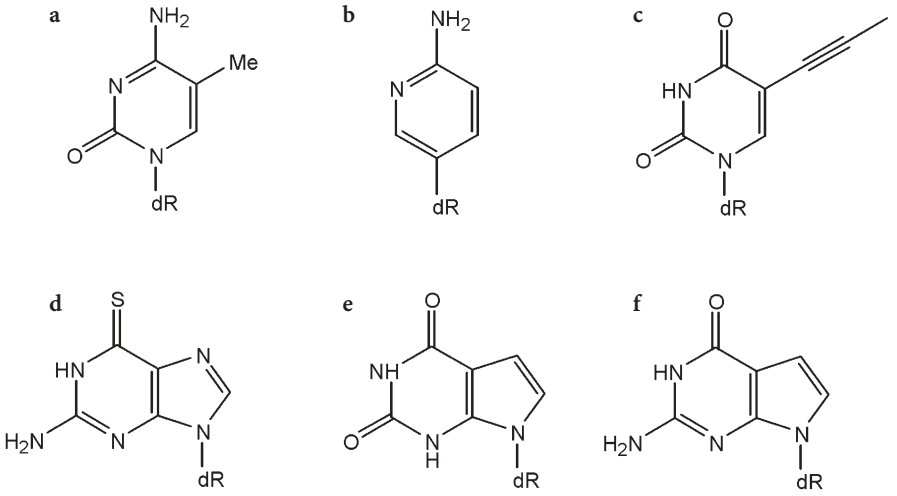


Fig. 3a–f Examples of modified nucleobases used in TFO: **a** 5-methylcytosine, **b** 2-aminopyridine, **c** 5-propynyluracil, **d** 6-thioguanine, **e** 7-deazaxanthosine and **f** 7-deazaguanine

tory [33]. At present, the most severe limitation of G-rich TFOs is their ability to form competing self-associated structures (intermolecular homoduplex or G-quadruplex) [34]. Limited incorporation of 6-thioguanine (Fig. 3d) and 7-deazaxanthine (Fig. 3e) [35] can prevent self-associated structures of purine-rich TFOs but slightly decreases TFO binding. Partial incorporation of 7-deazaguanine derivatives (Fig. 3f) within a block of consecutive Gs in the (T,C,G)-motif can also effectively prevent the formation of self-associated structures without notable effect on TFO binding [36]. Further chemical modifications should increase the intracellular potency of these G-rich TFOs.

2.3.2 Sugar Modifications

Modifications at the 2′-O position influence sugar pucker and the conformation of nucleic acids, and have been important in the development of bioactive TFOs. Structural studies have indicated that 2′-O-modifications stabilize the C3′-endo sugar conformation, which is optimal for triple helix formation by pyrimidine oligonucleotides and provokes the least perturbation of the underlying duplex [37]. It had been shown that RNA (2′-OH, Fig. 4a) and 2′-O-methyl (2′-OMe, Fig. 4b) TFOs formed more stable pyrimidine-motif triple helices than the corresponding DNA strands [38–40]. Other 2′-modifications such as 2′-O-aminoethyl (2′-AE, Fig. 4c) have been tested in combination with 2′-OMe residues [41]. It was found that TFOs containing 2′-AE were much better than 2′-aminopropyl residues. This is ascribed to the ammonium group in 2′-AE being well positioned to interact with the adjacent phosphate group in the

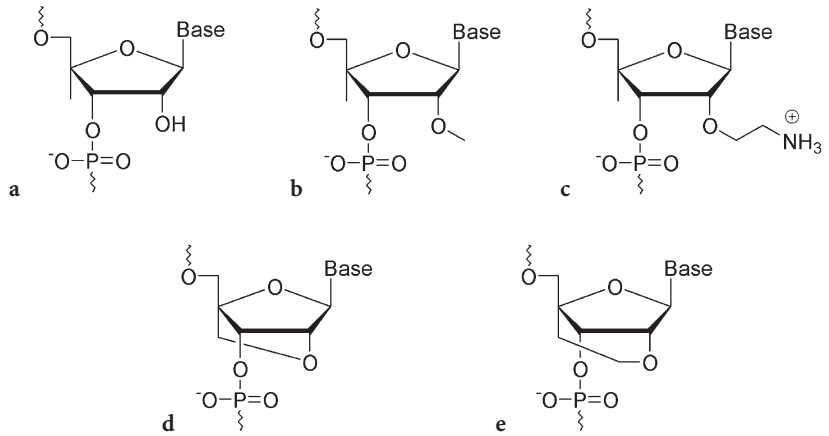


Fig. 4a–e Examples of modified sugars used in TFO: **a** 2′-hydroxyl (RNA), **b** 2′-O-methyl, **c** 2′-O-aminoethyl, **d** 2′-O,4′-C-methylene LNA and **e** 2′-O,4′-C-ethylene LNA

target duplex [42]. TFOs containing 2′-OMe and 2′-AE ribose substitutions in varying proportion were more stable than the underlying duplex, regardless of 2′-OMe/2′-AE ratio [43]. They exhibited enhanced kinetics of triple helix formation and greater stability of the resultant complex at physiological pH and low Mg^{2+} concentration. Enhanced biological activity of 2′-AE pyrimidine TFOs was observed, but extensive incorporation was deleterious [44].

LNA (Fig. 4d) is an RNA derivative in which the ribose ring is constrained by a methylene linkage between the 2′-oxygen and the 4′-carbon (see [45] for review). This bridge results in a locked 3′-*endo* sugar puckering and thus reduces the conformational flexibility of the ribose as well as increasing the local organization of the phosphate backbone. This conformational restriction enhances binding affinity for complementary sequences with high sequence selectivity. Partial incorporation of 2′-*O*,4′-*C*-methylene linked locked nucleic acid (LNA) residues in pyrimidine TFOs has been shown to enhance triple helix formation significantly, whereas the full length LNA TFO failed to form a stable triplex [46]. A modelling study suggests that this can be ascribed to the difference in helical parameters between the rather rigid and pre-organized LNA TFO and the DNA double helix, due to the short 2′-O,4′-C-methylene bridge which causes an unusually high amplitude of sugar puckering (about +50%) of LNA residues. It has been established that alternating LNA substitution every 2–3 nucleotides in pyrimidine TFOs is mandatory for enhancing TFO binding whereas the use of 5-methylcytosine LNA residues should be limited under neutral pH conditions due to their high sensitivity to pH change (Sun et al., data in publication).

Residues containing a longer bridge such as 2′-O,4′-C-ethylene (named ENA, Fig. 4e) have recently been synthesized. The ethylene bridge should reduce the amplitude of sugar puckering while retaining a locked C3′-*endo* conformation

which could lead to better base stacking interactions. A recent report has shown that a pyrimidine TFO fully substituted with 2′-O,4′-C-ethylene residues can indeed form a stable triple helix [47]. Further development of these conformationally constrained nucleotide derivatives will provide a useful chemical approach for DNA major groove targeting.

2.3.3 Phosphodiester Backbone Modifications

Modifications to the backbone of TFOs have often been made in attempts to enhance their affinity for the target duplex sequence and to improve their resistance toward nucleases. Some examples are outlined below.

N3′→P5′ phosphoramidate (Fig. 5a) oligonucleotides have been demonstrated to bind very strongly to double-stranded DNA by triplex formation in the Hoogsteen configuration, whereas reverse Hoogsteen interaction was not detectable with the investigated sequences [48]. A kinetic study showed that the triplex formed with a phosphoramidate TFO has a long lifetime (Sun et al., unpublished data). This can explain why phosphoramidate TFOs can efficiently inhibit transcription at the elongation step in vitro and in cell cultures.

Other modifications have been performed by making the backbone neutral or positive in order to eliminate electrostatic repulsion between the TFO and the duplex. These modifications often involve the replacement of a non-bridging oxygen in the phosphodiester backbone. Two diastereoisomers are possible. TFOs with stereo-uniform backbone are difficult to synthesize. As a result, n-mer oligonucleotides are a mixture of 2^{n-1} stereoisomers. In the pyrimidine-motif, one of the stereoisomers has a favourable effect on the interaction between the TFO and the duplex. In the purine-motif, both stereoisomers seem to be well tolerated. Oligonucleotides containing α anomers of nucleotides can also form triple helices. It is possible to combine the use of α anomers of the sugar with backbone modifications. In particular, α anomers tolerate well stereoisomer mixtures in the pyrimidine-motif.

Among oligonucleotides containing non-ionic phosphoramidate linkages, pyrimidine α-oligodeoxynucleotides with *N*-(methoxyethyl) or *N*-(dimethoxyethyl) phosphoramidate linkages (Fig. 5b) [49] and purine β-oligodeoxynucleotides with *N*-(methoxyethyl) phosphoramidate linkages (Fig. 5c) [50] are examples of backbone modifications which enhance the stability of triple helices.

Pyrimidine oligonucleotides containing alternating anionic and stereo-uniform cationic *N*-(dimethylaminopropyl)phosphoramidate linkages (zwitterionic TFO) were able to form stable triple helices even under low salt conditions. One stereoisomer bound with higher affinity, and the other stereoisomer with lower affinity, than the corresponding all-phosphodiester oligonucleotides. Interestingly, one of them preferred to form base triplets in the Hoogsteen configuration, while the other favoured base triplets in the reverse Hoogsteen configuration [51].

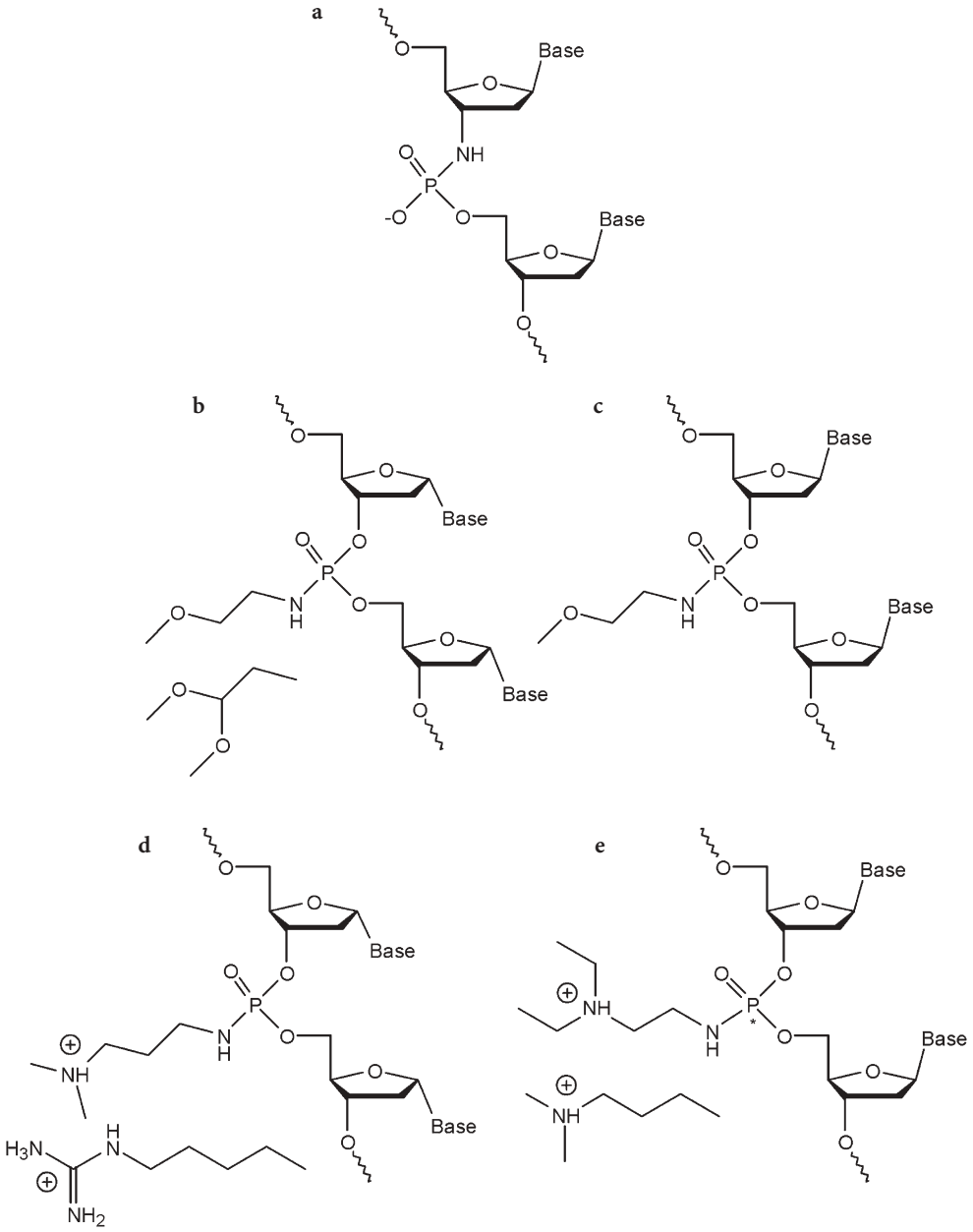


Fig. 5a–e Examples of modified sugar-phosphate backbones used in TFO: **a** β-oligonucleotide with N3′→P5′ phosphoramidate linkage, **b** α-oligonucleotide with *N*-(methoxyethyl) or *N*-(dimethoxyethyl)phosphoramidate linkage, **c** β-oligonucleotide with *N*-(methoxyethyl) phosphoramidate linkage, **d** α-oligonucleotide with *N*-(dimethylaminopropyl) or *N*-(guanidiniumbutyl)phosphoramidate linkage and **e** β-oligonucleotide with *N*-(diethylaminoethyl) or *N*-(dimethylaminopropyl)phosphoramidate linkage

As expected, the oligonucleotides having a cationic backbone form very stable triple helices. For example, pyrimidine α-oligodeoxynucleotides with *N*-(dimethylaminopropyl) and *N*-(guanidiniumbutyl) phosphoramidate linkages (Fig. 5d) in the pyrimidine-motif (Michel et al., personal communication) and purine β-oligodeoxynucleotides with *N*-(diethylaminoethyl) and *N*-(dimethylaminopropyl) phosphoramidate linkages (Fig. 5e) [50, 52] provide examples of backbone modifications which form more stable triplex structures than those formed with unmodified phosphodiester β-oligodeoxynucleotides.

As anticipated, triple helix formation by TFOs with neutral and cationic backbones are much less dependent upon ionic conditions, and less prone to form self-associated structures, than the corresponding TFO having a natural backbone. In addition, the oligonucleotides with a full length cationic backbone could be taken up into cells unaided, that is without transfection [53]. All these features make the cationic oligonucleotides very promising for future biological applications.

2.3.4
PNA and GPNA

Peptide (or polyamide) nucleic acid (PNA, Fig. 6a) is a mimic of nucleic acid in which the natural sugar-phosphate backbone has been replaced by achiral *N*-(2-aminoethyl) glycine units (Fig. 6). Binding of PNA to double-stranded DNA occurs mostly by a strand invasion mechanism at oligopyrimidine•oligopurine sites (see [54] for a review). Two PNA strands can form a triple helix by binding to a single strand of DNA. These two PNA strands can be linked together to form the so-called bis-PNAs. Triplex-forming bis-PNAs have elicited intracellular effects, which might indicate that PNA strand invasion is more efficient inside cells [55, 56]. Cellular uptake of PNA still presents major problems, however. Standard adjuvant procedures using cationic carriers are inefficient for uncharged PNA; permeabilization or electroporation seem to be more appropriate techniques but more original and less invasive approaches

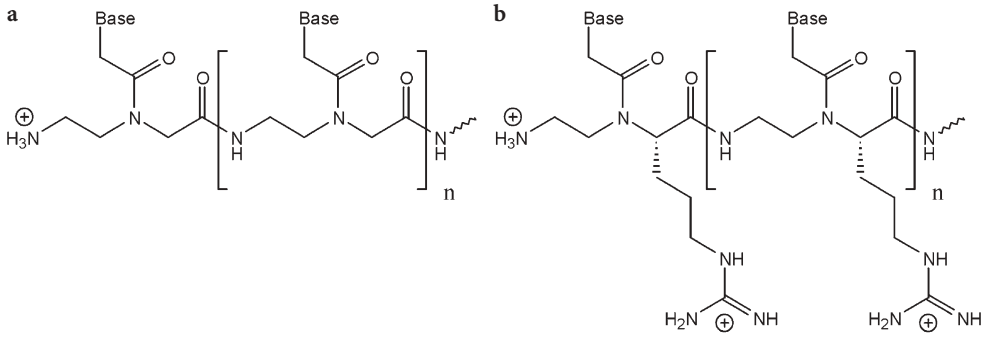


Fig. 6a,b **a** *N*-(2-aminoethyl) glycine PNA, **b** *N*-(2-aminoethyl) arginine PNA (GPNA)

are needed to fully exploit the properties of PNA, for example, the utilization of PNA-peptide conjugates [57].

Guanidine-based peptide nucleic acid (GPNA, Fig. 6b) in which arginine instead of glycine was used in the synthesis of the PNA backbone has recently been prepared [58]. GPNA is much more water-soluble than PNA. It has been shown that GPNA can form a triple helix at least as stable as PNA. Despite the positively charged side chains, GPNA retains the high level of sequence specificity normally exhibited by PNA. In addition, GPNA, like cationic oligonucleotides, enjoys remarkable cell uptake properties.

2.4
Expansion of DNA Sequence Repertoire for Triple Helix Formation

Oligonucleotide-directed triple helix formation is restricted to long oligopyrimidine•oligopurine sequences (at least 15-bp and longer). Such sequences are statistically over-represented in genomic DNA [59]. Nevertheless, in order to be able to selectively target a particular genomic site within a given gene, it is highly desirable to expand the sequence repertoire for triple helix formation. Several strategies have been explored which are outlined below.

Any interruption of oligopyrimidine•oligopurine sequences by even a single A•T or G•C base pair strongly destabilizes triple helix formation. The effect of base triplet mismatches on triple helix stability has been extensively investigated mostly in the (T,C)-motif [60–63] but also in the (G,T)-motif [64]. In general, the destabilization engendered by triplet mismatches depends upon: (1) the nature of base triplet; (2) the location of the mismatch; and (3) its neighbourhood. A central mismatch is more destabilizing than an end mismatch. A central single or double mismatch decreases the melting temperature of the triple helix by about 10°C or more than 20°C, respectively (at least). It was found that mismatched A•TxG and G•CxY (Y=T or C) triplets in the Hoogsteen (T,C)-motif, and A•TxT and G•CxT triplets in the reverse Hoogsteen (G,T)-motif, are the least destabilizing mismatched triplets. In addition, an A•TxG triplet in the (T,C)-motif and a G•CxT triplet in the (G,T)-motif still retain some elements of selectivity as judged by their free energy, which is about 0.8–1.5 kcal mol^{-1} lower than that of other mismatched triplets involving a similar purine•pyrimidine base pair inversion in the target sequence [63, 64]. The mismatched triplets can be stabilized by incorporating an intercalator in the TFOs. It has been shown that the incorporation of an acridine, a well known double-helical DNA intercalator, at the 3′-side of a mismatch site in TFOs (which corresponds to a 5′-pyrimidine-purine-3′ step in the target dsDNA, a favourable site for intercalators) reduced mismatch destabilization to about 5°C and 10°C for a single and double central mismatch, respectively [65, 66].

DNA sequences composed of oligopyrimidine•oligopurine tracts which alternate can also be targeted by TFOs. As several short triple helices form on each oligopyrimidine•oligopurine tract, these short TFOs can be linked together to achieve cooperative binding. Therefore, a TFO containing several triple

helix-forming domains can bind to a target double-stranded DNA composed of adjacent and alternating oligopyrimidine•oligopurine tracts, zig-zagging along the major groove and switching from one strand to the other at the 5′-purine-pyrimidine-3′ or 5′-pyrimidine-purine-3′ steps (see [67] for a review). It transpired that the binding of an alternate-strand TFO at the 5′RpY3′ junction is higher than that at the 5′YpR3′ junction, probably due to better base stacking interaction at the 5′RpY3′ junction. In general, the extent of cooperative binding between different TFO domains was weak. This was not well understood until recently when new insights were gained into the mechanism of triple helix formation: triplex formation at two adjacent and alternating oligopyrimidine•oligopurine tracts should require two nucleation events (either in a convergent or divergent manner) according to the 5′→3′ directional nucleation-zipping mechanism of triple helix formation [22].

Another ambitious approach to expand the dsDNA recognition code by TFOs is the design and synthesis of new nucleobases containing extended heterocycles which can form specific hydrogen bonds with all hydrogen bond-forming sites available in the major groove side of DNA double helix. Despite the fact that some such base analogues seem to recognize specifically A•T or G•C base pairs in the context of an oligopyrimidine•oligopurine sequence [68–70], it remains a great challenge to design and synthesize new nucleotide analogues and/or new backbones that would allow modified TFOs to target any sequence of double-helical DNA from the major groove by unambiguously recognizing all four base pairs.

2.5 Biological Applications of TFOs

As described above some TFOs have been so extensively modified, especially in their backbone, that it becomes more appropriate to designate them triplex-forming molecules (TFMs). The high sequence-specificity and affinity of TFMs have been exploited to down-regulate or up-regulate gene transcription, or to induce directed mutagenesis and to promote homologous recombination as well as to direct modification of genomic DNA at selected gene loci. Some outstanding achievements are outlined below (see [10, 71] for review).

2.5.1 Triplex-Mediated Modulation of Transcription

Inhibition of transcription mediated by TFOs has been reported in vitro and in cell cultures, either through competition with the binding of transcription factors at promoter sites during the initiation phase or by physically arresting the transcription machinery during the elongation phase [47, 72, 73]. The demonstration of the postulated triplex-mediated mechanism was provided by the use of mutant oligopyrimidine•oligopurine target sequences. Even though triplex formation is presently restricted to oligopyrimidine•oligopurine se-

quences, the fact that these sites can be located downstream of the transcription start site, including exons and introns, considerably increases the number of potential target genes for triplex-based strategies.

The accessibility of TFOs to transcriptionally active target genes in a chromatin environment was demonstrated by the use of psoralen-TFO or nitrogen mustard-TFO conjugates [74, 75]. The observation of triplex-mediated down-regulation of expression of genes integrated into cellular chromosomes or endogenous genes on native chromosomes furnished additional proof that TFOs can successfully access their target sequences [76, 77]. It was reasoned that chromatin is probably decondensed in the vicinity of transcriptionally active genes, thus facilitating the access of TFO to its target sequence(s). According to this view the accessibility of TFOs depends on the status of chromatin condensatio which is dynamic in nature, notably dependent upon the cell cycle.

Activation of transcription has been achieved by using a TFM covalently tethered to activation domains [78]. This suggests that a TFM-effector conjugate may be capable of selectively up-regulating expression of a target gene, provided that the conjugates are appropriately designed and optimized. Evidently triplex-mediated down- or up-regulation of expression of selectively targeted genes could be useful for dissecting biological mechanisms or altering phenotypes of cells and organisms at will.

2.5.2 Triplex-Directed Gene Manipulation

Directed mutagenesis in mammalian cells has been the focus of intense research because of its promising application for gene correction and engineering. Several methods based on gene targeting via TFO have been reported with the aim of providing rationally designed molecules as well as relevant protocols that can promote efficiently targeted homologous recombination and gene correction. TFMs have been used to target mutations or to promote homologous recombination at selected sites in cells and in vivo through the induction of a DNA repair-dependent process (see [79, 80] for review). It has been shown that TFMs either alone (both TFO and bis-PNA) or tethered to a DNA-damaging agent (e.g. TFO-psoralen conjugates) can induce a spectrum of point mutations and deletions/insertions around triplex sites located either in an extrachromosomal vector or on chromosomal DNA, with low but measurable frequency. They can also stimulate homologous recombination and gene conversion within and between extrachromosomal vectors at selected sites in mammalian cells in vivo [81]. Triple helix-induced mutagenesis seems to be associated with nucleotide excision repair or transcription-coupled repair pathways, though a full understanding of the precise mechanism(s) will require further investigations.

Recently it has been shown that the efficiency of targeted mutagenesis is sensitive to the cell cycle status and is variable across the cycle with the greatest activity in S phase [82]. This appeared to be the result of differential TFO

binding as measured by cross-link formation. Targeted cross-linking was low in quiescent cells but substantially enhanced in S phase cells with adducts in approximately 20–30% of target sequences. Some 75–80% of adducts were repaired faithfully, whereas the remaining adducts were converted into mutations (>5% mutation frequency).

TFOs have been used to guide homologous donor DNA (DD) to its intended target site on an extrachromosomal gene and to position it for efficient information transfer via homologous recombination and/or gene conversion [83, 84]. In this approach, TFO was covalently tethered to DD through a linker. The effectiveness of the TFO-DD conjugate could be explained by: (1) an increase in the local concentration of DD and (2) a stimulation of DNA repair by triple helix formation that could provoke recruitment of proteins involved in homologous pairing, strand exchange and/or recombination [85]. Recently, a new method named GOREC (for Guided hOmologous RECombination) has been described [86]. It shares a similar gene targeting strategy but differs by the noncovalent attachment (Watson–Crick base pairing) of the donor DNA to an adaptor oligonucleotide that is covalently linked to the TFO. This modular concept allows the experimenter to guide not only an oligonucleotide (ODN, RDO) but also a DNA fragment (either single- or double-stranded) to the target site for homologous replacement. Therefore, the target site is not restricted to the vicinity of the triplex site as is the case for the TFO-DD conjugate: it can be at a remote site. It has even been shown that the TFO accelerates D-loop formation between DD and target DNA in the presence of RecA protein, and both triplex and D-loop are formed in the joint molecule (an obligatory intermediate and a limiting step in the homologous recombination process) in vitro. Recent data have shown that PCR-amplified 759-nt single-stranded or 759-bp double-stranded donor DNA fragments guided by a TFO were able to restore the expression of a mutant eGFP gene by correcting a stop codon on a transfected plasmid in CHO cells [87]. It remains to determine quantitatively the efficiency of the GOREC method as a means of performing gene correction at endogenous genes in the context of chromatin.

2.5.3 Triplex-Based Molecular Tools

Triple helices are very attractive for biotechnological applications because they open new ways to manipulate double-stranded DNA in a sequence-specific manner (see [88] for a review). They have been successfully applied to purify plasmids used in gene therapy [89]. The method is based on a triple helix affinity chromatography technique. It allows elimination of toxic chemicals and purification of plasmids, provided they contain a triplex site. Plasmids can be modified for various purposes in gene therapy. Fluorescent labelled plasmids are used to investigate, in real time and in living cells, the factors affecting plasmid intracellular traffic and bio-distribution as well as those influencing the potency of gene delivery systems for gene transfer into cell cultures and in

vivo. Targeting moieties, such as nuclear localization signals or other peptides, can be linked to plasmid DNA in order to improve its in vivo bio-availability and expression. PNAs have been used for plasmid labelling [90] and targeting [91]. The circularization of an oligonucleotide around double-stranded DNA such as a plasmid is possible thanks to the triple helix: after triplex formation the free ends of the TFOs that are not involved in DNA binding can be ligated, thus creating a circular DNA molecule catenated to the plasmid [92]. Such 'padlock' oligonucleotides provide a new method for attaching a molecular tag in an irreversible way without covalent linkage to supercoiled plasmid DNA. The method has been used for the detection of specific sequences by electron [93] or optical [94] microscopy, as well as for linking peptides to plasmids [95].

It is now well established that a triplex can be formed within cells, but the intracellular fate of such a complex is still largely unknown. In particular, it would be interesting to direct efficient DNA damage to the site where a triplex is formed. This should enhance the efficacy of the antigene approach and could be considered as an equivalent to the RNase H-mediated RNA cleavage often involved in antisense technology. Different strategies have been contemplated. TFOs have been conjugated to DNA-reactive entities such as adduct inducers (psoralen), alkylating agents (nitrogen mustard) and cleaving reagents (EDTA•Fe^{2+}, OP•Cu^{2+}) (see [96] for a review). TFOs can be covalently tethered to a substance (e.g. a topoisomerase inhibitor) that is able to recruit a cellular enzyme such as topoisomerase I or II having DNA cleavage activity thereby inducing irreversible damage at the triplex site [97–99]. Covalent linkage between topoisomerase poisons and a short TFO may mitigate undesirable toxic side-effects of the free drugs when used in cancer chemotherapy, by conferring upon them sequence specificity.

3
Intramolecular DNA Triple Helix

Intramolecular triple helices, generally named H-DNA, were first found in plasmid DNA at the mirror repeats of oligopyrimidine•oligopurine sequences while subjected to physical constraints and/or low pH (see [100] for review). The plasmids underwent a conformational rearrangement involving the disruption of half of the symmetry-related double helix and the folding back of one of the resulting single strands so as to form an intramolecular triple-helical structure (Fig. 7). These structures can be divided into several different categories depending on the triplex structural motifs. It is believed that the essential constraint in natural DNA referred to as negative supercoiling is the driving force behind these rearrangements.

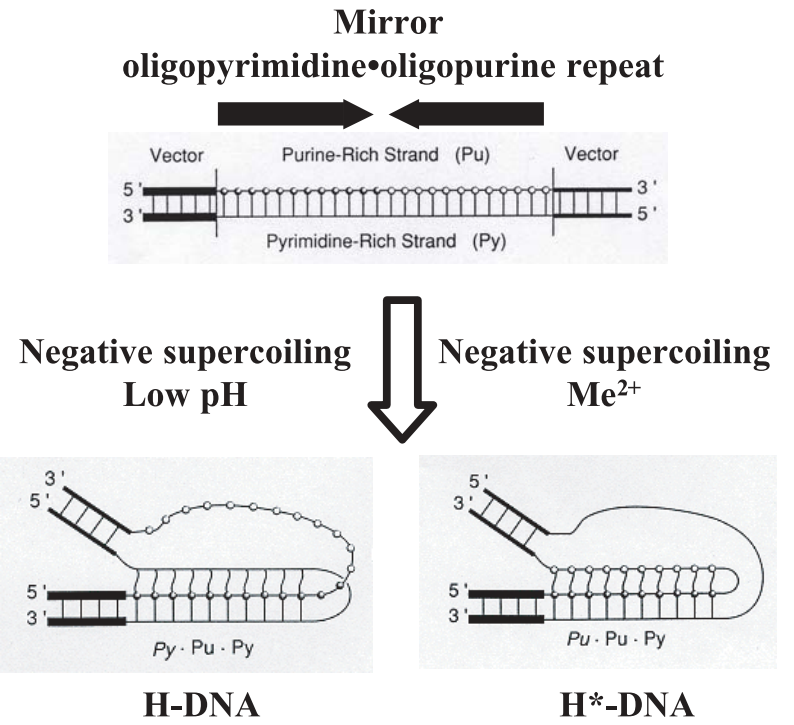


Fig. 7 Schematic description of the formation of an intramolecular triple helix. Under topological constraints, half of the duplex is disrupted into oligopyrimidine and oligopurine single strands. One of them subsequently folds back to form an intramolecular triplex with the other half of the duplex, while the other half is left single stranded. As the central part of the duplex acts as a linker, it can be any sequence of 3–10 bp. Different motifs, H- or H*-DNA, can be formed depending on the pH and concentration of divalent cations

3.1 H-DNA, H*-DNA and the Intrastrand Triplex

Hypersensitivity of supercoiled plasmids carrying oligopyrimidine•oligopurine sequences with mirror symmetry toward single strand-specific nucleases under acidic conditions, and the conformational transitions of DNA topoisomers unraveled by two-dimensional gel electrophoresis, have been interpreted in terms of a new DNA structure. This structure consists of folding back one half oligopyrimidine strand on a mirror symmetry-related oligopyrimidine•oligopurine sequence by forming T•AxT and C•GxC$^+$ base triplets in the Hoogsteen configuration, while the displaced half complementary oligopurine strand is left single-stranded (Fig. 7). It was called H-DNA due to the requirement for protonation of cytosine. Two isomers of this intramolecular structure can be identified where either the 5′- or the 3′-part of the oligopurine strand is exposed single-stranded.

A variant of H-DNA structure (called H*-DNA) has also been evidenced in which one half oligopurine strand folds back on the mirror symmetry-related oligopyrimidine•oligopurine sequence by forming T•AxA and C•GxG base triplets in the reverse Hoogsteen configuration, leaving the displaced half complementary oligopyrimidine strand single-stranded (Fig. 7). Again two isomers are possible as seen in H-DNA. The formation of H*-DNA depends on the sequence, the relative stability of each isomer, and the presence of divalent cations such as Mg^{2+} and Zn^{2+}.

It should be pointed out that the intramolecular triple helix could be formed in sequences that are not perfect oligopyrimidine•oligopurine mirror repeats due to the degeneracy of T•AxA and T•AxT base triplets in the reverse Hoogsteen configuration (i.e. the (G,T)- or mixed (G,T,A)-motifs). It can be anticipated that other imperfect oligopyrimidine•oligopurine mirror repeats could also form intramolecular triplexes with a mixed motif, such as the (T,C,G)-motif. Nevertheless, one half of the pseudo-mirror repeat must be a perfect oligopyrimidine•oligopurine sequence.

A different class of intramolecular triple-helical structures includes those arising from the folding of consecutive blocks of nucleotides along a single strand of DNA [101]. The structures formed have been referred to as intrastrand triplexes and the formation of the base triplets follows the same order as in the earlier mentioned structures, H- and H*-DNA. Unlike H- and H*-DNA, intrastrand triplex structures require three consecutive oligopyrimidine•oligopurine sequence domains with appropriate symmetry, and demand a higher level of negative supercoiling stress due to the requirement to disrupt two oligopyrimidine•oligopurine domains instead of only one in H- and H*-DNA.

3.2 Potential Occurrence

Two bioinformatic analyses regarding the natural occurrence of oligopurine•oligopyrimidine sequences have been reported on prokaryotic genomic DNA as well as a set of eukaryotic genes. In the first, an overabundance of long oligopurine tracts (15–30 purines) was found in all eukaryotic sequences included in the study, starting from yeast and worm through fruit fly, chicken, mouse and finally to human [59]. Whereas long oligopurine tracts were statistically over-represented in all of these eukaryotic genomes, very long oligopurine stretches (>30 contiguous purines) occurred much more frequently in mouse and human DNA than in the other eukaryotes. When prokaryotic genomic sequences of *E. coli* and *B. subtilis* were analysed, none of these trends were observed. In the second analysis, 157 human genes totalling 1,086,110 bp, the complete yeast (*Saccharomyces cerevisiae*) chromosome III which contains 182 ORFs (315,357 bp), and an *Escherichia coli* sequence containing 350 ORFs (324,146 bp) were analysed [102]. The occurrence of mirror repeats of DNA was compared in the three organisms and it was concluded that while both human and yeast genomes were highly enriched in the total number of mirror repeat

sequences, especially longer repeats (>12 bp), mirror repeats only occurred at random frequency in *E. coli*.

A recent and more detailed analysis of the complete yeast (*S. cerevisiae*) genome (12,294,083 bp, 6608 ORFs) has been carried out concerning the occurrence of H- or H*-DNA, including all possible triplex motifs, with the length of mirror repeat >10 bp (Polverari et al., manuscript in preparation). An empirical rule based on the predicted stability of intermolecular triple helix formation was applied to eliminate the intramolecular triplexes that are unlikely to be formed due to weak stability. In terms of frequency, the potential H- or H*-DNA sites in the promoter region defined as the proximal 500 bp upstream of the first exon (1/46,535) were estimated to be about 1.8-fold more common than the average occurrence of H-DNA in the yeast genome (1/83,068) and 3.3-fold more common than those in exons (1/152,903). This is in line with the early observation of hypersensitivity toward S1 nuclease at the promoter region of some eukaryotic genes that are associated with oligopyrimidine•oligopurine sequences.

Although the analyses presented above argue for the potential presence of H-like intramolecular triple helices in eukaryotic but not prokaryotic genomes, other types of intramolecular triplex structures might still be found in prokaryotes. One such example is the intrastrand triple-helical structure that was mentioned in the previous section. When the databases of several bacterial genomes were searched, potential intrastrand triplex elements were found in *E. coli*, *Synechocysis* and *H. influenzae* [103]. Remarkably, these elements were detected as multiple copies of a particular class of sequence, and in *E. coli* up to 25 copies of one particular purine-motif potential intramolecular triplex were identified.

3.3
In Vivo Evidence of Intramolecular Triple-Helices and Their Biological Relevance

A more direct in vivo probing of triplex formation has been achieved by immunodetection using triplex-specific antibodies. Agazie et al. prepared triplex-specific monoclonal antibodies, Jel 318 and Jel 466, by immunizing mice with a triplex-forming sequence of DNA [104, 105]. Binding of the two antibodies to chromosomes and cell nuclei was demonstrated by immunofluorescence and was inhibited by the addition of competing triplex DNA [106]. More recently, immunodetection by triplex-specific antibodies was combined with fluorescence in situ non-denaturating hybridization (N-FISH) in order to detect triplex-forming DNA in nuclei of fixed cells under mild conditions [107]. Employment of single-strand DNA fluorescent probes complementary to the single-stranded region associated with intramolecular triplexes led to the detection of sequence-dependent foci-type signals coincident with, or closely related to, triplexes that were immunolocalized by the triplex-specific antibodies. These results demonstrate that triple-helical structures could form in vivo, prompting the suggestion of several models of higher order structure

of chromatin in living cells wherein intramolecular as well as transmolecular triplex structures could participate.

Since DNA undergoes a characteristic conformational change upon H-DNA formation and forms a sharply bent double helix, which is often observed when regulatory proteins bind to DNA, it is tempting to postulate that H-DNA could act as a molecular switch to modulate gene expression in a structure-dependent manner in vivo, and that cellular proteins could specifically recognize triple-helical DNA and stabilize it. This speculation is supported by two facts: (1) mirror repeats of oligopyrimidine•oligopurine sequences have been found in several eukaryotic genomes, and were often located near regulatory regions and (2) DNA topological constraints, especially negative supercoiling, are an intrinsic and dynamic feature during the processing of DNA information such as replication, transcription, etc.

Notwithstanding a large number of biophysical and biochemical studies on the formation of H-DNA in vitro and in living cells during the last two decades, there is still a good deal of mystery surrounding the biological relevance played by these rather peculiar DNA structures in vivo. However, many independent investigations have supported the possible involvement of intramolecular triple-helical structures as regulatory elements, either as enhancers or repressors. A recent review has focused on the results that have been achieved in the search for intramolecular triple-helical structures in vivo and their possible biological relevance [108]. Nevertheless, the establishment of any detailed mechanism for intramolecular triplex-directed regulation of gene expression is still very challenging.

4
Triplex-Binding Proteins

Whereas sequence analyses of eukaryotic genomes lent early credence to the potential existence of intramolecular triplex structures in DNA of living cells, the chief difficulty in detecting H-DNA (and H*-DNA) in vivo lay in the lack of probes that could gain access to the nucleus, bind to DNA and act specifically on the triple-helical structures potentially formed there. More recently the identification of several proteins that can interact with triple-helical DNA supports the argument that triplexes can be formed and stabilized upon binding of triplex-specific and single strand-specific proteins in vivo, and therefore might ultimately have biological implications. Both inter- and intramolecular triplexes with different sequences have been used to select triplex-binding protein(s) from HeLa cell nuclear extracts [109, 110]. Using gel mobility shift assays, 2D gel electrophoresis and mass spectrometry, several proteins that showed affinity towards triplex structures have been identified. Nevertheless, the authors concluded that the identified triplex-binding proteins were not entirely specific to triplexes but showed some affinity for other types of nucleic acid structures [111]. Moreover, the biological relevance of these

proteins, besides their ability to bind triple-helical structures, has not yet been defined.

Another triplex-binding protein has been identified in *Saccharomyces cerevisiae* [112]. This protein was purified from whole-cell yeast extract using affinity chromatography and a psoralen-stabilized intermolecular DNA triple-helix. Electrophoretic mobility shift assay (EMSA) confirmed that the protein had a high preference for binding to purine-motif triplexes and sequence analyses revealed that it was the product of the STM1 gene. The STM1 gene has been identified as a multicopy suppressor of mutations in several genes involved in mitosis. In an effort to identify additional genes that encode triplex-binding proteins, Musso et al. screened an *S. cervisiae* genomic library using a psoralen cross-linked triplex probe and Southwestern blotting methods [113]. These experiments indicated that a second gene, CDP1, also encoded a triplex-binding protein. However, the screening identified only a small portion of this protein and no other yeast genes were found using the same method, including STM1.

It was independently reported that the GAGA factor which is a sequence-specific DNA-binding protein in *Drosophila*, could bind triple-helical structures of DNA in vitro using EMSA and footprinting by DNase I and DMS [114]. GAGA showed a similar affinity and specificity for intermolecular triplexes as for canonical double-stranded DNA. GAGA-binding sites are found in the promoter of several genes and they consist of repeated $d(GA/TC)_n$, which could potentially form triple-helical structures. Since the GAGA factor participates in the regulation of expression of several genes in *Drosophila*, it was suggested that its observed interaction with triplex DNA could play a role at that level. However, additional evidence demonstrating the presence of a GAGA-triplex complex in vivo is still necessary.

TnsC, a protein encoded by the bacterial transposon Tn7, has been reported to be able to recognize and bind inter- as well as intramolecular triple-helix structures formed in plasmid DNA in vitro [115, 116]. Interestingly, TnsC has been shown to control the target site of the transposon, leading it to insert preferentially adjacent to the formed inter- or intramolecular pyrimidine-motif triplex structure. Although these experiments outlined, in a remarkable way, the interaction of the protein TnsC with triple-helical DNA structures, as well as the consequence that it has on targeting the transposition event, it remains to be seen whether these findings provide hard evidence for triplex formation in vivo as well as the generality of the phenomenon in terms of potential involvement in different biological processes.

5
Triple Helix-Specific Ligands

5.1
General Considerations

The widespread interest in triple-helical nucleic acids has led the scientific community to look for drugs that could bind specifically to these structures.

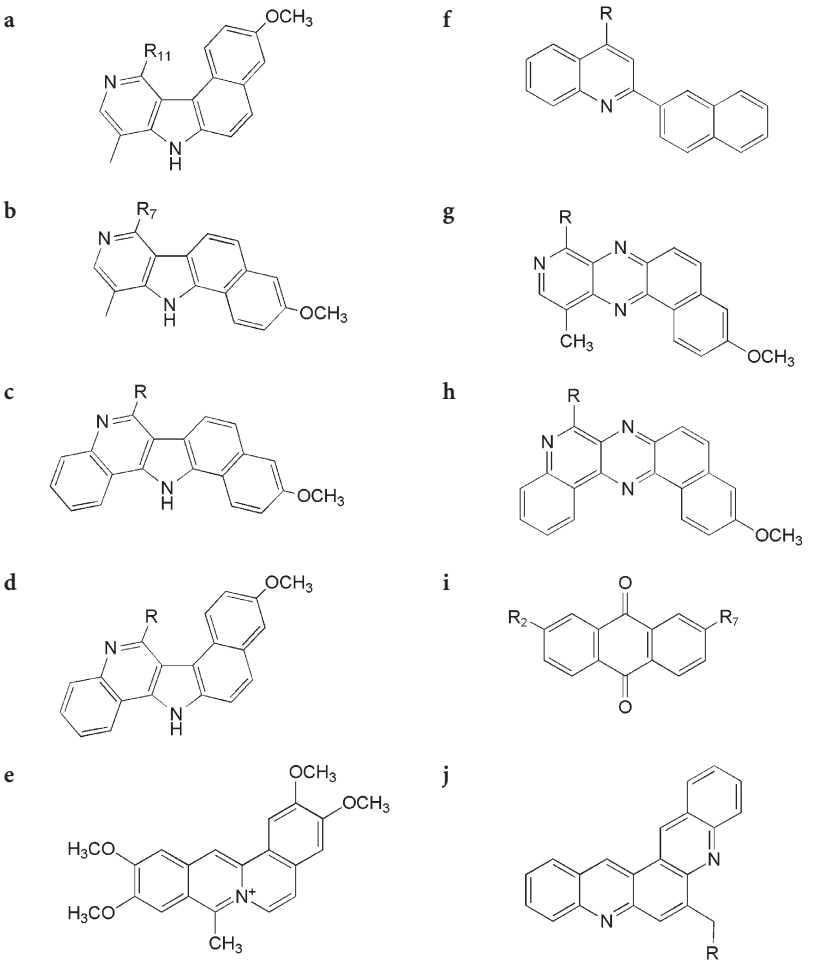


Fig. 8a–j Chemical structures of various triplex binding intercalators: **a** Benzo[e]pyridoindole, **b** benzo[g]pyridoindole, **c** benzo[f]indoloquinoline, **d** benzo[h]indoloquinoline, **e** coralyne, **f** naphthylquinoline, **g** benzo[f]pyridoquinoxaline, **h** benzo[f]quinoquinoxaline, **i** 2,7-disubstituted anthraquinone and **j** dibenzophenanthroline. For the sake of clarity, the structure of alkylamine side chains has been omitted. For **a**, **b**, **c**, **d**, **g** and **h** R=$NH(CH_2)_3NH_3^+$; for **f** R=$NH(CH_2)_2NH^+(CH_3)_2$; for **i** R=$NHCO(CH_2)_2NH^+(CH_3)_2$; for **j** R=$(CH_2)_3NH^+(CH_3)_2$

There are two main reasons for such an investigation. First, molecules that would bind preferentially to triple-helical DNA compared to double-stranded DNA should enhance the stability of both inter- and intramolecular triple-helical structures. Second, if the formation of H-DNA or any other triple-stranded nucleic acid structure has a biological role, compounds that bind to these structures would provide a new tool for studying their biological function and for the design of new probes for their detection.

The search for triplex binding drugs was initiated in the late 1980s. A first report showed that the classical DNA intercalator, ethidium bromide, could bind to DNA triple helices by intercalation [117]. This compound stabilized, although rather poorly, a polydA•2polydT triple helix, but destabilized triple helices containing C•GxC+ base triplets. This was attributed to electrostatic repulsion between ethidium and the protonated cytosines. Screening of a library of DNA intercalators led to the discovery of the first triple helix-specific DNA-binding agent, benzo[e]pyridoindole [118]. Since then, other polyaromatic compounds have been studied that bind and stabilize DNA triplexes by intercalation (reviewed in [119]) (Fig. 8). The most extensively studied are benzopyridoindole derivatives, coralyne, naphthylquinoline compounds and disubstituted anthraquinones.

The influence of minor groove binding agents on triple helices has also been investigated. The concomitant binding of both a TFO and the minor groove binding agent netropsin have been detected by CD spectroscopy and calorimetric methods. Other drugs that bind to DNA triple helices include Hoechst 33258, Berenil, DAPI, and Distamycin A (see [119] for a review). In general, binding of minor groove binders is associated with a decrease in the thermal stability of triple helices. But there have been several exceptions (see below).

5.2 Methodology

A review on the methods used to study interactions of drugs with triple-helical nucleic acids has already been published [120]. Triple helix formation can be followed by thermal denaturation experiments, electrophoretic mobility shift assays, DNase I or chemical footprinting, or inhibition of restriction enzyme cleavage. These methods therefore afford suitable means of assessing the effects of various ligands on triplex formation and stability. Melting temperature experiments represent the most convenient tool for investigating whether drugs can stabilize triple helices and have therefore been used in a large number of studies. This approach has also been processed in a high throughput way using a DNA cycler and molecular beacon for measuring triplex stability [121]; see Chap. 6 of this volume.

A very interesting competition dialysis method has also been developed which allows the experimenter to compare in a fast and easy assay the relative affinity of DNA-binding agents for different structural motifs [122]; see Chap. 3 of this volume. In this method, various nucleic acids structures are dialysed

against a ligand solution. After an incubation period, the amount of dye that has accumulated in each chamber is quantified. It is proportional to the affinity of the dye for the DNA structure in the chamber.

5.3 Structural Features

Benzo[e]pyridoindole is probably the ligand for which the mode of binding has been studied using the widest variety of techniques, including fluorescence quenching, fluorescence energy transfer, viscometry and linear dichroism experiments. All the results are consistent with an intercalative mode of binding. Intercalation has also been established for amidoanthraquinones, naphthylquinoline and a ruthenium derivative [123]. The preferential binding of these compounds to triple helices most likely results from strong stacking interactions between their large polyaromatic ring systems and base triplets, but hydrophobic and electrostatic interactions certainly also play an important role.

Two new grooves are present in triple helices compared to double-stranded DNA. In order to name them, we choose to call the strand containing the oligopyrimidine sequence the Watson strand, the one containing the oligopurine sequence the Crick strand, and the third strand the Hoogsteen strand. The grooves can then be named after the strands that create them. There are three: the Watson–Crick groove (formally equivalent to the minor groove of the double helix), the Watson–Hoogsteen and the Crick–Hoogsteen groove (see [119]). The small negative effect of many known minor groove-binding agents is probably due the fact that TFOs and minor groove-binders induce different distortions of the DNA double helix, which are not fully compensated by van der Waals interactions. Classical minor groove binding agents usually preserve their binding mode in triple helices, i.e. they bind in the Watson–Crick groove. Some exceptions have been reported. The Crick–Hoogsteen groove is very narrow, but the Watson–Hoogsteen groove may accommodate the binding of some ligands, like neomycin [124]. It has also been reported that DAPI, which is usually a groove binder, may intercalate into some triple helices [125].

No high-resolution structure for complexes of triple-helical DNA with binding agents is available so far. Therefore, molecular modelling has proved to be very helpful in trying to understand the interactions between polyaromatic molecules and triplex DNA, especially in the case of benzopyridoindole derivatives, naphthylquinolines and amidoanthraquinone. It has helped in the design of one of the most efficient triplex stabilizers to date, BQQ, by suggesting that addition of a new ring to the BfPQ molecule would enhance stacking interactions with the Watson strand of the triple helix [126]. A rationale has been provided for the differential effects of 2,6-, 2,7- and 1,4-disubstituted anthraquinone derivatives on triplex stabilization [127, 128]. Molecular modelling has also suggested that neomycin displays a good charge and shape complementarity with the Watson–Hoogsteen groove of triplexes [124, 129].

5.4 Thermodynamics

The apparent Kd for triple helix formation from double- and single-stranded DNA is usually much lower in the presence of triplex-stabilizing agents. This difference is the result of a higher affinity of the ligand for the triple-helical structure compared to the double- and single-strands. Accordingly, it is the binding affinity of each compound for various triple-helical structures compared to all other possible nucleic acid structures that is a key parameter.

Thermodynamic data relating to the binding of these compounds to triplex DNA may be obtained with the help of spectroscopic methods (absorbance or fluorescence), provided that the binding of the chromophore is associated with changes in the absorption or emission spectrum. However, this task is complicated by the fact that large hydrophobic molecules often self-associate and by the presence of different binding sites within the target triple-stranded nucleic acid. The dialysis method is the only one which allows direct analysis of the structural selectivity of a ligand in a fast and convenient way (see Chap. 6). However, one of its drawbacks is that the variety of structures for which the affinity is compared is chosen by the experimenter. Various polynucleotide sequences are usually selected representing a panel of known structures. Efforts to expand the repertoire of possible targets, including tetraplex and unnatural double-stranded DNA, are of obvious interest.

Stabilization of triplex DNA may result in the disproportionation of a DNA duplex into a triplex and a single-strand. This has been demonstrated with coralyne for example. In the presence of this compound, a polydA•polydT duplex can disproportionate into a polydA•2polydT triplex plus a free poly(dA) strand. This disproportionation has been studied using shorter oligonucleotides, and it was shown that the process depends greatly upon the length and the temperature [130]. The structure adopted by the oligo(dA) strand probably forms only in the presence of the ligand and has not yet been clearly characterized. Therefore, these processes are complicated. The relevance of these disproportionation events for natural sequences or in the formation of intramolecular triple helices has not yet been established.

5.5 Sequence Selectivity

Most of the triplex-stabilizing agents have been reported to stabilize triple helices containing consecutive T•AxT triplets. Some of them have been shown to stabilize triple helices that also contain C•GxC+ base triplets, but with a lower efficiency. A very recent study has investigated the ability of BePI, naphthylquinoline and anthraquinone to stabilize triplexes made of $(TC)_n$, $(CCT)_n$ and $(TTC)_n$. The only ligand that can stabilize triple helices in the absence of blocks of T•AxT triplets is amidoanthraquinone. This is most likely due to the absence of a positive charge on the aromatic portion of the ligand [131]. It would be

interesting to use competition dialysis to compare the binding of particular ligands to various triple-stranded structures containing a variety of binding sites.

Binding of potential ligands has been much more studied for DNA triple helices having a pyrimidine-rich third strand. There are only a few reports regarding the interaction of drugs with triple helices having a third strand containing purines. BePI has been shown to induce the formation of a triple helix formed with a (G,T)-containing antiparallel third strand. The triple helix was very stable and could be detected by thermal denaturation experiments. The absorbance of the ligand is probably responsible for part of the hypochromicity associated with triple helix formation. Some compounds are also able to stabilize triple helices with (G,A)-containing third strands, but the stabilization efficiency and the stability of the triple helices were generally lower than for (G,T)-motif triple helices. For this motif, there is also no detailed structural information about the drug–DNA complex. Intercalation may relieve part of the constraints on the DNA backbone, thus enhancing triplex stability. Other experiments carried out in our laboratory have shown that this compound can stabilize triple helices with a (G,T)-parallel third strand [132]. However, it is difficult to resolve the relative contribution of binding to the classical (T,C)-motif and to the stretch of C•GxG triplets.

It should be mentioned here that triple helices containing RNA strands can also form and be stabilized by triplex binding agents. DAPI has been shown to induce the formation of RNA triple helices which will not form in the absence of drug [125, 133]. Neomycin was also found to induce the formation of DNA•RNAxDNA and DNA•RNAxRNA triplexes [134]. However, these reports which involve polynucleotide structures have never been followed up by studies with short oligomers.

5.6
Dimers of Triplex Stabilizing Agents

The only triplex-stabilizing agents for which synthesis of dimers has been attempted are ligands belonging to the naphthylquinoline family [135, 136]. Very strong intercalators were obtained. Interestingly, the dimers were built using side chains linked at two different positions to the naphthylquinoline ring. The most active compounds were those with the linker located in the Watson–Crick groove and sufficiently long to be compatible with bisintercalation. Linkers in the Watson–Hoogsteen groove resulted in unfavourable steric interactions with the thymines.

Another interesting development has been described which consists of linking a triplex-specific intercalator to a minor groove-binding agent. The very efficient triplex intercalator BQQ was conjugated to neomycin, which had been shown to bind in the Watson–Hoogsteen groove of triple-helical DNA [137]. At low concentration, this chimera was more efficient than its individual components separately. This may be explained by an increased affinity for nucleic

acids. At higher concentration, the stabilization is so important that it becomes tricky to compare the compounds.

5.7 Conjugates of Triplex-Stabilizing Agents with TFOs

It has been known for a long time that covalent attachment of a molecule known to intercalate into double-stranded DNA to a TFO improves the binding affinity of this oligonucleotide for its target [25]. Based on the same idea, it has been suggested that covalent attachment of a triplex-specific intercalator to a TFO could engender stronger binding, provided that the intercalator can interact with base triplets. This could occur if the intercalator binds at the triplex-duplex junction and engages in favourable stacking interactions, but intercalation between base triplets may be possible. Some such conjugates have been synthesized, but up to now they have proved not much better than conjugates of oligonucleotides with double-stranded DNA intercalators [138–141]. A better knowledge of the geometry of the intercalated complex would probably help to optimize the nature and the position of the linker between the intercalator and the TFO. In the absence of structural data, new derivatives are being synthesized to explore new possibilities [142].

6 Applications

6.1 Stabilization of Intermolecular Triple Helices

The use of triplex binding agents to promote the formation of intermolecular triple helices, as has been done in several in vitro studies, could have numerous useful applications. An increase in the efficiency with which a TFO may block the progression of RNA polymerase has been reported [132]. The very efficient triplex stabilizer BQQ has found employ in order to form very stable triple helices for purposes such as DNA labelling prior to observation by electron or optical microscopy [93, 94] or plasmid modification for enhanced gene delivery [95]. In these applications, the triple helix is used to assemble circular oligonucleotides around double-stranded DNA. These studies were conducted with a (G,T)-triple helix whose formation is induced by the presence of the ligand. Interestingly, the BQQ ligand was shown to modulate the interactions between a circular oligonucleotide containing a sequence that can form a triple helix and its target site for triple helix formation (Fig. 9). In the absence of BQQ, the circular oligonucleotide was free to move from its binding site, whereas addition of BQQ locked the oligonucleotide around its target, resulting in inhibition of cleavage by a restriction enzyme [143] and blockage of transcription elongation by an RNA polymerase [144]. This

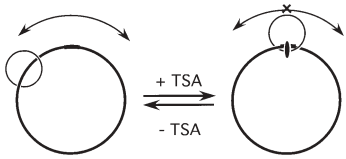


Fig. 9 Ligand-mediated control of the positioning of a padlock oligonucleotide on its target DNA. In the absence of ligand, the padlock oligonucleotide can move freely away from its binding site. In the presence of ligand (TSA), the circular oligonucleotide is locked on its target double-stranded sequence

approach could lead to new strategies aimed at regulating gene expression by small molecules.

Triple helix-specific ligands may also prove useful to expand the repertoire of possible target sequences. For example, it has been shown that the concomitant use of benzo[e]pyridoindole and some modified nucleobases enabled the targeting of a sequence containing a central mismatch with some sequence discrimination [145].

6.2 Binding to Intramolecular Triple Helices

Triplex binding agents display interesting properties per se as structure-specific DNA-binding agents but they open possibilities to study the biological effects of triple-helical DNA and to probe this structure. To date, only one study has addressed the question of the effect of triplex-binding ligands on intramolecular triple helices [146]. A mirror repeat sequence was inserted between the *E. coli* β-lactamase gene, which confers resistance to the antibiotic ampicillin, and its promoter in a plasmid, which also expresses the resistance gene for the antibiotic tetracycline. Growth of the cells containing the mirror repeat sequence was slower in the presence of ampicillin as compared with plasmids not containing this sequence, whereas both types of cells could grow at the same rate in the presence of tetracycline. Thus, transcription of the β-lactamase gene was inhibited by the mirror repeat sequence, whereas the plasmid could replicate normally. Adding the triplex-stabilizing agent BePI was shown to increase this inhibition specifically. The use of chemical probes such as chloroacetaldehyde also showed that BePI stabilized the H-DNA structure induced by supercoiling in the plasmid. This experiment suggests that mirror sequences that are able to form H-DNA structures could act as gene regulators, and that regulation of these genes could well be influenced by triplex-stabilizing agents. Further experiments are needed to confirm these results and to understand their mechanism at the molecular level. Other triple-helical structures, such as that formed by binding of a nascent transcribed RNA to double-stranded DNA, may also form and/or be enhanced by the binding of triplex-stabilizing agents. It remains to be investigated whether a triplex-stabilizing agent that can induce the formation of an

H-DNA structure within the promoter region of a specific cellular gene would down-regulate expression of that gene. If this is possible, it would raise the interesting prospect of a new mechanism of drug action which might even underlie the cytotoxic properties reported for some triplex-stabilizing agents.

6.3 Triple Helix-Specific DNA Cleaving Agents

TFOs have been used as tools for the sequence-specific targeting of irreversible modifications to double-stranded DNA, such as cleavage for example [147] or adduct formation [74]. These goals were achieved by linking a chemically reactive moiety to the TFO. Linking similar moieties to triplex-specific binding agents is another interesting approach to target irreversible modifications to inter- or intramolecular triplexes.

6.3.1 First Example of a Triplex-Specific Cleaving Agent

This concept has been tested by attaching EDTA to the triplex-stabilizing agent BQQ [148]. EDTA-Fe can generate hydroxyl radicals through the Fenton reaction and cleave nucleic acids. The selectivity of the BQQ–EDTA conjugate for a triplex structure was sufficiently high to induce oligonucleotide-directed DNA cleavage at a single site on a 2718 bp plasmid DNA upon addition of the appropriate TFO. When cleavage was studied at higher resolution within an 80 bp DNA fragment it was found to occur all along the triplex sequence, with peaks in the A.T-rich regions. This suggests that various sites exist within the triplex. However, as the hydroxyl radicals diffuse readily the method does not allow localization of the sites with sufficient accuracy. It remains to be determined whether such conjugates can be used as probes for the existence of triple-helical DNA structures in vivo.

6.3.2 Other Triplex-Specific Cleaving Agents

EDTA has been covalently attached to various triplex intercalators in our laboratory, including BePI and BgPI [149]. The BePI-EDTA conjugate was found to cleave duplex DNA, but not triplex DNA. BgPI-EDTA was able to cleave triplex DNA more efficiently than duplex DNA, but proved to be much less specific than BQQ-EDTA. Interestingly, these findings corroborate the results of previous melting temperature experiments showing that BQQ was considerably more specific than BgPI. They also provide experimental verification of a model that had been proposed on the basis of molecular modelling. According to this model, the side chain of BePI is located in the Watson–Hoogsteen groove, whereas that of BgPI and BQQ lies in the Watson–Crick (minor) groove. This model largely explains the different reactivity of BePI- and BgPI-EDTA

since it is known that cleavage by EDTA-Fe occurs in the minor groove [9]. The synthesis of conjugates between a triplex-stabilizing agent and a cleaving agent may therefore provide new insight into the structure and specificity of drug-triplex DNA interactions. Other conjugates have been synthesized, for example orthophenanthroline-BQQ (Zain et al., personal communication).

6.3.3 Photoactivable Compounds

Quinacridine derivatives are photosensitive compounds that have been shown to stabilize triple-helical DNA. They were shown to promote the photocleavage of a long double-stranded DNA in the vicinity of a TFO binding site only in the presence of this TFO [150]. Interestingly, the damage was located 6–9 bp away from the end of the triple helix site. There was experimental evidence for electron transfer between the quinacridine intercalated at the end of the triple helix and remote guanines. It should also be mentioned that, besides cleavage, triplex-specific intercalators may be able to induce other kinds of chemical damage such as photoaddition. The irradiation of oligonucleotides in the presence of a BePI derivative has been reported to result in the appearance of a retarded band on a denaturing polyacrylamide gel [118]. Piperidine treatment of these photoadducts may result in DNA cleavage, but this process was found to be rather inefficient with the compounds investigated in our laboratory (C. Escudé et al., unpublished results).

7 Perspectives

Chemical approaches to the design of DNA sequence- and structure-specific ligands, in particular major groove-binders such as triple helix-forming molecules (TFMs), have opened new avenues to exploit genomic information and to provide new tools for functional genomics as well as novel candidates for therapeutic development. Most of the work carried out so far has dealt with in vitro and cellular studies. In vivo applications of TFMs are still in their infancy and are lagging behind the development of antisense oligonucleotides, and now of siRNA. The recent results in mice [81] have provided a first hint at achieving the goal of activity in vivo.

Although the relatively simple use and high efficiency of siRNA seems to be the method of choice to knock down the expression of a targeted gene, TFMs that recognize and bind to specific genomic sequences in living mammalian cells would have their own potential for genetic manipulation, including gene knockout, strain construction and gene therapy – all goals that amply justify further in vivo investigations. The explosion of genomic information and genome analysis using the tools of bioinformatics provides a wide basis for developing such tools to hasten the day of chemical genomics, when science

may enable the ultimate goal of designing molecules that could act selectively on every single gene or gene product in cells and in vivo.

A better understanding of cellular traffic of oligonucleotides is necessary to design appropriate carriers that could increase oligonucleotide-specific activity in vivo. It can be anticipated that a substantial increase in intracellular efficacy of triplex-induced effects would be achieved by modifications that improve not only triplex stability but also nuclease resistance, cellular uptake and absence of notable toxicity of TFMs in the same way as has been described for antisense oligonucleotides. These optimizations should enhance the efficiency of TFMs so that they become powerful tools in the post-genome era, when new components of DNA-dependent pathways have to be characterized in molecular terms, or to probe pathways of cellular DNA metabolism such as chromatin remodelling, transcription or DNA repair.

Intercalating agents, which bind specifically to triplex DNA, can prove useful for promoting triple helix formation, but they merit interest also as structure-specific ligands. A better structural characterization of the complexes of various ligands with triplex DNA would be helpful for the design of more effective agents as well as new TFO conjugates and new cleaving agents. More work should also be devoted to the exploration of the biological effects of drug binding to triplex DNA, using new model cellular systems, as there is still a huge gap between in vitro data and biologically relevant phenomena.

After more than a decade of investigation, the accumulated data strongly suggest that intramolecular DNA triplexes may exist in vivo and be concerned in a number of DNA information flow processes involving a family of genes that accommodate the appropriate oligopyrimidine•oligopurine mirror repeat sequences in their regulatory region. However, it remains to be unambiguously established that these intramolecular DNA triple-helical structures can act as molecular switches to modulate gene expression or other DNA metabolism events in a structure-dependent manner, in addition to the well-known processes of sequence-specific regulation. Further identification and characterization of proteins that bind specifically to peculiar nucleic acid structures in conjunction with the development of efficient new structure-specific ligands could lead to enhanced understanding of the roles of DNA structure in vivo and contribute to the discovery of structure-related gene regulation mechanisms. The fast and still accelerating pace of deciphering the genomic information of a large number of organisms, together with the development of functional genomics, triplex-specific ligands and probes, will certainly contribute to further elucidating the formation of DNA triple-helical structures in vivo and their biological relevance.

AcknowledgmentsThis chapter is dedicated to the memory of Professor Claude Hélène who inspired and contributed so much to the field of nucleic acids in the last two decades. The authors thank all the members of the Laboratoire de Biophysique of the Muséum National d'Histoire Naturelle and our collaborators in other laboratories over the past two decades. Each of them has contributed to the development of oligonucleotide-based strategies for gene-directed modulation of cell function and genomic modification.

References

1. Watson JD, Crick FH (1953) Nature 171:737
2. Levine M, Tjian R (2003) Nature 424:147
3. Garvie CW, Wolberger C (2001) Mol Cell 8:937
4. Marmorstein R, Fitzgerald MX (2003) Gene 304:1
5. Johnson DS, Boger DL (1996) In: Murakami Y (ed) Comprehensive supramolecular chemistry, vol 4. Elsevier, p 73
6. Dervan PB (2001) Bioorg Med Chem 9:2215
7. Klug A (1999) J Mol Biol 293:215
8. Le Doan T, Perrouault L, Praseuth D, Habhoub N, Decout J-L, Thuong NT, Lhomme J, Hélène C (1987) Nucleic Acids Res 15:7749
9. Moser HE, Dervan PB (1987) Science 238:645
10. Giovannangéli C, Hélène C (2000) Curr Opin Mol Ther 2:288
11. Felsenfeld G, Davies DR, Rich A (1957) J Am Chem Soc 79:2023
12. Sun JS, Hélène C (1993) Curr Opin in Struct Biol 3:345
13. Lee JS, Johnson DA, Morgan AR (1979) Nucleic Acids Res 6:3073
14. Sun JS, de Bizemont T, Duval-Valentin G, Montenay-Garestier T, Hélène C (1991) C R Acad Sci III 313:585
15. Giovannangéli C, Montenay-Garestier T, Thuong NT, Hélène C (1992) Proc Natl Acad Sci USA 89:8631
16. Mills M, Arimondo PB, Lacroix L, Garestier T, Klump H, Mergny JL (2002) Biochemistry 41:357
17. Kool ET (1998) Acc Chem Res 31:502
18. Plum GE, Park YW, Singleton SF, Dervan PB, Breslauer KJ (1990) Proc Natl Acad Sci USA 87:9436
19. Mills M, Arimondo PB, Lacroix L, Garestier T, Hélène C, Klump H, Mergny JL (1999) J Mol Biol 291:1035
20. Rougée M, Faucon B, Mergny JL, Barcelo F, Giovannangeli C, Garestier T, Hélène C (1992) Biochemistry 31:9269
21. James PL, Brown T, Fox KR (2003) Nucleic Acids Res 31:5598
22. Alberti P, Arimondo PB, Mergny JL, Garestier T, Hélène C, Sun JS (2002) Nucleic Acids Res 30:5407
23. Arimondo PB, Barcelo F, Sun JS, Maurizot JC, Garestier T, Hélène C (1998) Biochemistry 37:16627
24. Cheng AJ, Vandyke MW (1994) Nucleic Acids Res 22:4742
25. Sun JS, François JC, Montenay-Garestier T, Saison-Behmoaras T, Roig V, Chassignol M, Thuong NT, Hélène C (1989) Proc Natl Acad Sci USA 86:9198
26. Singleton SF, Dervan PB (1992) Biochemistry 32:13171
27. Xodo LE, Manzini G, Quadrifoglio F, van der Marel GA, van Boom JH (1991) Nucleic Acids Res 19:5625
28. Sun JS, Garestier T, Hélène C (1996) Curr Opin Struct Biol 6:327
29. Hildebrandt S, Blaser A, Parel SP, Leumann CJ (1997) J Am Chem Soc 119:5499
30. Cassidy SA, Slickers P, Trent JO, Capaldi DC, Roselt PD, Reese CB, Neidle S, Fox KR (1997) Nucleic Acids Res 25:4891
31. Lacroix L, Lacoste J, Reddoch JF, Mergny JL, Levy DD, Seidman MM, Matteucci MD, Glazer PM (1999) Biochemistry 38:1893
32. Roig V, Asseline U (2003) J Am Chem Soc 125:4416
33. Debin A, Laboulais C, Ouali M, Malvy C, Le Bret M, Svinarchuk F (1999) Nucleic Acids Res 27:2699

34. Arimondo PB, Garestier T, Hélène C, Sun JS (2001) Nucleic Acids Res 29:E15
35. Olivas WM, Maher LJ, 3rd (1995) Nucleic Acids Res 23:1936
36. Aubert Y, Perrouault L, Helene C, Giovannangeli C, Asseline U (2001) Bioorg Med Chem 9:1617
37. Asensio JL, Carr R, Brown T, Lane AN (1999) J Am Chem Soc 121:11063
38. Roberts RW, Crothers DM (1992) Science 258:1463
39. Escudé C, Sun JS, Rougée M, Garestier T, Hélène C (1992) C R Acad Sci III 315:521
40. Shimizu M, Konishi A, Shimada Y, Inoue H, Ohtsuka E (1992) FEBS Lett 302:155
41. Cuenoud B, Casset F, Husken D, Natt F, Wolf RM, Altmann KH, Martin P, Moser HE (1998) Angew Chem Int Ed Engl 37:1288
42. Carlomagno T, Blommers MJ, Meiler J, Cuenoud B, Griesinger C (2001) J Am Chem Soc 123:7364
43. Puri N, Majumdar A, Cuenoud B, Natt F, Martin P, Boyd A, Miller PS, Seidman MM (2002) Biochemistry 41:7716
44. Puri N, Majumdar A, Cuenoud B, Natt F, Martin P, Boyd A, Miller PS, Seidman MM (2001) J Biol Chem 276:28991
45. Braasch DA, Corey DR (2001) Chem Biol 8:1
46. Obika S, Uneda T, Sugimoto T, Nanbu D, Minami T, Doi T, Imanishi T (2001) Bioorg Med Chem 9:1001
47. Koizumi M, Morita K, Daigo M, Tsutsumi S, Abe K, Obika S, Imanishi T (2003) Nucleic Acids Res 31:3267
48. Escudé C, Giovannangeli C, Sun JS, Lloyd DH, Chen JK, Gryaznov SM, Garestier T, Hélène C (1996) Proc Natl Acad Sci USA 93:4365
49. Sun BW, Geinguenaud F, Taillandier E, Laurent M, Debart F, Vasseur JJ (2002) J Biomol Struct Dyn 19:1073
50. Dagle JM, Weeks DL (1996) Nucleic Acids Res 24:2143
51. Chaturvedi S, Horn T, Letsinger RL (1996) Nucleic Acids Res 24:2318
52. Vasquez KM, Dagle JM, Weeks DL, Glazer PM (2001) J Biol Chem 276:38536
53. Michel T, Martinand-Mari C, Debart F, Lebleu B, Robbins I, Vasseur JJ (2003) Nucleic Acids Res 31:5282
54. Nielsen PE (2001) Curr Med Chem 8:545
55. Faruqi AF, Egholm M, Glazer PM (1998) Proc Natl Acad Sci USA 95:1398
56. Wang G, Xu X, Pace B, Dean DA, Glazer PM, Chan P, Goodman SR, Shokolenko I (1999) Nucleic Acids Res 27:2806
57. Cutrona G, Carpaneto EM, Ulivi M, Roncella S, Landt O, Ferrarini M, Boffa LC (2000) Nat Biotechnol 18:300
58. Zhou P, Wang M, Du L, Fisher GW, Waggoner A, Ly DH (2003) J Am Chem Soc 125:6878
59. Behe MJ (1995) Nucleic Acids Res 23:689
60. Griffin LC, Dervan PB (1989) Science 245:967
61. Mergny JL, Sun JS, Rougée M, Montenay-Garestier T, Barcelo F, Chomilier J, Hélène C (1991) Biochemistry 30:9791
62. Yoon K, Hobbs CA, Koch J, Sardaro M, Kutny R, Weis AL (1992) Proc Natl Acad Sci USA 89:3840
63. Kiessling LL, Dervan PB (1992) Biochemistry 31:2829
64. Greenberg WA, Dervan PB (1995) J Am Chem Soc 117:5016
65. Zhou BW, Puga E, Sun JS, Garestier T, Helene C (1995) J Am Chem Soc 117:10425
66. Kukreti S, Sun J, Garestier T, Hélène C (1997) Nucleic Acids Res 25:4264
67. Sun JS (1999) In: Pritchard LL (eds) Triple helix forming oligonucleotides. Kluwer, p 273
68. Lehmann TE, Greenberg WA, Liberles DA, Wada CK, Dervan PB (1997) Helv Chim Acta 80:2002

69. Prevot-Halter I, Leumann CJ (1999) Bioorg Med Chem Lett 9:2657
70. Li JS, Fan YH, Zhang Y, Marky LA, Gold B (2003) J Am Chem Soc 125:2084
71. Seidman MM, Glazer PM (2003) J Clin Invest 112:487
72. Grigoriev M, Praseuth D, Guyesse AL, Robin P, Thuong NT, Hélène C, Harel-Bellan A (1993) CR Acad Sci III 316:492
73. Ritchie S, Boyd FM, Wong J, Bonham K (2000) J Biol Chem 275:847
74. Giovannangéli C, Diviacco S, Labrousse V, Gryaznov S, Charneau P, Hélène C (1997) Proc Natl Acad Sci USA 94:79
75. Belousov ES, Afonina IA, Kutyavin IV, Gall AA, Reed MW, Gamper HB, Wydro RM, Meyer RB (1998) Nucleic Acids Res 26:1324
76. Faria M, Wood CD, Perrouault L, Nelson JS, Winter A, White MR, Hélène C, Giovannangeli C (2000) Proc Natl Acad Sci USA 97:3862
77. Bailey C, Weeks DL (2000) Nucleic Acids Res 28:1154
78. Kuznetsova S, Ait-Si-Ali S, Nagibneva I, Troalen F, Le Villain JP, Harel-Bellan A, Svinarchuk F (1999) Nucleic Acids Res 27:3995
79. Gorman L, Glazer PM (2001) Curr Mol Med 1:391
80. Knauert MP, Glazer PM (2001) Hum Mol Genet 10:2243
81. Vasquez KM, Narayanan L, Glazer PM (2000) Science 290:530
82. Majumdar A, Puri N, Cuenoud B, Natt F, Martin P, Khorlin A, Dyatkina N, George AJ, Miller PS, Seidman MM (2003) J Biol Chem 278:11072
83. Chan PP, Lin M, Faruqi AF, Powell J, Seidman MM, Glazer PM (1999) J Biol Chem 274:11541
84. Culver KW, Hsieh WT, Huyen Y, Chen V, Liu J, Khripine Y, Khorlin A (1999) Nat Biotechnol 17:989
85. Datta HJ, Chan PP, Vasquez KM, Gupta RC, Glazer PM (2001) J Biol Chem 276:18018
86. Biet E, Maurisse R, Dutreix M, Sun J (2001) Biochemistry 40:1779
87. Maurisse R, Feugeas JP, Biet E, Kuzniak I, Leboulch P, Dutreix M, Sun JS (2002) Gene Ther 9:703
88. Potaman VN (2003) Expert Rev Mol Diagn 3:481
89. Wils P, Escriou V, Warnery A, Lacroix F, Lagneaux D, Ollivier M, Crouzet J, Mayaux JF, Scherman D (1997) Gene Ther 4:323
90. Zelphati O, Liang X, Hobart P, Felgner PL (1999) Hum Gene Ther 10:15
91. Branden LJ, Mohamed AJ, Smith CI (1999) Nat Biotechnol 17:784
92. Escudé C, Garestier T, Hélène C (1999) Proc Natl Acad Sci USA 96:10603
93. Roulon T, Coulaud D, Delain E, Lecam E, Hélène C, Escudé C (2002) Nucleic Acids Res:e12
94. Géron-Landre B, Roulon T, Desbiolles P, Escudé C (2003) Nucleic Acids Res 31:e125
95. Roulon T, Hélène C, Escudé C (2002) Bioconjug Chem 13:1134
96. Zarytova V, Levina A (1999) In: Pritchard LL (eds) Triple helix forming oligonucleotides. Kluwer, p 87
97. Matteucci MD, Lin HY, Huang T, Wagner R, Sternbach DD, Mehrotra M, Besterman JM (1997) J Am Chem Soc 119:6939
98. Arimondo PB, Bailly C, Boutorine A, Sun JS, Garestier T, Hélène C (1999) C R Acad Sci III 322:785
99. Arimondo PB, Boutorine A, Baldeyrou B, Bailly C, Kuwahara M, Hecht SM, Sun JS, Garestier T, Hélène C (2002) J Biol Chem 277:3132
100. Mirkin SM, Frank-Kamenetskii MD (1994) Ann Rev Biophy Biomol Struct 23:541
101. Hoyne PR, Gacy AM, McMurray CT, Maher LJ, 3rd (2000) Nucleic Acids Res 28:770
102. Schroth GP, Ho PS (1995) Nucleic Acids Res 23:1977
103. Hoyne PR, Edwards LM, Viari A, Maher LJ, 3rd (2000) J Mol Biol 302:797

104. Lee JS, Burkholder GD, Latimer LJP, Haug BL, Braun RP (1987) Nucleic Acids Res 15:1047
105. Agazie YM, Lee JS, Burkholder GD (1994) J Biol Chem 269:7019
106. Agazie YM, Burkholder GD, Lee JS (1996) Biochem J 316(2):461
107. Ohno M, Fukagawa T, Lee JS, Ikemura T (2002) Chromosoma 111:201
108. Zain R, Sun JS (2003) Cell Mol Life Sci 60:862
109. Guieysse AL, Praseuth D, Hélène C (1997) J Mol Biol 267:289
110. Musso M, Nelson LD, Van Dyke MW (1998) Biochemistry 37:3086
111. Guillonneau F, Guieysse AL, Le Caer JP, Rossier J, Praseuth D (2001) Nucleic Acids Res 29:2427
112. Nelson LD, Musso M, Van Dyke MW (2000) J Biol Chem 275:5573
113. Musso M, Bianchi-Scarra G, Van Dyke MW (2000) Nucleic Acids Res 28:4090
114. Jimenez-Garcia E, Vaquero A, Espinas ML, Soliva R, Orozco M, Bernues J, Azorin F (1998) J Biol Chem 273:24640
115. Rao JE, Miller PS, Craig NL (2000) Proc Natl Acad Sci USA 97:3936
116. Rao JE, Craig NL (2001) J Mol Biol 307:1161
117. Scaria PV, Shafer RH (1991) J Biol Chem 266:5417
118. Mergny JL, Duval-Valentin G, Nguyen CH, Perrouault L, Faucon B, Rougée M, Montenay-Garestier T, Bisagni E, Hélène C (1992) Science 256:1681
119. Escudé C, Garestier T (1999) In: Pritchard LL (eds) Triple helix forming oligonucleotides. Kluwer, p 257
120. Escudé C, Garestier T, Sun J-S (2001) In: Chaires JB, Waring MJ (eds) Drug-nucleic acid interactions (Methods in Enzymology), vol 340. Academic, p 340
121. Darby RA, Sollogoub M, McKeen C, Brown L, Risitano A, Brown N, Barton C, Brown T, Fox KR (2002) Nucleic Acids Res 30:e39
122. Ren J, Chaires JB (1999) Biochemistry 38:16067
123. Choi S-D, Kim M-S, Kim SK, Lincoln P, Tuite E, Norden B (1997) Biochemistry 36:214
124. Arya DP, Micovic L, Charles I, Coffee RL, Jr., Willis B, Xue L (2003) J Am Chem Soc 125:3733
125. Xu Z, Pilch D, Srinivasan A, Olson W, Geacintov N, Breslauer K (1997) Bioorg Med Chem 5:1137
126. Escudé C, Nguyen C, Kukreti S, Janin Y, Sun J, Bisagni E, Garestier T, Hélène C (1998) Proc Natl Acad Sci USA 95:3591
127. Fox KR, Polucci P, Jenkins TC, Neidle S (1995) Proc Natl Acad Sci USA 92:7887
128. Kan Y, Armitage B, Schuster GB (1997) Biochemistry 36:1461
129. Arya D, Coffee R, Willis B, Abramovitch A (2001) J Am Chem Soc 123:5385
130. Jain SS, Polak M, Hud NV (2003) Nucleic Acids Res 31:4608
131. Keppler MD, James PL, Neidle S, Brown T, Fox KR (2003) Eur J Biochem 270:4982
132. Giovannangéli C, Perrouault L, Escudé C, Thuong NT, Hélène C (1996) Biochemistry 35:10539
133. Pilch DS, Breslauer KJ (1994) Proc Natl Acad Sci USA 91:9332
134. Arya DP, Coffee RL, Jr., Charles I (2001) J Am Chem Soc 123:11093
135. Keppler M, Zegrocka O, Strekowski L, Fox KR (1999) FEBS Lett 447:223
136. Strekowski L, Say M, Zegrocka O, Tanious FA, Wilson WD, Manzel L, Macfarlane DE (2003) Bioorg Med Chem 11:1079
137. Arya DP, Xue L, Tennant P (2003) J Am Chem Soc 125:8070
138. Silver GC, Sun JS, Nguyen CH, Boutorine AS, Bisagni E, Hélène C (1997) J Am Chem Soc 119:263
139. Silver GC, Nguyen CH, Boutorine AS, Bisagni E, Garestier T, Hélène C (1997) Bioconjug Chem 8:15

140. Grimm GN, Boutorine AS, Lincoln P, Norden B, Hélène C (2002) Chembiochem 3:324
141. Vinogradov S, Roig V, Sergueeva Z, Nguyen CH, Arimondo P, Thuong NT, Bisagni E, Sun JS, Hélène C, Asseline U (2003) Bioconjug Chem 14:120
142. Schmitt P, Nguyen CH, Sun J-S, Grierson DS, Bisagni E, Garestier T, Hélène C (2000) Chem Comm:763
143. Roulon T, Hélène C, Escudé C (2001) Angew Chem Int Ed 40:1523
144. Bello-Roufaï M, Roulon T, Escudé C (2004) Chem Biol 11:509
145. Kukreti S, Sun J, Loakes D, Brown D, Nguyen C, Bisagni E, Garestier T, Hélène C (1998) Nucleic Acids Res 26:2179
146. Duval-Valentin G, Debizemont T, Takasugi M, Mergny JL, Bisagni E, Hélène C (1995) J Mol Biol 247:847
147. François JC, Saison-Behmoaras T, Barbier C, Chassignol M, Thuong NT, Hélène C (1989) Proc Natl Acad Sci USA 86:9702
148. Zain R, Marchand C, Sun J-S, Nguyen CH, Bisagni E, Garestier T, Hélène C (1999) Chem Biol 6:771
149. Marchand C, Nguyen CH, Ward B, Sun J-S, Bisagni E, Garestier T, Hélène C (2000) Chem Eur J 6:1559
150. Teulade-Fichou MP, Perrin D, Boutorine A, Polverari D, Vigneron JP, Lehn JM, Sun JS, Garestier T, Hélène C (2001) J Am Chem Soc 123:9283

Top Curr Chem (2005) 253: 149–178
DOI 10.1007/b100446

Aminoglycoside–Nucleic Acid Interactions: The Case for Neomycin

Dev P. Arya (✉)

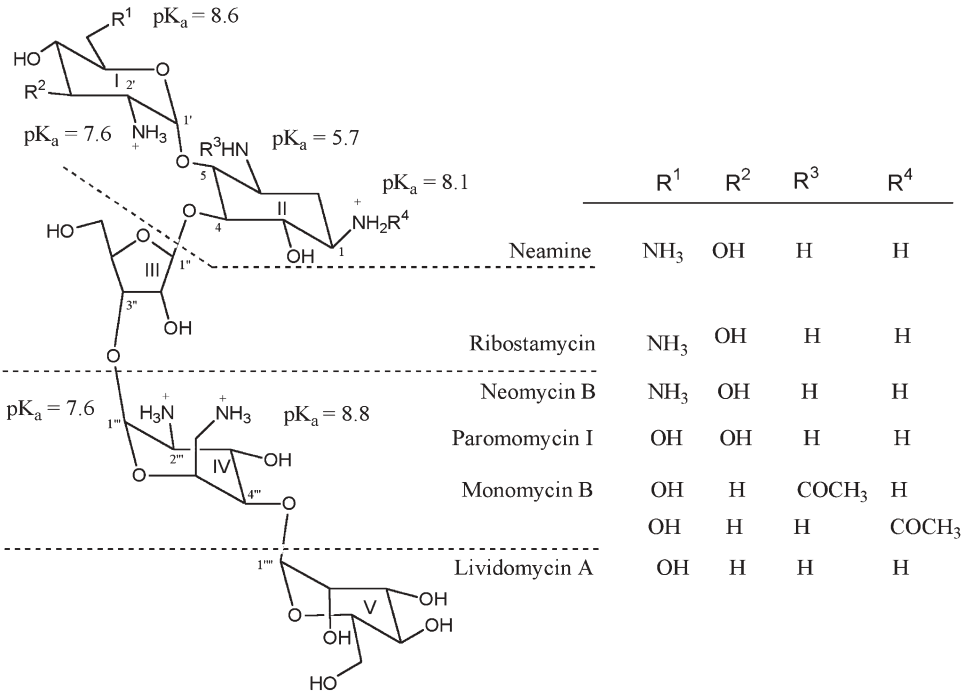
461 Hunter, Laboratory of Medicinal Chemistry, Department of Chemistry, Clemson University, Clemson, SC 29634, USA
dparya@clemson.edu

Abstract Aminoglycoside antibiotics are bactericidal drugs that have been at the forefront of antimicrobial therapy for almost five decades. The past decade (1990–2000) saw a resurgence in aminoglycoside-based drug development as their chemistry/mechanism of action became better understood. This work, however, had almost exclusively focused on targeting RNA. This review summarizes new developments (past 4–5 years) in aminoglycoside–nucleic acid interactions in the broader context of nucleic acid selectivity, not just RNA. Aminoglycoside binding to A-form nucleic acid structures is discussed, as is the development of novel conjugates for major-minor groove recognition of B-form DNA. Neomycin is chosen as the representative aminoglycoside and is revealed here to be an underutilized scaffold in nucleic acid recognition.

Keywords Aminoglycoside · Neomycin–Hoechst · A-form nucleic acids · B-form nucleic acids · PNA

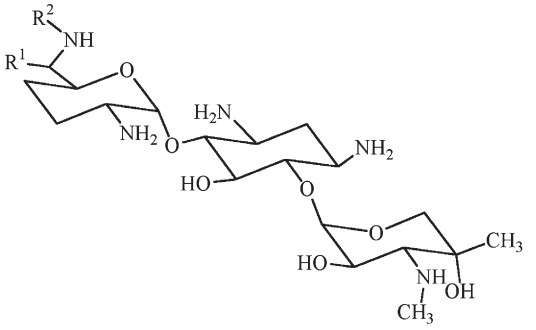
1 Aminoglycosides: An introduction

Aminoglycoside antibiotics (Scheme 1 and Scheme 2) are bactericidal agents that are comprised of two or more amino sugars joined in glycosidic linkage to

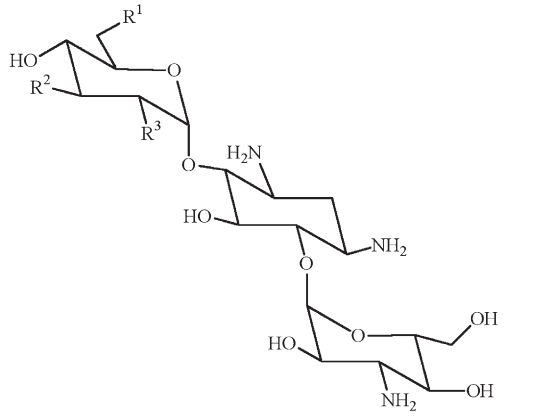


	R^1	R^2	R^3	R^4
Neamine	NH_3	OH	H	H
Ribostamycin	NH_3	OH	H	H
Neomycin B	NH_3	OH	H	H
Paromomycin I	OH	OH	H	H
Monomycin B	OH	H	$COCH_3$	H
	OH	H	H	$COCH_3$
Lividomycin A	OH	H	H	H

Scheme 1 Structures/pKas of aminoglycosides with a central ribose. Slightly different pKa values have been recently reported [6]



	R^1	R^2
Gentamicin C1	CH_3	CH_3
Gentamicin C2	CH_3	H
Gentamicin C1a	H	H



	R^1	R^2	R^3
Kanamycin A	NH_2	OH	OH
Kanamycin B	NH_2	OH	NH_2
Kanamycin C	OH	OH	NH_2
Tobramycin	NH_2	H	NH_2

Scheme 2 Structures of aminoglycosides (kanamycin and gentamicin families)

a hexose nucleus [1]. Though they exhibit a narrow toxic/therapeutic ratio, their broad antimicrobial spectrum, rapid bactericidal action, and ability to act synergistically with other drugs makes them highly effective in the treatment of nosocomial (hospital acquired) infections [2]. They are clinically useful in the treatment of urinary tract infections [3], lower respiratory infections, bacteremias, and other superinfections by resistant organisms [4]. Their greatest potential has been in combination drug regimens for the treatment of infections that are difficult to cure with single agents and for use in patients who are allergic to other classes of drugs [5]. Aminoglycosides (Schemes 1 and 2) contain a unique polyamine/carbohydrate structure, and have attracted considerable attention because of their specific interactions with RNA [6]. The bactericidal action of aminoglycosides is attributed to the irreversible inhibition of protein synthesis following their binding to the 30S subunit of the bacterial ribosome and thus interfering with the mRNA translation process. The miscoding causes membrane damage, which eventually disrupts the cell integrity, leading to bacterial cell death [7–10].

2
Aminoglycosides and Nucleic Acids: The Attraction for RNA?

After the discovery of streptomycin and other aminoglycosides by Selman Waksman in the early to mid-1940s as life-saving antibacterials against tuberculosis [11, 12], considerable efforts were focused on understanding their mechanism of action. In the decades that followed, and through seminal work by Davies and others [13, 14], ribosomal RNA came to be accepted as the target biopolymer responsible for drug action. In the early to late 1990s, as developments in nucleic acid synthesis, combinatorial biosynthesis, and the need for new drugs/targets for infectious diseases emerged, aminoglycosides were shown to bind to various RNA molecules. These include the 5′-untranslated region of thymidylate synthase mRNA [15], both Rev response element and transactivating response element RNA motifs [16–18] of HIV-1, a variety of catalytic RNA molecules such as group I introns [1, 19], ribonuclease P RNA [20], hairpin ribozyme [21, 22], hammerhead ribozyme [23–25], and hepatitis delta virus ribozyme [26, 27]. Aminoglycosides binding to HIV-1 RNA molecules have been shown to prevent binding of the cognate viral proteins Tat and Rev to TAR [28] and RRE [16], respectively. The glucose residues present in glycosylated DNA render the DNA inaccessible for enzymes, and thus help the pathogen escape degradation by host restriction enzymes [29, 30]. The literature of the past decade is rife with a large number of different RNA structures that aminoglycosides have been shown to bind. The reason for this RNA-centered development was understandable: aminoglycosides exhibit their antibacterial action through rRNA binding and show high affinity binding (K_d in the nanomolar range) to such RNAs. RNA rapidly became a target of drug development and discovery of such functional RNAs for drug development was a logical extension of an exploration of their extended activity. What was remarkable, however, was the almost complete absence of reports on the non-RNA structures targeted by aminoglycosides.

2.1
The Need for New Approaches: DNA vs RNA Recognition

RNA recognition has proven to be more challenging than DNA recognition by small molecules. Recognition of DNA•RNA hybrids by small molecules was virtually unexplored at the beginning of this century [31]. DNA-based intercalators and groove binders were the first to be examined for RNA recognition. These approaches met with limited success, due in large part to the different 3-D structures of functional RNA molecules. Sequence-specific RNA recognition has more similarities to recognition principles used in targeting proteins than to DNA duplexes. As with proteins, a distribution of charged pockets can provide a 3-D pattern that can be targeted specifically by compounds exhibiting structural electrostatic complementarity. Aminoglycosides have been shown

to provide complementary scaffolds where the positively charged ammonium groups displace several Mg^{2+} ions from their RNA binding sites [32–38]. An intriguing question in this regard was whether these scaffolds complementary to hairpin RNA structures could be extended for recognition of higher ordered RNA structures (triplexes, tetraplexes), DNA•RNA hybrids (duplex/triplex), and even purely DNA structures (duplex, triplex, tetraplex). Therefore, at the outset of our investigations, we wanted to investigate whether there were any other nucleic acids that aminoglycosides were capable of targeting and to try to understand their recognition principles in the larger context of nucleic acid selectivity. In this paper, using neomycin as the key example, I wish to show that our preliminary work of the past few years provides convincing evidence of the underutilized nature of aminoglycoside scaffolds in nucleic acid targeting.

3
The Nucleic Acid Triplex: Role of Aminoglycosides

The biochemical access to a living organism's genetic information (stored in DNA) is based on specific protein–DNA interactions. Predictive chemical principles for protein–DNA recognition are still considered complex, despite the recent progress using biological selection methods [39–43]. Recognition of duplex DNA by small molecules (minor groove binders-polyamides) [44–49] and oligonucleotides (major groove binders-DNA triple helices) [50–52] are promising alternate approaches to a chemical solution for DNA recognition. Triple strand formation has also been exploited to facilitate the delivery and enhance the sequence specificity of DNA-cutting reagents [51, 53, 54] and drugs [52, 55]. In addition, triple strand formation has been used to modify enzyme cutting patterns by selectively blocking enzyme binding sites in the major groove [56, 57]. In short, appropriately designed and constructed third strand oligonucleotides that hybridize to targeted duplex domains can be used to control gene-expression, serve as artificial endonucleases in gene mapping strategies, dictate or modulate the sequence specificity of DNA-binding drugs, and selectively alter the sites of enzyme activity. Ligands that increase the rates of association of a triplex forming oligonucleotide (TFO) to a target duplex thus have enormous potential in drug development and as tools for molecular biology.

Triple helix formation (see Fig. 1 for H-bonding in different types of triple helical structures) has been the focus of considerable interest because of possible applications in developing new molecular biology tools as well as therapeutic agents [58–64], and the possible relevance of H-DNA structures in biological systems [51, 65–67]. Intermolecular triplexes have aroused considerable interest as potential inhibitors of the expression of particular genes, since a sequence of either third-strand pyrimidines or purines, when 16–18 base pairs long, can be sufficient to be unique for recognition and binding to

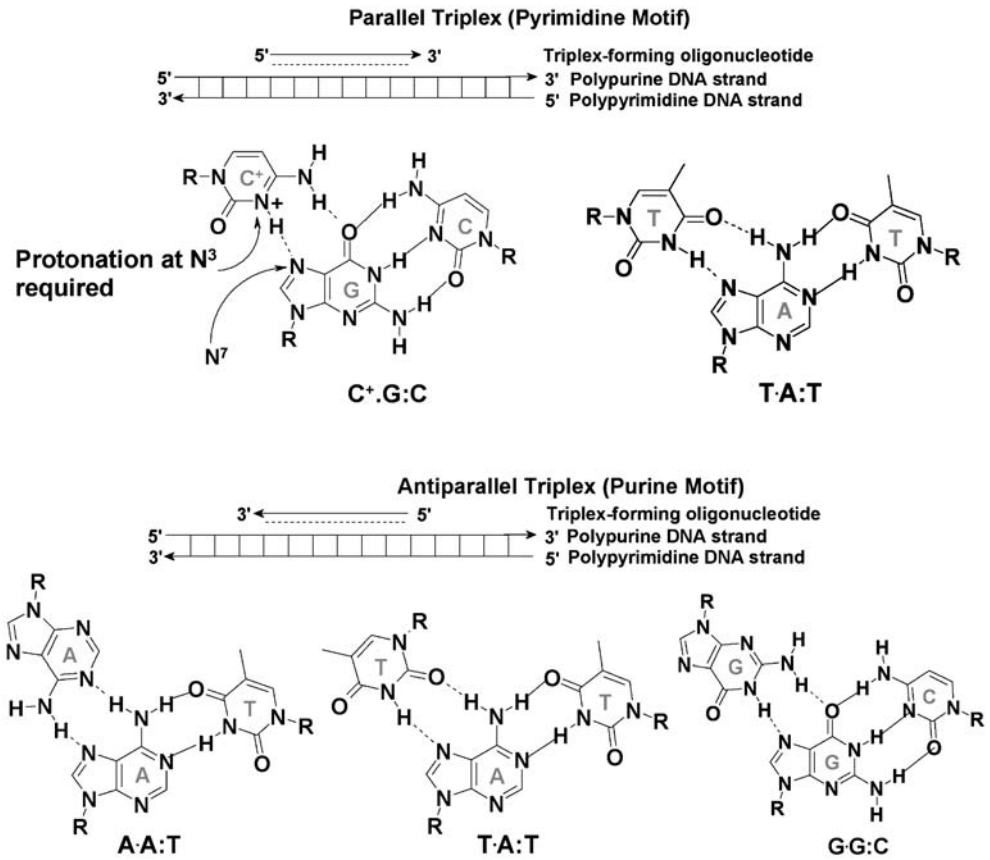


Fig. 1 Base interactions in parallel (pyrimidine motif *top*) and antiparallel (purine motif *bottom*) triple helices

defined single sites in a genome [52]. A number of experiments have now been reported that demonstrate the feasibility of the concept [68, 69].

$$\mathrm{dA{\bullet}dT} + \mathrm{dT} \rightleftharpoons \mathrm{dA{\bullet}2dT} \tag{1}$$

Association of a third strand with a duplex, however, is a thermodynamically weaker and a kinetically slower interaction than duplex formation itself (Eq. 1) [70, 71].

A number of intercalators, groove binders and polyamines have been used to stabilize triple helices [72–111]. The design of ligands that bind strongly to triple-helical structures and have a high discrimination between triplexes and duplexes opens new possibilities to control gene expression at the transcriptional level. There is a significant amount of high-resolution information on complexes of compounds that bind to both DNA and RNA by intercalation, and on compounds that bind in the DNA minor groove [48, 112]. Good models

exist for proteins and peptides that bind in the major groove of DNA and RNA [40–42]. There is, however, little information available for antibiotics that selectively bind DNA triplex grooves or RNA triplex grooves. We have recently reported on the importance of neomycin in narrowing the disparity between groove recognition of duplex versus triplex nucleic acids [113, 114]. Neomycin was shown as one of the first examples that bridge this gap and thus may lead to a novel understanding of the recognition principle(s) involved in selective targeting of triplex grooves. These results have shown that neomycin selectively stabilizes the DNA triplex without any effect on the DNA duplex.

3.1
Effect of Neomycin on a Polynucleotide Triplex

Neomycin selectively stabilizes DNA triplex without affecting the duplex [113–116]. Increasing the molar ratios of neomycin from 0–25 μM, r_{db} (ratio drug [neomycin]/base triplet)=1.67, increases the triplex melting point by nearly 25°C, whereas the duplex is virtually unaffected (Fig. 2).

3.2
Thermal Denaturation Studies with Poly(dA)•2poly(dT) in the Presence of Other Aminoglycosides and Diamines

Thermal analysis of poly(dA)•2poly(dT) in the presence of other aminoglycosides is shown as a bar graph in Fig. 3. At high concentrations (r_{db}=0.66–1.67, Fig. 3), most aminoglycosides with five or more amines are able to stabilize the triple helix (increasing $\Delta T_{m3\rightarrow2}$, without significantly affecting the $\Delta T_{m2\rightarrow1}$ values). The difference between the effectiveness of paromomycin and neomycin is quite remarkable. The structural difference between the two is a positively

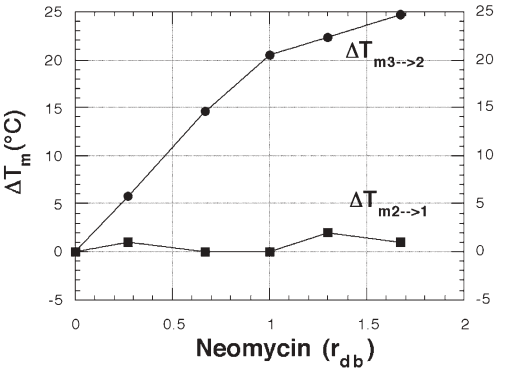


Fig. 2 Variation of triplex melting ($T_{m3\rightarrow2}$) and duplex melting ($T_{m2\rightarrow1}$) of poly(dA)•2poly(dT) as a function of increasing neomycin concentration; r_{db}=drug [neomycin]/base triplet ratio. Reprinted with permission from J Am Chem Soc (2001) 123(23):5385

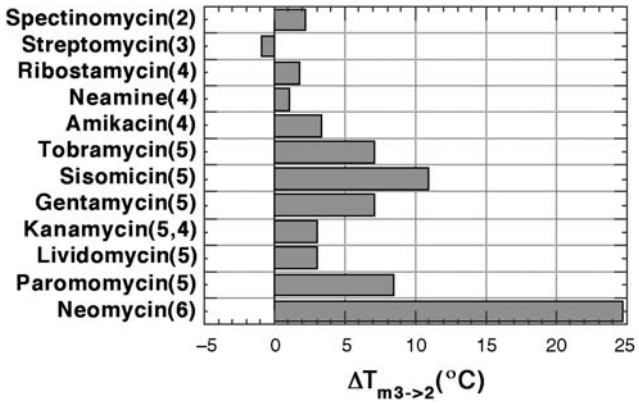
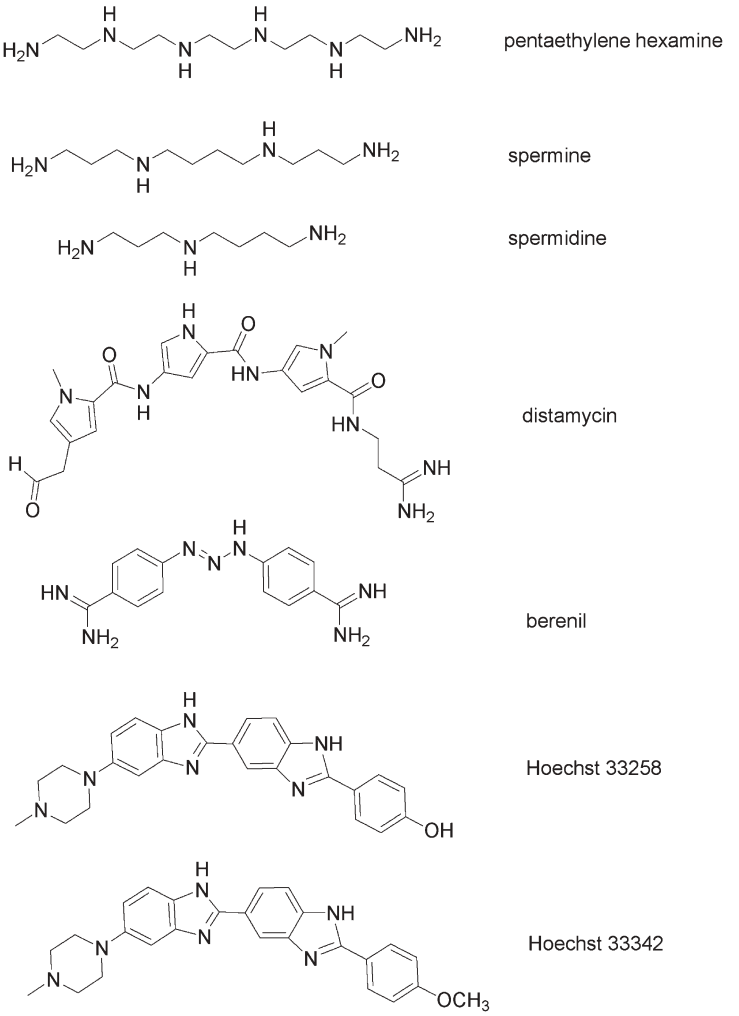


Fig. 3 Effect of aminoglycoside antibiotics on the melting of poly dA•2polydT triplex (r_{db}=1.67). Number of amines in each antibiotic is shown in parenthesis. Reprinted with permission from J Am Chem Soc (2001) 123(23):5385

charged amino group (present in neomycin), replacing a neutral hydroxyl (present in paromomycin). This leads to a difference of 10°C in $T_{m3\rightarrow2}$ values (r_{db}=0.66) and a difference of 16°C at r_{db}=1.67 [114]. At lower concentration of antibiotics (r_{db}=0.26), paromomycin has little effect on the stability of the triplex. Lividomycin, a paromomycin analog with a polyhydroxy hexose tether, is slightly less effective than paromomycin in increasing $T_{m3\rightarrow2}$ values under these conditions.

3.3 Stabilization of DNA Triple Helix Poly(dA)•2poly(dT) by Other DNA Groove Binders

In order to assess how neomycin compares to other ligands in stabilizing triplexes, thermal denaturation analyses of poly(dA)•2poly(dT) triplex in the presence of previously studied DNA minor groove binders (Scheme 3) has also been performed (Fig. 4). A comparison with groove binders, shown in Fig. 5, indicates that neomycin is much more active than the minor groove binders (berenil, DOC, DODC, DAPI, Hoechst 33258, Hoechst 33342). The minor groove binders previously studied have little preference for triple helix (berenil, distamycin and Hoechst dyes). Most groove binders stabilize the duplex as well as the triplex (Hoechst, berenil, distamycin) and some even destabilize the triplex (berenil, distamycin). The groove-binding ability of neomycin was extremely unique and presented a novel mode of triplex recognition. Neomycin, as opposed to other groove binders, differentiated the triplex grooves from those present in the duplex (Fig. 4).



Scheme 3 Structures of some groove binders known to bind duplex DNA

3.4 Thermodynamics of Drug Binding to the DNA Triplex (ITC)

An ITC-derived thermodynamic profile for neomycin binding to 12-mer intramolecular DNA triplex gave a binding constant of 2.0×10^5 M^{-1} (Fig. 5 and Table 1). The complexation is enthalpy-driven (81%), with little entropic contributions. The binding is salt-dependent, with higher salt leading to a decrease in the association equilibrium constant. A much higher binding constant of neomycin is observed with other nucleic acids-RNA triplex/DNA tetraplex (unpublished results).

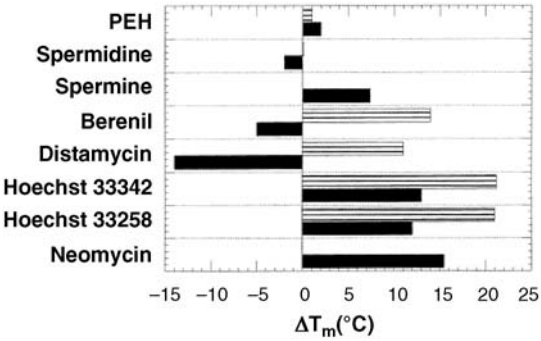


Fig. 4 Effect of 10 μM (r_{db}=0.66) groove binders on the DNA triplex melt, poly(dA)•2poly(dT) (*black bars*) and the duplex melt, poly(dA)•poly(dT) (*striped bars*). Distamycin does not show $T_{m3\rightarrow2}$ transition (20°C). *PEH* Pentaethylene hexamine. Reprinted with permission from J Am Chem Soc (2001) 123(23):5385

Table 1 ITC-derived thermodynamic profiles for the binding of neomycin to 5′–dA_{12}-x-dT_{12}-x-dT_{12}–3′ triple helix in 10 mM sodium cacodylate 0.5 mM EDTA, 150 mM KCl, pH 6.8 at 20°C

T (K)	ΔK ($\times10^5$ M^{-1})	ΔH (kcal•mol^{-1})	TΔS (kcal•mol^{-1})	ΔG (kcal•mol^{-1})	N (drug/triplex)
293	1.96±0.13	−6.9±0.3	0.18	−7.1±0.04	2.17±0.09

Reprinted with permission from J Am Chem Soc (2003) 125(13):3733.

3.5 CD/Molecular Modeling

Previous studies of neomycin have shown that it has a marked preference for binding to the larger Watson–Hoogsteen (W-H) groove of the triplex [116]. Ring I/II amino groups and Ring IV amines were proposed to be involved in the recognition process. CD/ITC studies indicate a five base triplet/drug binding site. The novel selectivity of neomycin was shown to be a function of its charge and shape complementarity to the triplex W-H groove (Fig. 6) [116].

A large number of molecules had previously been studied for triplex recognition. Before we began our investigations on aminoglycoside–triplex interactions, the goal of triplex-selective groove recognition had remained elusive. Neomycin has been shown to be the first molecule to selectively stabilize DNA triplex structures that include polynucleotides, small homopolymer, as well as mixed base triplexes [116]. This stabilization was shown to be based on neomycin's ability to bind triplexes in the groove with high affinity (based on viscometric and ITC titrations). Modeling/physicochemical results suggested a

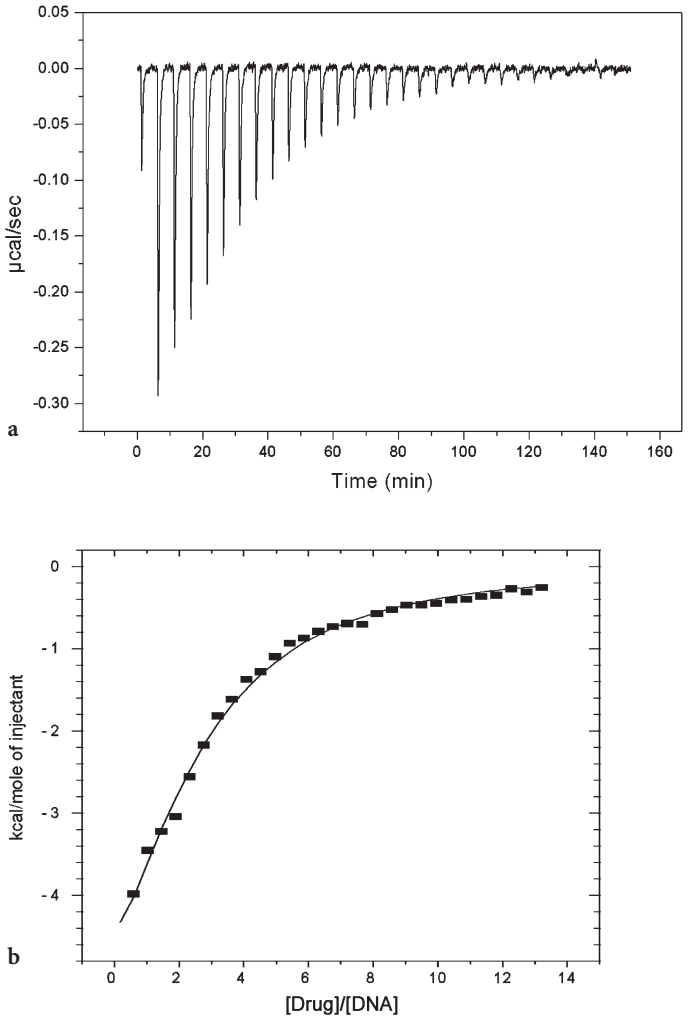


Fig. 5 **a** ITC profile of 5′-dA$_{12}$-x-dT$_{12}$-x-dT$_{12}$-3′ (4 µM/strand) titrated with neomycin (500 µM) in 10 mM sodium cacodylate, 0.5 mM EDTA, 150 mM KCl, pH 6.8 at 20°C. **b** Corrected injection heats plotted as a function of the [drug]/[DNA] ratio. The corrected injection heats were derived by integration of the ITC profile shown in Fig. 5a, followed by subtraction of the corresponding dilution heats derived from control titrations of drug into buffer alone. The *data points* reflect the experimental injection heats, while the *solid line* reflects calculated fit of the data. Reprinted with permission from J Am Chem Soc (2003) 125(13):3733

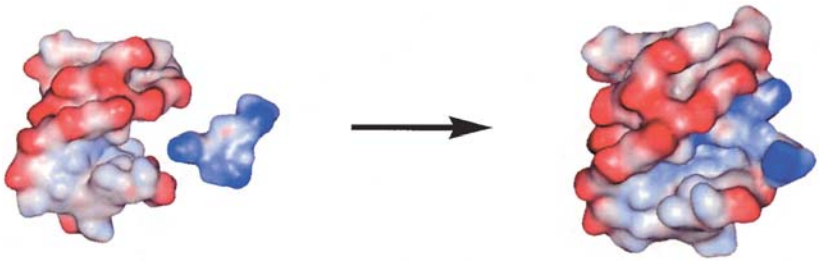


Fig. 6 Charge/shape complementarity of neomycin to the triplex W-H groove: Electrostatic surface potential maps of neomycin approaching the W-H groove of the triplex (*left*), and neomycin buried in the triplex groove (*right*). Reprinted with permission from J Am Chem Soc (2003) 125(13):3733

further preference of neomycin binding to the larger W-H groove. These findings will further contribute to the development of a new series of triplex-specific (DNA/RNA and hybrid) ligands, which may contribute to either antisense or antigene therapies.

4 DNA•RNA Hybrids

RNA•DNA hybrid duplexes are the primary targets for important enzymes that include ribonuclease-H and reverse transcriptase [117, 118]. Stable RNA•DNA triplexes normally adopt an A-type conformation and have been shown to inhibit RNA polymerase [119], DNAase-I, and RNAse [120]. Only six of the eight possible combinations of triplexes are stable under physiological conditions [121, 122]. Stabilization of poly(rA)•2poly(dT) and 2poly(rA)•poly(dT) triplexes can only be achieved under molar salt conditions [123]. Since these two triplexes could not be studied under the lower salt conditions used in competition dialysis assay, we investigated the effect of neomycin on these two triplexes using UV and CD thermal denaturation studies. Neomycin has been shown to stabilize the hybrid poly(rA)•poly(dT) duplex [113], and even induce poly(rA)•2poly(dT) triplex formation [113], much more effectively than previously reported ligands [100]. The effect of aminoglycosides on hybrid duplex and triplex structures showed that almost all aminoglycosides stabilized the hybrid poly(dA)•poly(dT) duplex (see Fig. 7). It is noteworthy that formation of these triple helices require molar salt in the absence of the drug, whereas micromolar neomycin concentration can induce the triplex formation. Recently, work from the Pilch lab [124] has also corroborated these findings and shown a high binding constant (10^7 M^{-1}) for aminoglycoside binding to small RNA•DNA hybrids.

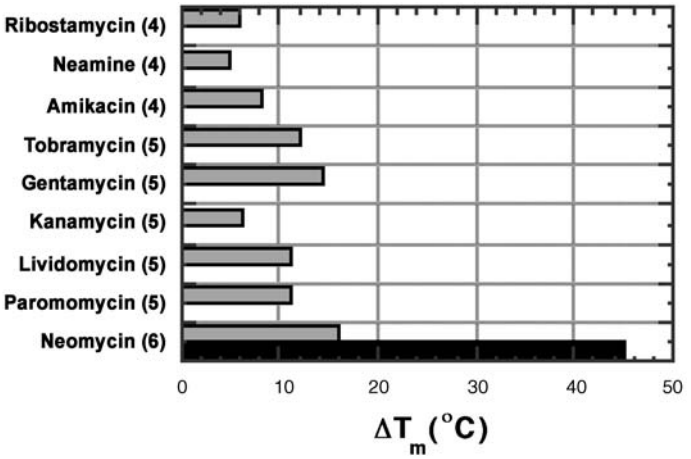
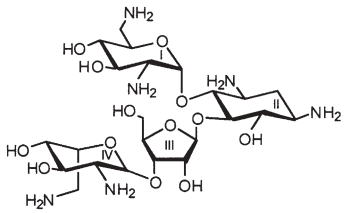


Fig. 7 Effect of added aminoglycoside (r_{db}=0.66) on the stabilization of rA•dT duplex (*gray*) and on inducing rA•2dT triplex (*black*). Number of amines in each aminoglycoside is shown in parenthesis. $\Delta T_{m3\rightarrow2}$ is calculated by assuming a $T_{m3\rightarrow2}$ of 10°C in the absence of neomycin (no transition seen). Reprinted with permission from J Am Chem Soc (2001) 123(44):11093

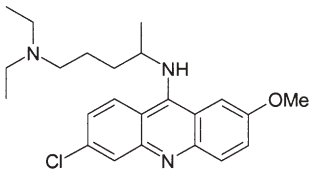
5
The A-Form Nucleic Acids

5.1
Competition Dialysis of Neomycin–Acridine Conjugate with Nucleic Acid Forms

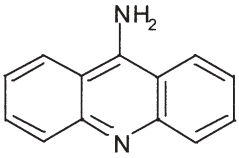
The remarkable ability of neomycin and other aminoglycosides to stabilize DNA, RNA, and hybrid triple helices has been reported and discussed above [113–116]. Neomycin has also been shown to induce the stabilization of hybrid duplexes as well as hybrid triple helices [113]. This significantly added to the number of nucleic acids (other than RNA) that aminoglycosides have been shown to target. A clear requirement then arose for a quantitative assay to determine the relative binding affinities for host triplex, duplex DNA, single-stranded (SS) DNA/RNA and other possible nucleic acid targets (tetraplex) for a given aminoglycoside ligand. Fortunately, a rapid technique has been established by Chaires for this exact purpose, using a thermodynamically rigorous competitive equilibrium dialysis method that exploits therapeutically useful drug concentrations [91, 125]. In the assay, solutions consisting of identical concentrations of different nucleic acid structures were dialysed simultaneously against a common ligand dissolved in appropriately buffered conditions. After equilibration, the amount of ligand bound to each DNA was measured by spectrophotometry. More ligand accumulated in the dialysis tube containing the structural form of highest binding affinity and, since all of the DNA samples were in equilibrium with the same free ligand concentration, the amount of ligand bound was directly proportional to the binding constant for each con-



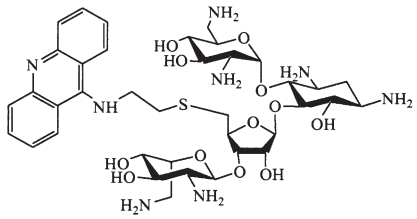
Neomycin



Quinacrine



9-aminoacridine



neomycin-acridine conjugate

Fig. 8 Structures of neomycin, aminoacridines, and the neomycin–acridine conjugate

formational form. Thus, comparison among the DNA samples gave a rapid and thermodynamically reliable indication of structural selectivity for any given ligand (an updated review on competition dialysis also appears in this volume).

Since aminoglycosides do not have a chromophore for spectrophotometric analysis, competition dialysis of three acridines with increasing positive charge was used to decipher aminoglycoside specificity (Fig. 8). Competition dialysis studies were carried out using 9-aminoacridine, quinacrine, and a neomycin–acridine (neo-acridine) conjugate [126] against 14 different nucleic acids. Going from acridine to neo-acridine, we were able to parse the effect of neomycin conjugated to the acridine chromophore. At first sight, dialysis of neo–acridine (Fig. 9) showed highly promiscuous binding with little preference for any specific nucleic acid structure, except for a clear preference for RNA triplex. Among comparable single strand, duplex, and triplex structures, maximum binding was always observed with the triplexes. This seemingly promiscuous binding yielded a different picture upon careful analysis of the dialysis data. All three drugs showed comparable binding to one nucleic acid: calf thymus DNA. Calf thymus DNA also represents a standard duplex DNA. This observation was used to replot the dialysis results to emphasize differences relative to that standard. These results, shown in Fig. 10, better illustrate the change in specificity of the different acridines toward different nucleic acids. While 9-aminoacridine and quinacrine showed a clear preference for DNA triplex, neo-acridine binding to RNA triplex is much greater than DNA triplex and even better than the natural aminoglycoside RNA target: eubacterial 16S A-site. Drug binding was

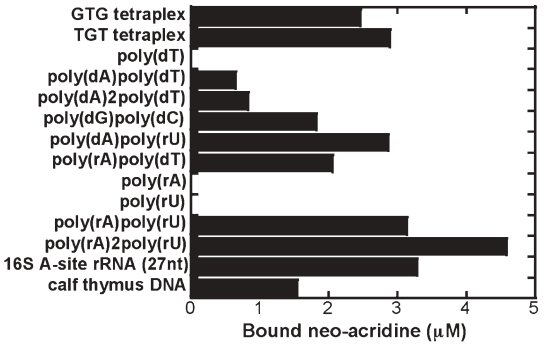


Fig. 9 Competition dialysis results of neo-acridine (1 μM) with various nucleic acids; 180 μL of different nucleic acids (75 μM per monomeric unit of each polymer) were dialyzed with 400 mL of 1 μM neo-acridine in BPES buffer (6 mM Na_2HPO_4, 2 mM NaH_2PO_4, 1 mM Na_2EDTA, 185 mM NaCl, pH 7.0) solution for 72 h. Reprinted with permission from J Am Chem Soc (2003) 125(34):10148

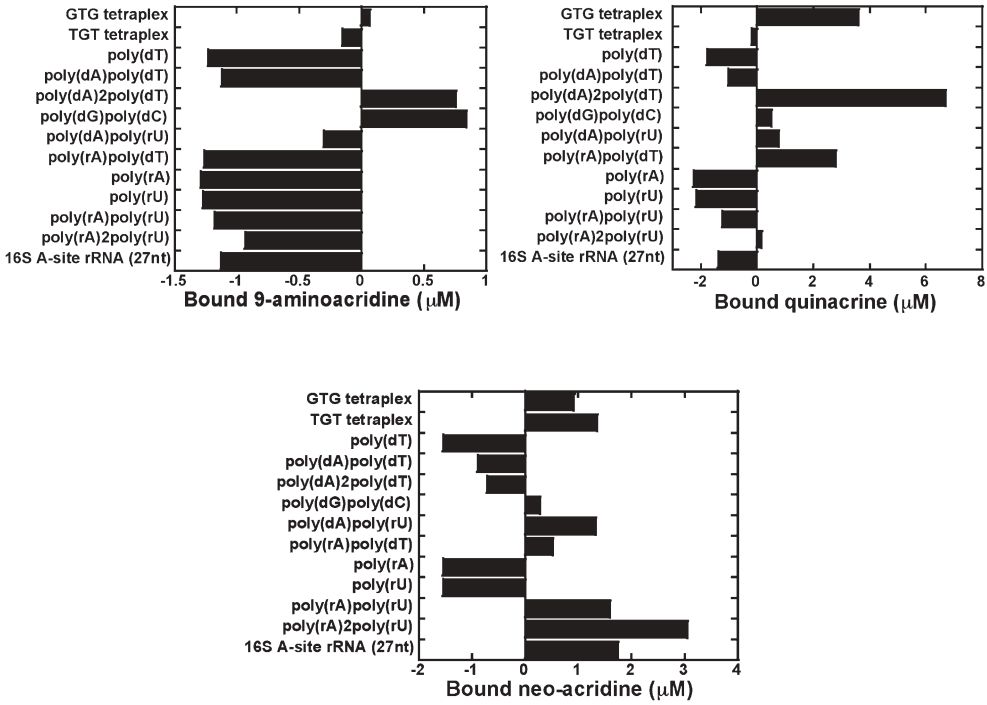


Fig. 10 Competition dialysis results (difference plots, with calf thymus DNA as reference) of 9-aminoacridine, quinacrine, and neo-acridine (1 μM) with various nucleic acids. Experimental conditions were identical to those for Fig. 6. Maximum binding of neo-acridine is observed with nucleic acids that can adopt the A-type conformation. Reprinted with permission from J Am Chem Soc (2003) 125(34):10148

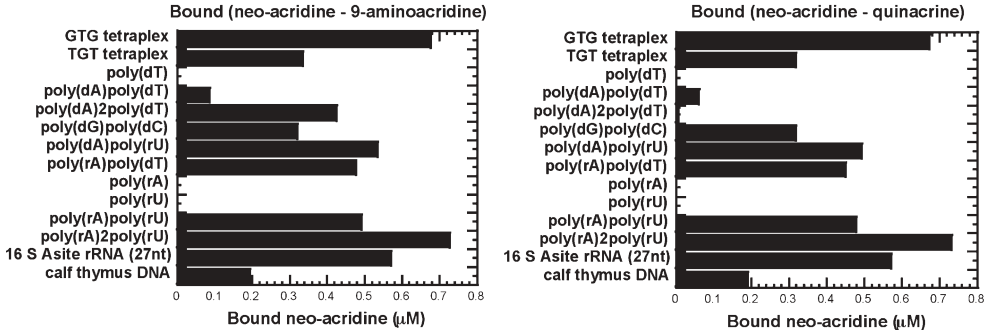


Fig. 11 Competition dialysis results of 100 nM drug: difference plots, neo-acridine minus 9-aminoacridine (*left*) and neo-acridine minus quinacrine (*right*). 180 μL of different nucleic acids (7.5 μM per monomeric unit of each polymer) were dialyzed with 400 mL of 100 nM ligand in BPES buffer (6 mM Na_2HPO_4, 2 mM NaH_2PO_4, 1 mM Na_2EDTA, 185 mM NaCl, pH 7.0) solution for at least 24 h

also observed with DNA as well as RNA duplex, and even with DNA tetraplex. The binding to DNA tetraplex was still lower than to the RNA triplex. RNA•DNA duplexes were better targets than DNA homoduplexes; poly(dA)•poly(rU) hybrid duplex being comparable in binding to the tetraplexes. Previous studies with aminoglycoside natural products have shown no effect on the stability of A•T-rich duplex DNA (in the presence of salt), suggesting weaker, nonproductive binding. Triplexes, then, are the targets of choice for neomycin. Neo-acridine shows a remarkable binding preference to RNA triplex that has not previously been observed. A big surprise, however, was the significant binding observed with the poly(dG)•poly(dC) duplex.

A competition dialysis assay using tenfold (100 nM) and 100-fold (10 nM) lower concentrations (nanomolar range) was also carried out. Results from dialysis under 100 nM drug concentration (Fig. 11) showed that neo-acridine favors nucleic acid forms that can adopt an A-type conformation. However, reliable results could not be obtained at 1 nM and 10 nM concentrations due to the low fluorescence intensity of the neo-acridine conjugate.

Neo-acridine binding to RNA triplex was also investigated by UV thermal melts, ITC, viscometric and CD titrations. Thermal denaturation in the presence of neo-acridine showed an increase in T_{m3-2} at low drug concentrations. At higher drug concentrations, the duplex was stabilized as well. We have previously shown neomycin to be one of the best stabilizers of an RNA triple helix [114]. Viscosity measurements showed a clear groove binding (as seen by shortening of RNA triplex length) upon titration of neomycin as well as neo-acridine into the triplex [127].

5.2
The Common Thread that Holds Together RNA Duplex/Triplex, DNA-RNA Hybrid Duplexes, DNA Tetraplexes, and the Poly(dG)•poly(dC) Duplex is the Propensity Towards an A-type Conformation

RNA duplex structures are known to adopt an A-type conformation, as are hybrid duplexes [128]. dG•dC-rich DNA duplex sequences [129] have also been shown to have a high propensity for A-form in the presence of cations, including neomycin [130], and CD studies have suggested the A-like solution conformation of G4 tetraplexes [131]. Further evidence of A-type preference was observed with the change in the CD spectrum of poly(dG)•poly(dC) upon inclusion of neo-acridine. A shift in λ_{max} from 257 nm to 267 nm, and increased signal in this range, in the presence of this drug, was strongly indicative of a B–A transformation, as observed by Wang [130] as well as Kypr [129, 131] in similar CD experiments. Additionally, the differences in binding to DNA•RNA hybrids can be attributed to the fact that poly(dA)•poly(rU) has been known to adopt an A-type conformation whereas poly(rA)•poly(dT) can exist in the B-form [128].

5.3
Importance of A-Form DNA and its Recognition

The polymorphism of DNA was noticed early after the discovery of its double helical structure [132]. The conformations of DNA have since been limited to two major distinctions: A-DNA and B-DNA (other less-well-known structures do exist). Both structures are of identical topology and hydrogen bonding patterns, but they differ largely in their overall shape (Fig. 12). B-DNA has long

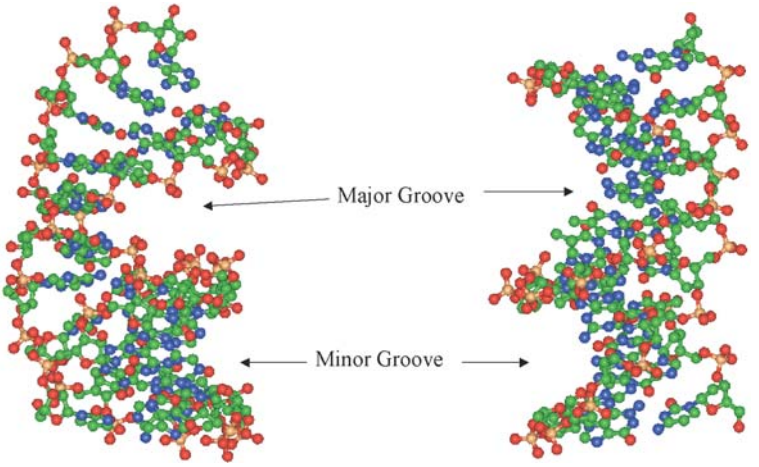


Fig. 12 Conformations of an A-type duplex (*left*) and a B-type duplex (*right*), generally seen for RNA•RNA and DNA•DNA duplexes, respectively. The B-form duplex has a much wider major groove

been believed to be the dominant biological conformation, implementing water molecules and biological cations appropriately within its structure. A-DNA, on the other hand, requires dehydrated conditions. The transition of B- to A-DNA is a reversible and cooperative process [133], in which the A-form is considered the higher-energy state. The underlying factors for this instability have been addressed, but with little success [134, 135].

Native DNA, which comprises the genetic information of all known free organisms, mostly adopts B-form under physiological conditions because it is associated with high humidity in fibers or with aqueous solutions of DNA. However, it is important to switch B-DNA into the A-form in a living organism since constitutive conformation of double-stranded RNA is predominantly A-type. RNA probably preceded DNA in evolution [136], so the basic mechanisms of genetic information copying are likely to have evolved on an A-form rather than B-form. In fact, the template DNA is induced by many polymerases into A-form at positions of genetic information copying in the microenvironment. Thus, DNA switching into A-form may influence replication and transcription of the genomes.

Understanding aminoglycoside A-form nucleic acid interactions then has underlying importance to the area of drug development as well as to the fundamental understanding of A-form recognition because:

1. Novel nucleic acid therapeutic targets can be identified with a better understanding of the thermodynamics and kinetics of molecular recognition involved in aminoglycoside specificity. We have already initiated such a program in the development of novel antimicrobial agents targeting novel RNA and DNA sequences [137, 138].
2. As opposed to B-form DNA recognition, very few small molecules (multivalent cations) [133, 139, 140] are known that select for A-form structural features. Aminoglycosides present a novel scaffold for groove recognition of A-form structures.
3. Aminoglycoside binding to such higher order structures (H-DNA triplex) has also been implicated in their toxic side effects [114]. A better understanding of aminoglycoside binding and selectivity can also help in a better understanding of toxic side-effects of these broad spectrum antibiotics.

6
From A- to B-Form Nucleic Acids: Using Organic Chemistry to Tune Aminoglycoside Selectivity

Aminoglycosides most likely bind in the major groove of A-form structures (much like RNA, as the A-form nucleic acids have a narrower major groove) [127]. The B-form duplex has a much larger major groove and does not provide a good shape complementarity for aminoglycoside binding (see Fig. 13). These findings have complemented the success in development of DNA duplex-spe-

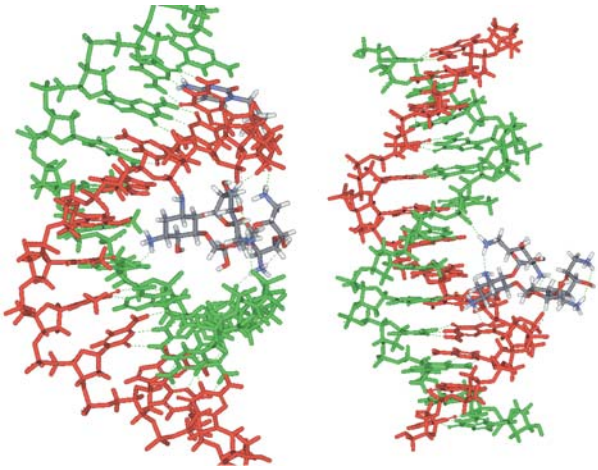
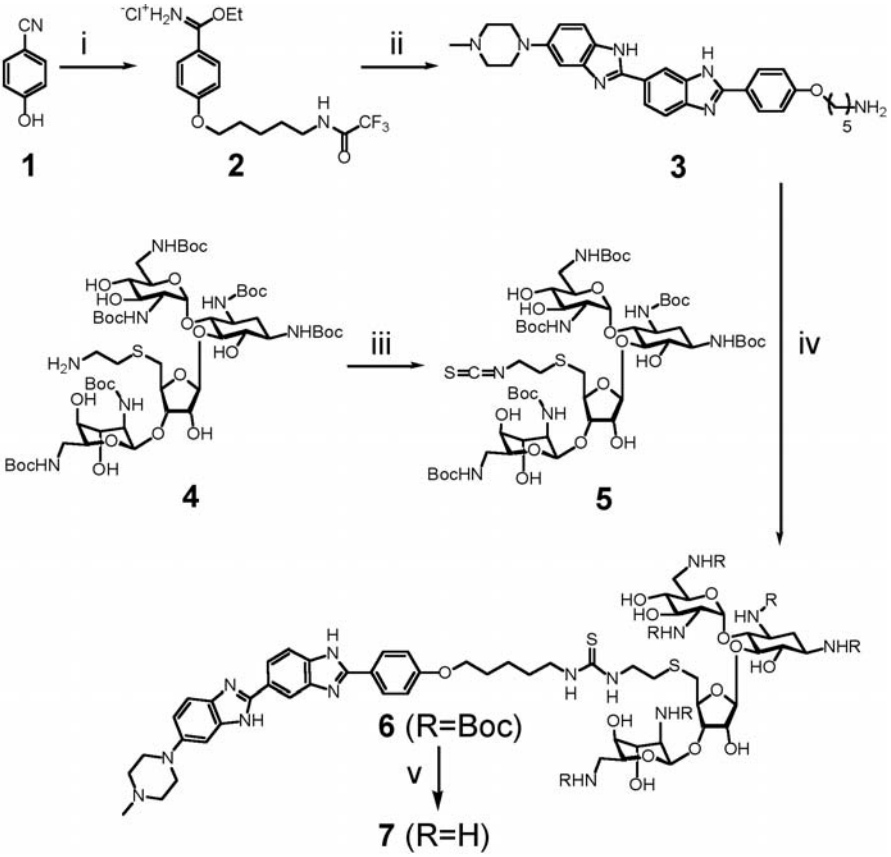


Fig. 13 Charge and shape complementarity of neomycin to the A-form major groove: Computer models of neomycin docked in the major groove of A-form DNA (*left*), and neomycin buried in the B-form major groove (*right*)

cific groove binders in the past few decades, among which netropsin, distamycin and Hoechst 33258 have been the lead compounds. We wished to investigate whether a molecule like neomycin could be forced into the B-form DNA major groove. Therefore, conjugation of neomycin to Hoechst 33258 was accomplished. Another intriguing question in this regard was whether the binding would be driven by Hoechst 33258 (duplex-selective groove binder) or neomycin (triplex-selective groove binder). Such ligands with minor/major groove recognition are promising drug candidates for development of inhibitors of transcription factors [141]. To answer these questions, the synthesis and nucleic acid binding of a novel neomycin–Hoechst 33258 conjugate has been recently reported. The conjugate showed remarkable stabilization of DNA duplexes and destabilization of the DNA triplex.

Starting from the natural product neomycin B, which is commercially available as the tri-sulfate salt, Boc (*t*-butoxycarbonyl) protection of the six amino groups followed by conversion to 2,4,6-triisopropylbenzenesulfonyl derivative, and subsequent substitution by aminoethanethiol, gave rise to the protected neomycin amine [126] compound **4**. Treatment of **4** with 1,1′-thiocarbonyldi-2(1*H*)-pyridone using a catalytic amount of DMAP gave isothiocyanate derivative **5**, which was coupled with bis(benzimidazole) **3** and deprotected to give conjugate **7** (Scheme 4).

The thermal stability of DNA triple and double helices in the presence of neomycin, Hoechst 33258, and neomycin–Hoechst 33258 conjugate **7** was investigated using thermal denaturation monitored by UV absorbance. It was found that **7** displayed a marked effect on the stability of poly(dA)•poly(dT) duplex when compared to both neomycin (which is known to have no effect on



Scheme 4 Reagents and conditions: *i a* 5-trifluoroacetamido-1-pentanol, PPh_3, DIAD, dioxane, r.t., 2 h, 84%; *i b* HCl, EtOH, 0°C, quant.; *ii a* 2-(3,4-diaminophenyl)-6-(1-methyl-4-piperazinyl) benzimidazole, HOAc, reflux, 4 h, 38%; *ii b* K_2CO_3 in 5:2 MeOH:H_2O, r.t., overnight, 94%; *iii* 1,1′-thiocarbonyldi-2(1*H*)-pyridone, cat. DMAP, CH_2Cl_2, r.t. 20 h, 95%; *iv* **3**, pyridine, r.t., overnight, 72%; *v* 1:1 CH_2Cl_2, TFA, r.t., 3 h, quant. Reprinted with permission from J Am Chem Soc (2003) 125(41):12398

the thermal stability of duplex DNA) and Hoechst 33258, which displayed some degree of stabilization of duplex DNA (Fig. 14).

In the absence of ligand, the melting profile of poly(dA)•2poly(dT) is biphasic with $T_{m3\rightarrow2}$=34°C and $T_{m2\rightarrow1}$=72°C. As depicted in Fig. 14, the dissociation of duplex DNA in the presence of **7** occurs at a higher temperature (>95°C) than that of DNA in the presence of Hoechst 33258 (86°C) and neomycin (72°C, unchanged when compared to native duplex melting). This suggests that **7** stabilizes the duplex better than the individual parent compounds. Samples containing both neomycin and Hoechst 33258 displayed no difference in T_m from that observed with the individual molecules. It is important to note that triplex melting was not observed for poly(dA)•2poly(dT) in the presence

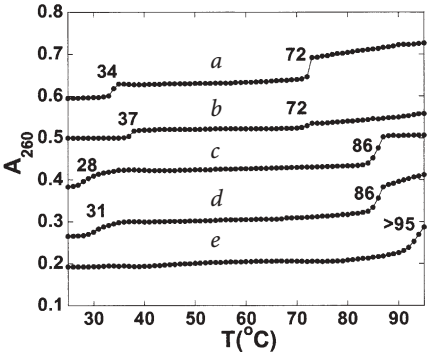


Fig. 14 UV melting profile of poly(dA)•2poly(dT) in the absence (*a*) and presence of 2 μM neomycin (*b*), 2 μM Hoechst 33258 (*c*), 2 μM Neomycin+2 μM Hoechst 33258 (*d*), and 2 μM Hoechst–neomycin conjugate (*e*). Samples of DNA (15 μM/base triplet) in buffer (10 mM Na cacodylate, 0.5 mM EDTA, 150 mM KCl, pH 7.20) containing ligand were analyzed for UV absorbance at 260 nm from 20–95°C using a temperature gradient of 0.2°C min^{-1}. Reprinted with permission from J Am Chem Soc (2003) 125(41):12398

of **7**, suggesting that drug binding prevents the third strand polypyrimidine from binding in the major groove. A comparison was then made with a self-complementary DNA duplex $d(CGCAAATTTGCG)_2$ well known for Hoechst 33258 affinity [142]. UV melting showed increased stability of the duplex in the presence of **7**, with a ΔT_m=25°C, compared to ΔT_m=14°C for Hoechst 33258 [142].

Further studies of numerous duplex DNA 22-mers of varying G/C content (breaking up stretches of A/T base pairs) were carried out. In all cases where stretches of at least 4 base pairs were present, ΔT_m for **7** was at least 10°C higher than that for Hoechst 33258. Duplex stabilization by **7** followed the selectivity shown by Hoechst 33258 (Fig. 15a), whereas neomycin had no effect on the stabilization of any duplex. Hoechst 33258 is well known to have a primary preference for A/T stretches as low as four base pairs, suggesting that the binding-induced thermal stabilization by **7** is largely controlled by the Hoechst 33258 moiety's ability to bind to its required stretch of A/T base pairs. A model depicting the possible binding of **7** to a 12-mer duplex is shown in Fig. 15b. Computer modeling suggests that electrostatic and H-bonding contacts between neomycin and sites within the major groove compete somewhat with the otherwise deep minor groove binding of Hoechst 33258 (Fig. 15b). As Hoechst 33258 binds in the minor groove, neomycin is unable to be completely buried in the major groove (due to the linker size). Despite this constraint, conjugate **7** prefers the duplex, suggesting that neomycin can be forced into the major groove of a B-form DNA duplex. In retrospect, this could be due primarily to the larger binding constants observed between Hoechst 33258 and duplex DNA [142] (~10^8 M^{-1}) as opposed to neomycin binding to triplex (10^5–10^6 M^{-1}) [116]. Conjugates of different linker sizes can then perhaps be designed to target

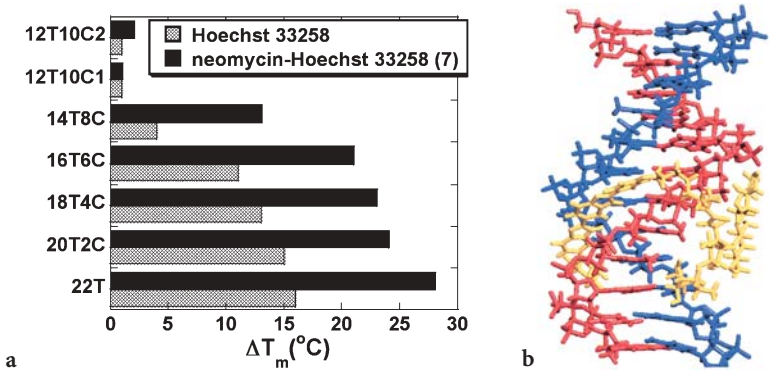


Fig. 15 **a** Bar graph of ΔT_m for 22-mer duplexes in the presence of 4 μM Hoechst 33258 and 4 μM neomycin–Hoechst 33258 7 obtained from UV melting profiles (solution conditions were identical to those for Fig. 14). Reprinted with permission from J Am Chem Soc (2003) 125(41):12398. **b** Computer model of neomycin–Hoechst 33258 (*yellow*, linker atoms shown in *white*) docked in the DNA major-minor grooves. Reprinted with permission from J Am Chem Soc (2003) 125(41):12398

a structure of preference and should aid in the development of even more selective and potent conjugates. Development of such dual recognition ligands opens up new avenues in targeting nucleic acids and is being further explored in our laboratories.

7
Targeting Nucleic Acids with Aminoglycoside–DNA and PNA Conjugates

7.1
RNA Sequence-Specific Aminoglycoside–ODN Conjugates

RNA has now become a well-established drug target [36, 143, 144]. Small molecules and antisense oligonucleotides are now being used to down-regulate gene expression. Vitravene, the first antisense drug, was approved by the FDA at the end of the 20th century [145]. RNA has distinct advantages in antibacterial and antiviral treatment. Primarily, appearance of drug resistance through point mutations in a conserved RNA motif among bacteria or viral strains is likely to be slow. Bacteria become resistant to ribosomal RNA-binding antibiotics through exchange of genetic material encoding RNA-modifying enzymes (typically methyltransferases and phosphotransferases), drug-modifying enzymes, or enzymes that affect drug transport [146, 147]. Therefore, if the structure of the nucleic acid-binding drug is novel, the emergence of resistance is likely to be slower than for protein targets (barring any novel efflux pump mechanisms). Antisense/antigene therapy can offer a viable alternative in tackling such resistance mechanisms.

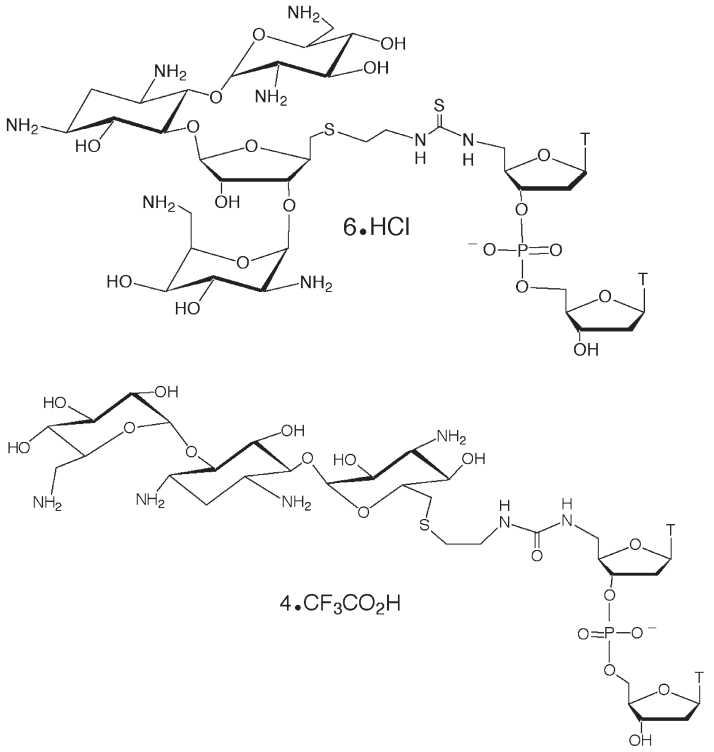
Recent findings that aminoglycosides can stabilize DNA/RNA triplexes [114, 115], hybrid duplexes [113], and that neomycin can even induce hybrid triplex formation [113] suggested that aminoglycoside–DNA conjugates could be effective models for targeting nucleic acids sequence specifically (via a hybrid duplex or triplex formation). Conjugation of an aminoglycoside to an ODN can assist in the following processes:

1. Delivery of aminoglycoside to a specific DNA/RNA site
2. Increasing the stabilization inferred by these hybrid duplex/triplex stabilizing agents
3. The unique structure of aminoglycosides can aid in cellular permeability/site-specific delivery of the ODN

To investigate the advantage of nucleic acid-based specificity coupled with aminoglycoside charge/shape complementarity, we have described a general strategy for the synthesis of covalently attaching aminoglycosides (neomycin) to nucleic acid analogs (DNA/PNA). Recently [148–150], ultrarapid functional genomics technologies have helped identify approximately 4,000 essential gene drug targets in 11 clinically relevant bacterial and fungal pathogens. In contrast, most antimicrobials prescribed today inhibit only a small fraction of this number of targets within bacterial and fungal pathogens. A comprehensive approach to identifying such essential drug targets in multiple pathogens can be combined with a complementary approach of developing antimicrobial agents that are sequence-specific to previously known, as well as rapidly identified, new RNA targets. Interestingly, a shotgun antisense technology was used as the key tool to identify these 4,000 essential genes, suggesting that oligos binding to these RNA targets will be able to selectively shut down protein synthesis. Additionally, the finding that aminoglycosides can stabilize DNA/RNA triplexes and DNA•RNA duplexes [113, 114, 116] suggests that neomycin–ODN conjugates could also be effective models for targeting nucleic acids sequence-specifically via triplex or hybrid duplex formation.

7.1.1
Synthesis of Aminoglycoside Isothiocyanates/ODN–Aminoglycoside Coniugate

The amino groups on rings I, II, and IV (neomycin) are necessary in recognizing and in stabilizing various nucleic acid forms (aminoglycosides without any of these amines do not stabilize DNA tniplexes as efficiently) [137]. The conjugates based on aminoglycosides must then retain these amines. The 5″-OH on ring III (neomycin) was chosen to provide the linkage to the nucleic acids (for ring numbering, please see Scheme 1). We recently reported the synthesis of neomycin isothiocyanate as a stable reagent that can be coupled to a variety of amines [137]. Scheme 4 shows the synthesis of neomycin isothiocyanate, starting from neomycin amine. The use of this isothiocyanate in the synthesis of a DNA 5′–aminothymidine dimer conjugated to neomycin and kanamycin also has been recently reported (Scheme 5) [137].

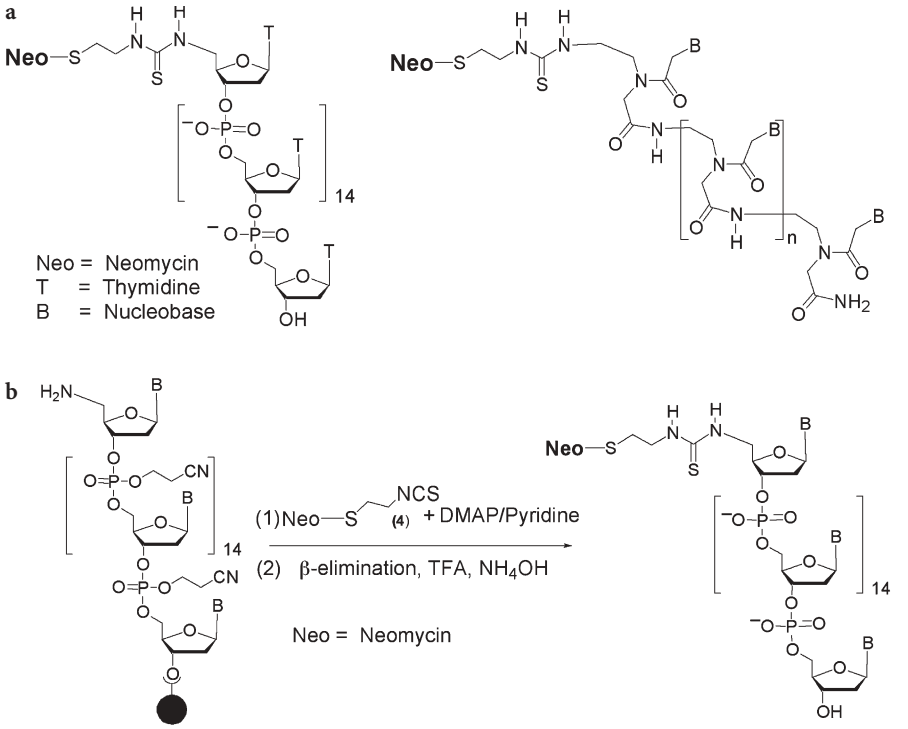


Scheme 5 Structures of neomycin–DNA and kanamycin–DNA dimers

7.1.2 Synthesis of Oligomeric Neomycin–ODN Coniugates

7.1.2.1 Neomycin–DNA/PNA Conjugate

The structures of generic neomycin DNA and PNA conjugates is shown in Scheme 6a and has been recently reported [138]. The synthesis of neomycin conjugated to 5′-end of a oligonucleotide $dT_{(16)}$ is shown in Scheme 6b. Neomycin is linked to the DNA via a thiourea linkage. Neomycin isothiocyanate (Scheme 6b) has been coupled to a 5′-amino-5′-deoxy ODN, which is easily prepared by incorporation of 5′-amino-5′ deoxythymidine (or cytidine) in a growing ODN chain. The synthesis of neomycin linked to a 16-mer DNA $dT_{(16)}$ has been reported [138]. The reactive amine at the 5′-end of $dT_{(16)}$ (Scheme 6b) was treated with a pyridine solution containing neomycin isothiocyanate and 4-dimethylaminopyridine (DMAP) for 12 h at room temperature, washed with trifluoroacetic acid (TFA), and deprotected from solid support with NH_4OH [138]. Having established that these conjugates can be synthesized on solid phase using conventional DNA synthesis, attention can now be devoted to syn-



Scheme 6 **a** Structure of a generic aminoglycoside–DNA/PNA conjugate. **b** Synthesis of neomycin–DNA conjugate on the solid phase

thesizing ODNs for targeting anticancer and antimicrobial DNA sequences of interest [138].

Although considerable advances have been made in antisense technology over the last few decades, there are still some issues that warrant active investigation [151]. These include:

1. Increasing the binding affinities (kinetic/thermodynamic) of ODNs to their target duplexes and single strands
2. Improving the delivery and uptake of oligonucleotides into any cell or tissue of interest
3. Enhancing the stabilities of the oligonucleotides inside the cells

Our preliminary work shows that aminoglycosides can considerably enhance the binding affinities of the ODNs to their duplex DNA target as well as to the single strand RNA targets. Further investigations of the molecular basis of such stabilizations and then using it to synthesize aminoglycoside–DNA/PNA conjugates with improved stability is being carried out in our laboratories. The approach could open up doors for developing sequence-specific anticancer and antimicrobial drugs.

8 Summary

Electrostatic complementarity has been successfully used to explain the structural basis of RNA binding to their aminoglycoside substrates [33–37]. Perhaps the best complementarity for aminoglycosides with a natural target is observed with eubacterial ribosomal 16S A-site. The structural basis of A-form specificity may be related to the closeness of the two negatively charged sugar-phosphate backbones along the major groove in A-DNA, which can be effectively neutralized by the multivalent positively charged amine functions of aniinoglycosides. Groove recognition of triplexes and tetraplexes has been an elusive feat, where such charged polyamine binding factors may be the key to opening this Pandora's box. While these findings do not question that aminoglycoside's mechanism of drug action involves binding to rRNA, they reevaluate, as a matter of biochemical principles, the common belief that aminoglycoside specificity is simply for RNAs, and subsequently unveil new targets for aminoglycoside-based drug development.

Acknowledgements I am greatly indebted to my students, postdoctoral researchers, and collaborators for their hard work and patience. Funding for this work was made available by Clemson University and the NSF (CHE/MCB-0134792).

References

1. Chow CS, Bogdan FM (1997) Chem Rev 97:1489
2. Kotra LP, Haddad J, Mobashery S (2000) Antimicrob Agents Chemother 44:3249
3. Santucci R, Krieger J (2000) J Urol 163:1076
4. Forge A, Schacht J (2000) Audio Neurootol 5:3
5. Gerding D (2000) Infect Control Hosp Epidemiol 21: S12
6. Kaul M, Barbieri CM, Kerrigan JE, Pilch DS (2003) J Mol Biol 326:1373
7. Moazed D, Noller HF (1987) Nature 327:389
8. Purohit P, Stern 5 (1994) Nature 370:659
9. Recht MI, Fourmy D, Blanchard SC, Dablquist KD, Puglisi JD (1996) J Mol Biol 262:421
10. Miyaguchi H, Narita H, Sakamoto K, Yokoyama 5 (1996) Nucleic Acids Res 24:3700
11. Waksman SA (1945) Science 102:40
12. Waksman SA, Lechevalier HA (1949) J Antibiot 109:305
13. Davies J (1965) Antimicrob Agents Chemother (1961–70): 1001
14. Davies J, Davis BD (1968) J Biol Chem 243:3312
15. Tok JBH, Cho J, Rando RR (1999) Biochemistry 38:199
16. Zapp ML, Stem 5, Green MR (1993) Cell 74:969
17. Hermann T, Westhof E (1998) Biopolymers 48:155
18. Mei H-Y, Galan AA, Halim NS, Mack DP, Moreland DW. Sanders KB, Hoa N. Truong, Czarnik AW (1995) Bioorg Med Chem Lett 5:2755
19. von Ahsen U, Noller HF (1993) Science 260:1500
20. Mikkelsen NE, Brannvall M, Virtanen A, Kirsebom LA (1999) Proc Natl Acad Sci USA 96:6155

21. Walter F, Murchie Al, Thomson JB, Lilley DM (1998) Biochemistry 37:14195
22. Eamshaw DJ, Gait MJ (1998) Nucleic Acids Res 26:5551
23. Tor Y, Hermann T, Westhof E (1998) Chem Biol 5: R277
24. Clouet-d'Orval B, Stage TK, Uhlenbeck OC (1995) Biochemistry 34:11186
25. Stage TK, Hertel KJ, Uhlenbeck OC (1995) RNA 1:95
26. Chia JS, Wu HL, Wang HW, Chen DS, Chen PJ (1997) J Biomed Sci 4:208
27. Rogers J, Chang AH, von Ahsen U, Schroeder R, Davies J (1996) J Mol Biol 259:916
28. Mei H-Y, Mack DP, Galan AA, Halim NS, Heldsinger A, Loo JA, Moreland DW, Sannes-Lowery KA, Sharmeen L, Truong HN, Czarnik AW (1997) Bioorg Med Chem 5:1173
29. van Leeuwen F, de Kort M, van der Marel GA, van Boom JH, Borst P (1998) Anal Biochem 258:223
30. Gao Y-g, Robinson H, Wijsman ER, Marel GAvd, Boom JHv, Wang AH-J (1997) J Amer Chem Soc 119:1496
31. Ren J, Qu X, Dattagupta N, Chaires J (2001)
32. Tor Y, Hermann T, Westhof E (1998) Chem Biol 5: R277
33. Hermann T (2000) Angew Chem Int Ed Engl 39:1890
34. Hermann T, Westhof E (1998) Biopolymers 48:155
35. Hermann T, Westhof E (1998) J Mol Biol 276:903
36. Hermann T, Westhof E (1998) Curr Opin Biotechnol 9:66
37. Hermann T, Westhof E (1999) J Med Chem 42:1250
38. Henry CM (2000) Chem Eng News 78:41
39. Greisman HA, Pabo CO (1997) Science 275:657–661
40. Wolfe SA, Nekludova L, Pabo CO (2000) Annu Rev Biophys Biomol Struct 29:183
41. Choo Y, Klug A (1997) Cur Opin Struct Biol 7:117–125
42. Nagai K (1996) Curr Opin Struct Biol 6:53
43. Ellenberger T (1994) Curr Opin Struct Biol 4:12
44. Dervan PB (1986) Science 232:464
45. Dervan PB, Burli RW (1999) Cur Opin Chem Biol 3:688
46. Gottesfeld JM, Turner JM, Dervan PB (2000) Gene Expr 9:77
47. Geierstanger BH, Wemmer DE (1995) Annu Rev Biophys Biomol Stnuct 24:463
48. Wemmer DE (2000) Annu Rev Biophys Biomol Struct 29:439
49. Kopka ML, Yoon C, Goodsell D, Pjura P, Dickerson RE (1985) Proc Natl Acad Sci USA 82:1376–1380
50. Felsenfeld U, Rich A (1957) Biochim Biophys Acta 26:457
51. Moser HE, Dervan PB (1987) Science 238:645
52. Praseuth D, Guieysse AL, Helene C (1999) Biochim Biophys Acta 1489:181
53. Francois J-C, Saison-Behmoaras T, Barbier C, Chassignol M, Thuong NT, Helene C (1989) Proc Natl Acad Sci USA 86:9702
54. Strobel SA, Moser HE, Dervan PB (1988) J Am Chem Soc 110:7927
55. Sun J-S, Francois J-C, Montenay-Garestier T, Saison-Behmoaras T, Roig V, Thuong NT, Helene C (1989) Proc Natl Acad Sci USA 86:9198
56. Maher III U, Wold B. Dervan PB (1989) Science 245:725
57. Hanvey JC, Shimizu M, Wells RD (1989) Nucleic Acids Res 18:157
58. Frank-Kamenetskii MD, Mirkin SM (1995) Annu Rev Biochem 64:65
59. Ganesh KN, Kumar VA, Barawkar DA (1996) In: Hamilton AD (ed) Supramolecular control of structure and reactivity. Wiley, p 263
60. Kool ET (1997) New J Chem 21:33
61. Radhakrishnan I, Patel DJ (1994) Biochemistry 33:11405
62. Shafer RH (1998) Prog Nucleic Acid Res Mol Biol 59:55
63. Thuong NT, Helene C (1993) Angew Chem Int Ed Engl 32:666

64. Kool ET (1998) Acc Chem Res 31:502
65. Wells RD, Collier DA, Hanvey JC, Shimizu M, Wohlrab F (1988) FASEB J 2:2939
66. Htun H, Dahlberg JE (1989) Science 243:1571
67. Htun H, Dahlberg JE (1988) Science 241:1791
68. Grigoriev M, Praseuth D, Robin P, Hemar A, Saison-Behmoaras T, Dautry-Varsat A, Thuong NT, Helene C, Harel-Bellan A (1992) J Biol Chem 267:3389
69. Grigoriev M, Praseuth D, Guieysse AL, Robin P, Thuong NT, Helene C, Harel-Bellan A (1993) Proc Natl Acad Sci USA 90:3501
70. Craig ME, Crother DM, Doty P (1971) J Mol Biol 62:383
71. Rougee M, Faucon B, Mergny JL, Barcelo F, Giovannageli C, Garestier T, Helene C (1992) Biochemistry 31:9269
72. Escude C, Sun JS, Nguyen CH, Bisagni E, Garestier T, Helene C (1996) Biochemistry 35:5735
73. Escude C, Nguyen CH, Mergny J-L, Sun J-S, Bisagni E, Garestier I, Helene C (1995) J Amer Chem Soc 117:10212
74. Escude C, Mohammadi S, Sun JS, Nguyen CH, Bisagni E, Liquier J, Taillandier E, Garestier T, Helene C (1996) Chem Biol 3:57
75. Escude C, Nguyen CH, Kukreti 5, Janin Y, Sun J-S, Bisagni E, Garestier T, Helene C (1998) Proc Natl Acad Sci USA 95:3591
76. Kim SK, Sun J-S, Garestier T, Helene C, Nguyen CH, Bisagni E, Rodger A, Norden B (1997) Biopolymers 42:101
77. Nguyen CH, Marchand C, Delage 5, Sun J-S, Garestier T, Helene C, Bisagni E (1998) J Am Chem Soc 120:2501
78. Tarui M, Doi M, Ishida T, lnoue M, Nakaike 5, Kitamura K (1994) Biochem J 304 (1): 271
79. Wilson WD, Mizan S, Tanious FA, Yao S, Zon U (1994) J Mol Recognit 7:89
80. Cassidy SA, Strekowski U, Wilson WD, Fox KR (1994) Biochemistry 33:15338
81. Wilson WD, Tanious FA, Mizan 5, Yao 5, Kiselyov AS, Zon U, Strekowski U (1993) Biochemistry 32:10614
82. Mergny JL, Duval-Valentin U, Nguyen CH, Perrouault U, Faucon B, Rougee M, Montenay-Garestier T, Bisagni E, Helene C (1992) Science 256:1681
83. Choi S-D, Kim M-S, Kim SK, Lincoln P, Tuite E, Norden B (1997) Biochemistry 36:214
84. Scaria PV, Shafer RH (1991) J Biol Chem 266:5417
85. Mergny J-U, Collier D, Rougee M, Montenay-Garestier T, Helene C (1991) Nucleic Acids Res 19:1521
86. Latimer LIP, Payton N, Forsyth U, Lee JS (1995) Biochem Cell Biol 73:11
87. Lee JS, Uatimer UP. Hampel KJ (1993) Biochemistry 32:5591
88. Denny WA (1989) Anticancer Drug Des 4:241
89. Pilch DS, Waring MJ, Sun J-S, Rougee M, Nguyen C-H, Bisagni E, Garestier T, Helene C (1993) J Mol Biol 232:926
90. Pilch DS, Shafer RH (1993) J Am Chem Soc 115:2565
91. Ren J, Chaires JB (2000) J Am Chem Soc 122:424
92. Strekowski U, Gulevich Y, Baranowski TC, Parker AN, Kiselyov AS, Lin S-Y, Tanious FA, Wilson WD (1996) J Med Chem 39:3980
93. Fox KR, Thurston DE, Jenkins TC, Varvaresou A, Tsotinis A, Siatra-Papastaikoudi T (1996) Biochem Biophys Res Cominun 224:717
94. Fox KR, Po]ucci P, Jenkins TC, Neidle 5 (1995) Proc Natl Acad Sci USA 92:7887
95. Haq I, Ladbury JE, Chowdhry BZ, Jenkins TC (1996) J Amer Chem Soc 118:10693
96. Kan Y, Armitage B, Schuster GB (1997) Biochemistry 36:1461
97. Durand M, Maurizot JC (1996) Biochemistry 35:9133
98. Durand M, Thuong NT, Maurizot JC (1994) J Biomol Struct Dyn 11:1191

99. Pilch DS, Kirolos MA, Breslauer KJ (1995) Biochemistry 34:16107
100. Pilch DS. Breslauer KJ (1994) Proc Natl Acad Sci USA 91:9332
101. Durand M, Thuong NT, Maurizot JC (1992) J Biol Chem 267:24394
102. Antony T, Thomas T, Shirahata A, Sigal LH, Thomas Ti (1999) Antisense Nucleic Acid Drug Dev 9:221
103. Basu HS, Marton U (1987) Biochem J 244:243
104. Musso M, Thomas T, Shirahata A, Sigal LH, Dyke MWV, Thomas TJ (1997) Biochemistry 36:1441
105. Thomas TJ, Kulkarni GD, Greenfield NJ, Shirahata A, Thomas T (1996) Biochem J 319:591
106. Thomas I, Thomas TJ (1993) Biochemistry 32:14068
107. Pallan PS, Ganesh KN (1996) Biochem Biophys Res Commun 222:416
108. Nagamani D, Ganesh KN (2001) Org Lett 3:103
109. Rajeev KG, Sanjayan GJ, Ganesh KN (1997) J Org Chem 62:5169
110. Potaman VN, Sinden RR (1995) Biochemistry 34:14885
111. Mamyama A, Katoh M, Ishihara T, Akaike T (1997) Bioconjug Chem 8:3
112. Dervan PB (2001) Bioorg Med Chem 9:2215
113. Arya DP, Coffee RL Jr, Charles I (2001) J Am Chem Soc 123:11093
114. Arya DP, Coffee RL Jr, Willis B, Abramovitch Al (2001) J Am Chem Soc 123:5385
115. Arya DP, Coffee RL Jr (2000) Bioorganic Med Chem Lett 10:1897
116. Arya DP, Micovic L, Charles I, Coffee RL Jr, Willis B, Xue U (2003) J Am Chem Soc 125:3733
117. Kohlstaedt LI, Wang J, Friedman J, Rice P, Steitz T (1992) Science 256:1783
118. Stein CA, Cohen JS (1988) Cancer Res 48:2659
119. Morgan AR, Wells RD (1968) J Mol Biol 37:63
120. Murray NL, Morgan AR (1973) Can J Biochem 51:436
121. Han H, Dervan PB (1993) Proc Natl Acad Sci USA 90:3806
122. Wang 5, Kool ET (1995) Nucleic Acids Res 23:1157
123. Riley M, Maling B, Chamberlin MJ (1966) J Mol Biol 20:359
124. Barbieri CM, Li T-K, Guo 5, Wang U, Shallop AJ, Pan W, Yang U, Gaffney BL, Jones RA, Pilch DS (2003) J Am Chem Soc 125:6469
125. Ren J, Chaires JB (2001) Methods Enzymol, vol 340. Academic, New York, p 99
126. Kirk SR, Luedtke NW, Tor Y (2000) J Am Chem Soc 122:980
127. Arya DP, Xue U. Willis B (2003) J Am Chem Soc 125:10148
128. Sanger W (1983) In: Cantor CR (ed) Principles of nucleic acid structure. Springer, Berlin Heidelberg New York, p 242
129. Stefl R, Trantirek U, Vorlickova M, Koca J, Sklenar V, Kypr J (2001) J Mol Biol 307:513
130. Robinson H, Wang AHJ (1996) Nucleic Acids Res 24:676
131. KyprJ, Fialova M, Chladkova J, Tumova M, Vorlickova M (2001) Eurbiophysi 30:555
132. Franklin RE, Goslin RU (1953) Acta Crystallography 6:673
133. Lvanov VI, Minchenkova LE (1995) Molekulyamaya Biologiya (Moscow) 28:780
134. Calladine CR, Drew HR (1984) J Mol Biol 178:773
135. Marky NU, Olson WK (1994) Biopolymers 34:121
136. Jeffares DC, Poole AM, Penny D (1998) J Mol Evol 46:18
137. Charles I, Xue U, Arya DP (2002) Bioorg Med Chem Lett 12:1259
138. Charles I, Arya DP (2004) Bioorg Med Chem Lett (in press)
139. Lu X-J, Shakked Z, Olson WK (2000) J Mol Biol 300:819
140. Robinson H, Gao Y, Sanishvili R, Joachimiak A, Wang AHJ (2000) Nucleic Acids Res 28:1760
141. White CM, Satz AL, Bruice TC, Beerman TA (2001) Proc Natl Acad Sci USA 98:10590
142. Haq I, Ladbury JE, Chowdhry BZ, Jenkins TC, Chaires JB (1997) J Mol Biol 271:244

143. Ecker DJ, Griffey RH (1999) Drug Discovery Today 4:420
144. Gallego JVU (2001) Acc Chem Res 34:836
145. Grillone LR (2001) In: Crooke ST (ed) Antisense drug technology: principles, strategies and applications. Dekker, New York, p 725
146. Neu HC (1992) Science 257:1064
147. Azucena E, Mobashery S (2001) Drug Resist Update 4:106
148. Elitra (2000) Bioworld Week 8:2
149. Elitra (1999) Bioworld Week 7:4
150. Haselbeck R, Wall D, Jiang B. Ketela T, Zyskind J, Bussey H, Foulkes JG, Roemer I (2002) Cur Pharm Des 8:1155
151. Vazquez ICM, Glazer PM (2002) Q Rev Biophys 35:89

Top Curr Chem (2005) 253: 179–204
DOI 10.1007/b100447

Ribosomal RNA Recognition by Aminoglycoside Antibiotics

Daniel S. Pilch (✉) · Malvika Kaul · Christopher M. Barbieri

Department of Pharmacology, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, NJ 08854–5635, USA
pilchds@umdnj.edu

Abstract 2-Deoxystreptamine (2-DOS) aminoglycosides are a family of structurally related broad-spectrum antibiotics that are used widely in the clinic. Their antibiotic activities are ascribed to their abilities to bind a highly conserved sequence (termed the A site) in the 16S rRNA of the 30S ribosomal subunit and interfere with protein synthesis. The aminoglycosides represent a paradigm for both drug–RNA and drug–ribosome interactions, and information gleaned from their study has relevance with regard to other RNA- and ribosome-directed drugs of acute clinical importance. This contribution provides an integrated overview of structural and thermodynamic studies of the rRNA binding of aminoglycosides. The results of these studies have enhanced our understanding of the molecular forces that govern aminoglycoside recognition of the rRNA A site, and underlie the mechanism and specificity of action of these drugs. Such knowledge provides the type of predictive capabilities that are essential for the devlopment of a rational basis for future drug design strategies.

Keywords Thermodynamics of rRNA binding · Binding-linked drug protonation · Binding-induced conformational change · Isothermal titration calorimetry · Osmotic stress · Polyelectrolyte contributions to binding

Abbreviations

rRNA	Ribosomal RNA
2AP	2-Aminopurine
TAPS	*N*-Tris[hydroxymethyl]methyl-3-aminopropanesulfonic acid
EPPS	*N*-[2-Hydroxymethyl]piperazine-N′-[3-propanesulfonic] acid
ITC	Isothermal titration calorimetry
2-DOS	2-Deoxystreptamine
Osm	Osmolality
DMS	Dimethylsulfate

1 Introduction

The aminoglycosides are a structurally related group of broad-spectrum bactericidal antibiotics that are used widely in the treatment of infections caused by aerobic gram-negative bacteria, particularly in cases of bacteremia and sepsis [1]. They have predictable pharmacokinetic properties and often act in synergy with other antibiotics (e.g., β-lactams and vancomycin). In addition, they exhibit a significant postantibiotic effect in which antibacterial activity persists beyond the time that measurable drug is present. These characteristics all serve to bolster the clinical value of aminoglycosides [1, 2].

Aminoglycosides derive their name from their structures, which consist of an aminocyclitol that is linked glycosidically to various amino sugars. In streptomycin, the aminocyclitol functionality is streptidine, while being 2-deoxystreptamine (2-DOS) in other aminoglycosides. There are two major classes of 2-DOS aminoglycosides, the 4,5-disubstituted 2-DOS class, which includes neomycin, paromomycin, and ribostamycin, and the 4,6-disubstituted 2-DOS class, which includes tobramycin, kanamycins A and B, amikacin, geneticin (G418), and the gentamicins (see Fig. 1 for a representative structure from each

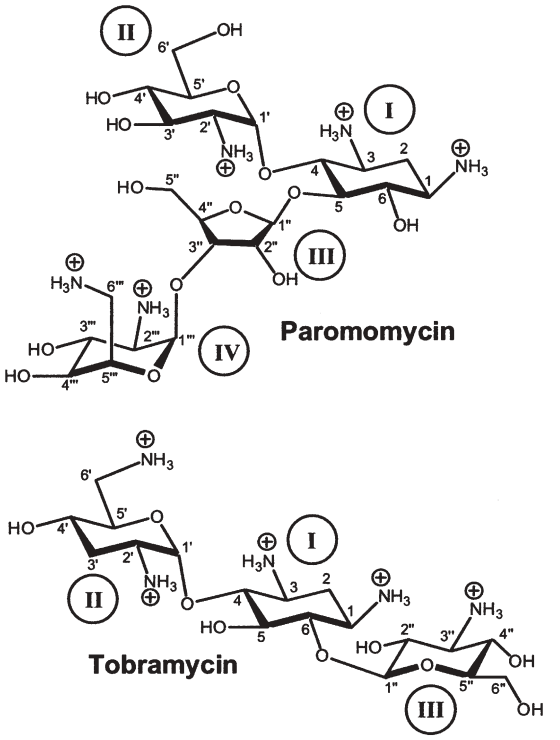


Fig. 1 Structure of the 4,5-disubstituted 2-deoxystreptamine (2-DOS) aminoglycoside paromomycin (*top*) and the 4,6-disubstituted 2-DOS aminoglycoside tobramycin (*bottom*), with the atomic and ring numbering systems being denoted in Arabic and Roman numerals, respectively. The 2-DOS functionality is ring I in each structure

2-DOS class). Reported pK_a values for the amino groups of various aminoglycosides range from approximately 5.7 to 10.1 [3–7]. Thus, aminoglycosides exist as oligocations at physiologically relevant values of pH.

In prokaryotes, the intracellular target of aminoglycosides is the 30S ribosomal subunit [8, 9]. For 2-DOS aminoglycosides, the specific target within the 30S ribosomal subunit is a highly conserved sequence in the decoding region of the 16S rRNA. This conserved RNA sequence forms the site, termed the A site, at which the interaction between the anticodon of the aminoacyl-tRNA and the mRNA codon occurs [10]. The bactericidal activities of the 2-DOS aminoglycosides are attributed to their abilities to interfere with this crucial step in the translation process [9]. The deleterious impact of aminoglycosides on protein synthesis includes both a reduction in translational fidelity (i.e., mistranslation) as well as inhibition of the translocation step, which, in turn, results in the formation of truncated polypeptides [9, 11, 12]. The abilities of the aminoglycosides to recognize a specific subdomain of a large RNA molecule make these compounds archetypical models for RNA-targeting drugs.

To date, eight structures of aminoglycoside–rRNA complexes are on deposit in the Protein Data Bank [13]. The drug target in six of these structures is an rRNA model oligonucleotide, with the target in the other two structures being a 30S ribosomal subunit. These structures, which were derived either crystallographically or by NMR techniques, have revealed key insights into the structural biology of aminoglycoside–rRNA recognition. While invaluable, high-resolution structural information alone cannot provide a complete understanding of the molecular forces that govern aminoglycoside–rRNA recognition. Such an understanding requires the integration of structural information with thermodynamic, dynamic, and kinetic data.

The purpose of this contribution is to provide an integrated overview of both structural and thermodynamic characterizations of aminoglycoside–rRNA interactions. Particular emphasis will be placed on paromomycin, since its physical, chemical, and rRNA binding properties are the best characterized among the aminoglycosides.

2 Structural Database of Aminoglycoside–rRNA Complexes

The central portion of the 16S rRNA A site contains an asymmetric internal loop formed by nucleotides A1408, A1492, and A1493 (see Fig. 2). Footprinting studies have indicated that this region of 16S rRNA is essentially free of contacts with ribosomal proteins [14], an observation later confirmed by the recently reported crystal structure of the 30S ribosomal subunit of *Thermus thermophilus* [15, 16]. In the aggregate, these results suggested that appropri-

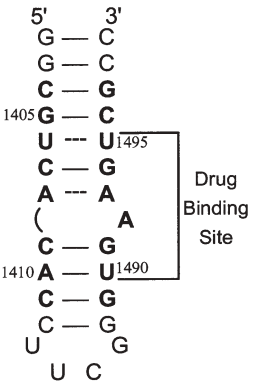


Fig. 2 Secondary structure of a 27-mer 16S rRNA A site model oligonucleotide (*Ec*WT), as determined by NMR [18, 19, 53]. Watson–Crick base pairs are denoted by *solid lines*, while noncanonical base pairs are denoted by *dashed lines*. Bases present in *E. coli* 16S rRNA are depicted in *bold face*, and are numbered as they are in 16S rRNA. The aminoglycoside binding site, as revealed by NMR and footprinting studies [17, 18, 53], is as indicated

ately designed RNA oligonucleotides might be able to recapitulate the local structure and/or conformation that exists in the decoding region A site within the ribosome. The Puglisi group took such a reductionist approach by designing a 27-mer hairpin oligonucleotide (whose sequence and secondary structure are shown in Fig. 2) intended to mimic the decoding region A site of *E. coli* 16S rRNA [17, 18]. Hereafter, this oligonucleotide will be referred to as *Ec*WT. Significantly, they demonstrated that the pattern of aminoglycoside-induced protection of the RNA bases from methylation by DMS was virtually identical in the oligonucleotide construct as it was in the 30S ribosomal subunit [17, 18].

2.1 NMR-Derived Structures of a 16S rRNA A Site Model Oligonucleotide in Complex with Paromomycin and Gentamicin C1a – The First High-Resolution Structures of Aminoglycosides in Complex with a rRNA Sequence

Having validated the reductionist approach as described above, Puglisi and coworkers used NMR techniques to determine the solution structure of *Ec*WT [19], as well as its complex with paromomycin [18] (see Fig. 3 for representations of the two structures). In addition to the solution structure of the complex between *Ec*WT and paromomycin, the Puglisi group also determined the corresponding solution structure of the gentamicin C1a complex [20]. These collective NMR studies revealed the first structural insights into some of the key molecular interactions that underlie the specificities of representative members of both the 4,5- and 4,6-disubstituted 2-DOS classes of aminoglycosides for the 16S rRNA A site. Some of the key structural features to emerge from these studies are summarized as follows:

1. Both drugs bind in the major groove of the RNA at the site of the internal loop formed by A1408, A1492, and A1493 (Fig. 2)
2. Both drugs induce a similar conformational change in the RNA, which includes a displacement of A1492 and A1493 towards the minor groove by approximately 3 and 4 Å, respectively
3. The drugs bind in a pocket created, in part, by the displacement of A1492 and A1493 where they form an array of hydrogen-bonding, van der Waals, and electrostatic contacts with the host RNA.

2.2 Crystal Structure of the 30S Ribosomal Subunit of *Thermus thermophilus* in Complex with Paromomycin – Comparison with the NMR Structure of the Paromomycin–Oligonucleotide Complex

Subsequent to the NMR studies noted above, Ramakrishnan and coworkers reported a crystal structure at 3.0 Å resolution of the 30S ribosomal subunit of *Thermus thermophilus* in complex with both paromomycin and streptomycin [15]. Significantly, the crystal structure of the paromomycin complex (see

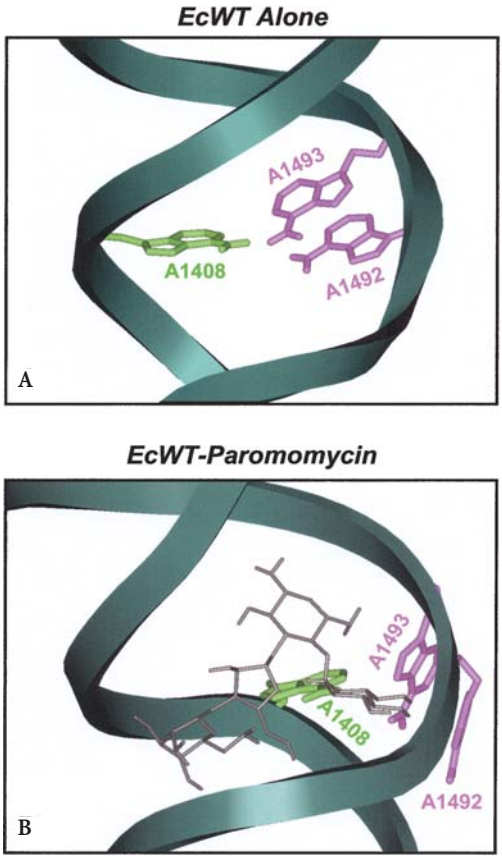


Fig. 3 **A** Solution structure of *Ec*WT in the absence of drug (PDB code 1A3M) [19]. The RNA backbone is depicted in ribbon format. The adenines at positions 1492 and 1493 are depicted in *magenta*, while the adenine at position 1408 is depicted in *green*. **B** Solution structure of *Ec*WT in complex with paromomycin (PDB code 1PBR) [18], with the drug being depicted in *gray*

Fig. 4) revealed many of the same basic features described above for the complex between paromomycin and the A site model oligonucleotide, highlighting the usefulness and validity of using oligonucleotide sequences to model the rRNA A site. While the two complexes exhibited similar global features, they differed in the following three ways (compare Fig. 3B and Fig. 4):

1. The extents to which A1492 and A1493 were displaced toward the minor groove (which were greater in the crystal structure than the NMR structure)
2. The orientation of ring IV of paromomycin
3. The positions of U1406 and C1407 and their associated backbones.

These differences were attributed to the conformational dynamics of the drug-RNA complex in solution, with the NMR structure (refined against residual

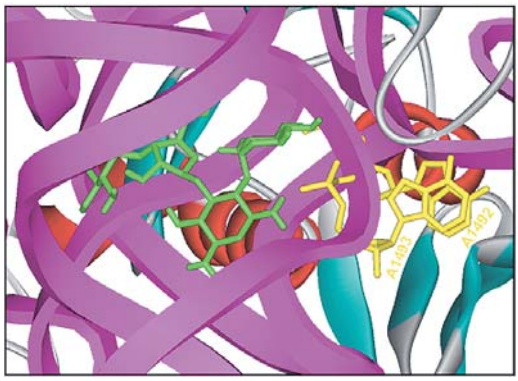


Fig. 4 Crystal structure of the 30S ribosomal subunit of *Thermus thermophilus* in complex with paromomycin (PDB code 1FJG) [15]. The rRNA is depicted in ribbon format and colored magenta. The adenines at positions 1492 and 1493 are depicted in *yellow*, and the paromomycin molecule is depicted in *green*. The ribosomal proteins are depicted according to their secondary structural motifs

dipolar coupling restraints) yielding a range of low energy conformations and the crystal structure isolating a single conformation that is stabilized in part by packing forces [21].

2.3 Crystal Structures of 16S rRNA A Site Model Oligonucleotides in Complex with Paromomycin, Tobramycin, and Geneticin

Westhof and coworkers have recently reported the crystal structures of 16S rRNA A site model oligonucleotides in complex with paromomycin, tobramycin, and geneticin [22–24]. These structures emulate the features revealed by the crystal structure of the paromomycin complex with the 30S ribosomal subunit, but also reveal additional drug–RNA contacts by virtue of their superior resolution (≈2.4–2.5 Å). In particular, the extents to which A1492 and A1493 are destacked in the crystal structures of the three aminoglycoside–oligonucleotide complexes are similar to those observed in the crystal structure of the paromomycin complex with the 30S ribosomal subunit. Another significant contribution of the crystal structures solved by the Westhof group was the indication that water molecules can play a role in mediating aminoglycoside–rRNA contacts [22, 23].

3 pH Dependence of Paromomycin Binding to the 16S rRNA A Site

One of the hallmarks of aminoglycosides is the pH dependence of their RNA binding properties [7, 25, 26]. In this connection, we have shown that pH modulates the extent to which aminoglycosides thermally stabilize the host RNA

[7, 25, 26]. As an illustrative example, Fig. 5 shows the UV melting curves for *Ec*WT in the absence and presence of paromomycin at a drug to duplex ratio of 1.0. Note that at pH 6.0 and a Na^+ concentration of 60 mM, the presence of paromomycin enhances the thermal stability (T_m) of the host RNA duplex by 9.7°C. Further note that, at a constant Na^+ concentration of 60 mM, the paromomycin-induced change in thermal enhancement (ΔT_m) decreases from 9.7 to 2.3°C with a pH increase from 6.0 to 7.0 (compare panels A and B of Fig. 5). Thus, increasing pH results in a concomitant decrease in ΔT_m.

Reductions in ΔT_m often (but need not necessarily) coincide with corresponding reductions in binding constant (K_a). Isothermal titration calorimetry (ITC) provides a means for investigating whether the observed pH-induced reduction in ΔT_m that is associated with the binding of paromomycin to the 16S

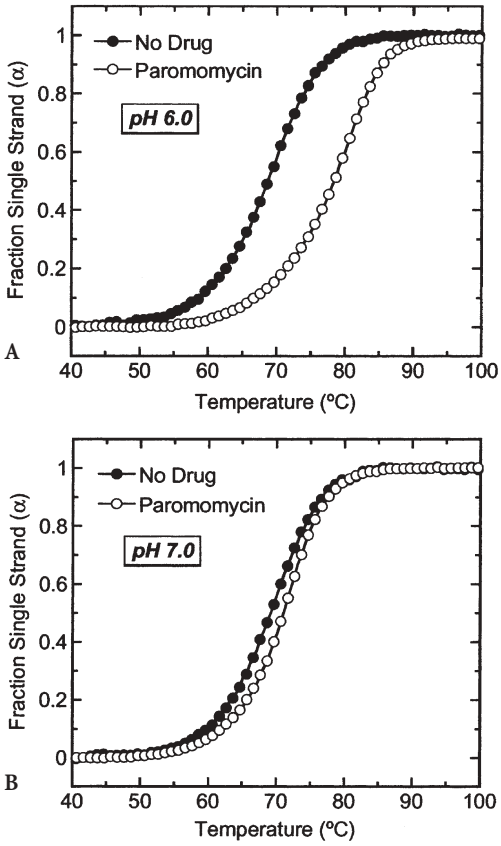


Fig. 5 UV melting profiles for *Ec*WT and its paromomycin complex at a [drug]/[RNA] ratio of 1.0. The Na^+ concentration is 60 mM and the pH values are 6.0 (**A**) and 7.0 (**B**). For clarity of presentation, the melting curves were normalized by subtraction of the upper and lower baselines to yield plots of fraction single strand (α) versus temperature [54]. The UV melting profiles were acquired as described in [26]

rRNA A site correlates with a corresponding reduction in K_a. Figure 6 shows representative ITC profiles resulting from the injection of paromomycin into a solution of *Ec*WT at a constant Na^+ concentration of 150 mM and a pH of either 5.5 (panels A and B) or 6.3 (panels C and D). Each of the heat burst curves in panels A and C of Fig. 6 corresponds to a single drug injection. The areas under the heat burst curves were determined by integration to yield the associated injection heats. These heats were subsequently corrected by subtraction of the corresponding dilution heats derived from the injection of identical amounts of drug into buffer alone. Panels B and D of Fig. 6 show the resulting corrected injection heats plotted as a function of the [drug]/[duplex] ratio.

The binding parameters that emerged from the fits of the ITC profiles in Fig. 6 are summarized in Table 1. Inspection of these binding parameters reveals that paromomycin binds to the host RNA with a 5.2-fold higher affinity at pH 5.5 than it does at pH 6.3: $K_a=(3.7\pm0.7)\times10^7$ and $(7.1\pm0.7)\times10^6$ M^{-1} at pH 5.5 and 6.3, respectively. Thus, increasing pH results not only in a reduction in ΔT_m, but also a reduction in K_a. Further inspection of the data in

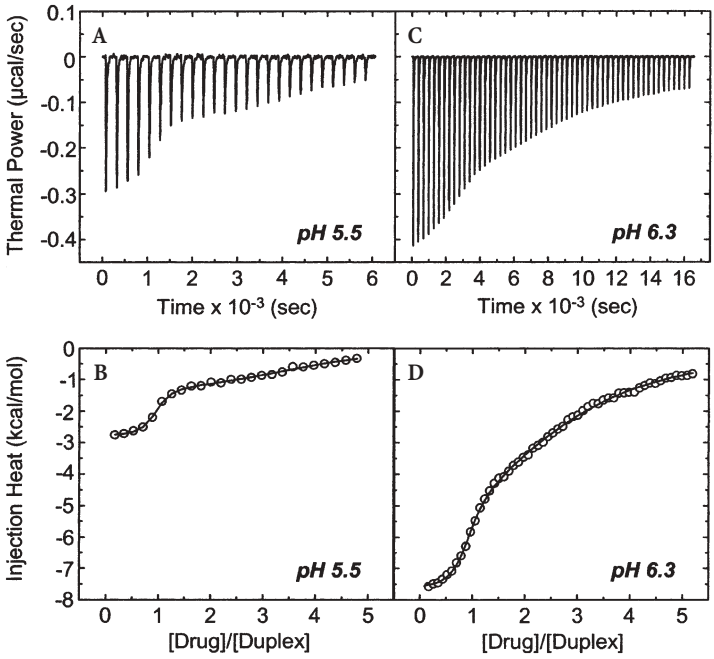


Fig. 6 ITC profiles at 25°C for the titration of paromomycin into a solution of *Ec*WT at pH 5.5 (**A** and **B**) and 6.3 (**C** and **D**). Each heat burst curve in panel *A* is the result of a 10 µL injection of 250 µM drug, with each heat burst curve in panel *B* being the result of a 5 µL injection of 250 µM drug. The RNA concentration was 10 µM in strand. Solutions contained 10 mM cacodylate, 0.1 mM EDTA, and sufficient NaCl to bring the total Na^+ concentration to 150 mM. The data were obtained and analyzed as described in [7]

Table 1 ITC-derived parameters at 25°C for the binding of paromomycin to *Ec*WT at pH 5.5 and 6.3 and Na^+ concentration of 150 mM

pH	K_a (M^{-1})	ΔH_{obs} (kcal mol^{-1})	N
5.5	$(3.7\pm0.7)\times10^7$	−2.9±0.1	0.9±0.1
6.3	$(7.1\pm0.7)\times10^6$	−7.7±0.1	1.0±0.1

Solutions contained 10 mM sodium cacodylate, 0.1 mM EDTA, and sufficient NaCl to bring the total Na^+ concentration to 150 mM. Values for the association constant (K_a), the observed binding enthalpy (ΔH_{obs}), and the binding stoichiometry (N) are derived from fits of the ITC profiles shown in Fig. 6.

Table 1 reveals that the observed binding enthalpy (ΔH_{obs}) is more exothermic (favorable) at pH 6.3 (ΔH_{obs}=−7.7±0.1 kcal mol^{-1}) than at pH 5.5 (ΔH_{obs}= −2.9±0.1 kcal mol^{-1}). In other words, the reduction in binding affinity that accompanies the increase in pH from 5.5 to 6.3 is entropic in origin. As detailed in the sections that follow, the inverse relationship between pH and K_a arises from the linkage of drug protonation to drug–RNA complex formation.

3.1 pH-Induced Effects on Paromomycin Binding to the 16S rRNA A Site Arise from the Linkage of Complex Formation to Drug Protonation

Recall that paromomycin contains five amino groups, which exist in equilibria between their noncharged NH_2 and positively charged NH_3^+ states (see Fig. 1). Increasing pH will tend to shift these equilibria toward the noncharged NH_2 states. The protonation of NH_2 groups, such as the 2-NH_2 group on D-glucosamine and the α-amino groups of amino acids, are known exothermic reactions. Such protonation events, while being enthalpically favorable, are entropically costly. Hence, it is likely that the pH dependence of ΔH_{obs} and K_a noted above reflects binding-induced protonation of one or more drug NH_2 groups. One can probe for a linkage between drug–RNA binding and the release or uptake of protons by conducting ITC experiments over a range of pH values using buffers that differ with respect to their heats of ionization (ΔH_{ion}). If the pH dependence of ΔH_{obs} reflects binding-induced uptake or release of protons, then its value at a given pH should vary with the ΔH_{ion} value of the buffer. Furthermore, the number of protons (Δn) linked to binding at a specific pH, as well as the intrinsic binding enthalpy (ΔH_{int}), a value that differs from ΔH_{obs} in that it excludes enthalpic contributions from ionization of the buffer, can be determined by simultaneous solution of the following two equations [27]:

$$\Delta H_{obs1} = \Delta H_{int} + \Delta H_{ion1}\Delta n \quad (1a)$$

$$\Delta H_{obs2} = \Delta H_{int} + \Delta H_{ion2}\Delta n \quad (1b)$$

In these equations, the numerical subscripts refer to the different buffers. A positive value of Δn is indicative of a net uptake of protons, while a negative value of Δn is indicative of a net release of protons.

Figure 7 shows representative ITC profiles resulting from the injection of paromomycin sulfate into a solution of *Ec*WT in bicine (A and B) and TAPS (C and D) buffers at pH 9.0. A comparison of panels A and B with panels C and D reveals that the magnitude of the exothermic signal is substantially greater in bicine (ΔH_{ion}=+6.47 kcal mol^{-1}) than in TAPS (ΔH_{ion}=+9.92 kcal mol^{-1}). This observation is indicative of binding-linked proton uptake and is therefore consistent with binding-linked drug protonation.

Table 2 summarizes the values of ΔH_{obs} derived from the ITC experiments shown in Fig. 7, as well as from other ITC experiments (not shown) conducted at pH 6.3 and 7.5. In addition, Table 2 also includes the values of ΔH_{int} and Δn, as calculated by solution of Eq. 1a and Eq. 1b. Inspection of these data reveals Δn values of +0.53, +1.47, and +3.25 at pH 6.3, 7.5, and 9.0, respectively. Note that these values of Δn reflect contributions from all the drug amino groups whose protonation is linked to RNA complex formation. In fact, the number of

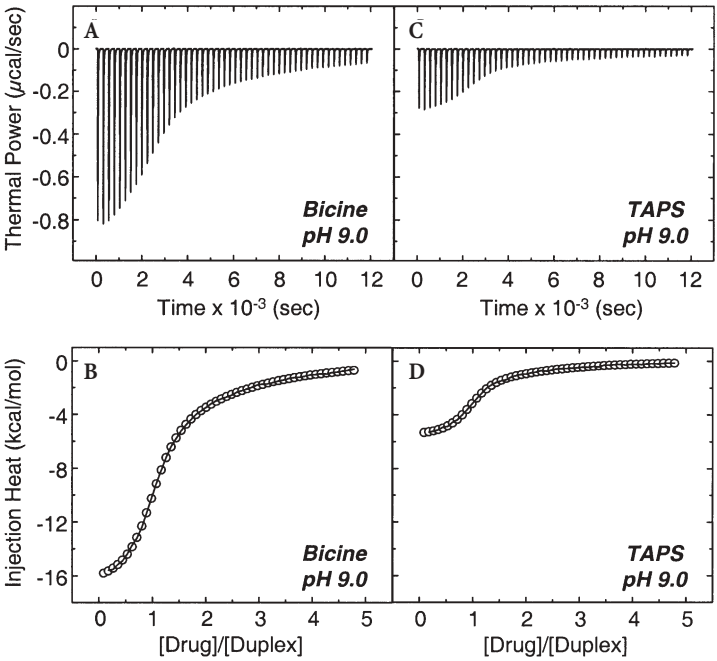


Fig. 7 ITC profiles at 25°C for the titration of paromomycin into a solution of *Ec*WT at pH 9.0 in bicine (**A** and **B**) and TAPS (**C** and **D**) buffer. Each heat burst curve in panels *A* and *C* is the result of a 5 μL injection of 250 μM drug. The RNA concentration was 10 μM in strand, with each experimental solution containing 10 mM buffer, 0.1 mM EDTA, and sufficient NaCl to bring the total Na^+ concentration to 11 mM. The data were obtained and analyzed as described in [7]

Table 2 Number of protons linked to the binding of paromomycin to *Ec*WT at pH 6.3, 7.5, and 9.0

Buffer	pH	ΔH_{ion} (kcal mol^{-1})	ΔH_{obs} (kcal mol^{-1})	ΔH_{int} (kcal mol^{-1})	Δn
Cacodylate	6.3	−0.47	−7.7±0.1	−7.5±0.2	0.53±0.04
MES	6.3	+3.71	−5.5±0.1	−7.5±0.2	0.53±0.04
EPPS	7.5	+5.15	−9.8±0.1	−17.4±0.2	1.47±0.04
TAPS	7.5	+9.92	−2.8±0.1	−17.4±0.2	1.47±0.04
Bicine	9.0	+6.47	−16.8±0.1	−37.8±0.5	3.25±0.06
TAPS	9.0	+9.92	−5.6±0.1	−37.8±0.5	3.25±0.06

The Na^+ concentration was 150 mM at pH 6.3 and 7.5, but 11 mM at pH 9.0. All parameters were determined at 25°C. Heats of ionization (ΔH_{ion}) were obtained from [52], with the remaining data being obtained and analyzed as described in [7].

drug amino groups involved in RNA binding-linked proton uptake reactions at a given pH can exceed the magnitude of the observed Δn at that pH. In order to define the relevant number of drug amino groups that participate in RNA binding-linked protonation reactions, as well as their identities, one must determine the pK_a value of each drug amino group.

3.2 Using Natural Abundance ^{15}N NMR to Determine the pK_a Values of the Amino Groups of Paromomycin in Both Its Free Base and Sulfate Salt Forms

Natural abundance ^{15}N NMR can be used to determine the pK_a values of the paromomycin amino groups by monitoring the pH dependence of the ^{15}N chemical shifts at which the amino groups resonate [3, 4, 7, 28]. Table 3 summarizes the pK_a values so obtained for the free base and sulfate salt forms of paromomycin. Note that the pK_a values of the paromomycin amino groups range from 6.50 to 9.13 in the free base form of the drug. Further note that the

Table 3 ^{15}N NMR-derived pK_a value sat 25°C for the amino groups of paromomycinin its sulfate salt and free base forms

Amino group	$pK_a^{Sulfate}$	$pK_a^{Free\ Base}$
1	8.65±0.05	8.20±0.07
3	7.07±0.01	6.50±0.03
2′	8.33±0.03	8.07±0.04
2‴	8.25±0.01	7.91±0.02
6‴	9.46±0.03	9.13±0.03

The data were obtained and analyzed as described in [7].

pK_a values of the amino groups are 0.26–0.58 pH units lower in the free base than in the sulfate salt form, with the difference in pK_a being the smallest for the 2′-amino group and the largest for the 3-amino group. Thus, the presence of sulfate counterions modulates ^{15}N NMR-determined values of pK_a. Note that the Δn values listed in Table 2 were determined from ITC experiments that used the sulfate salt form of paromomycin. Thus, it is tempting to ascribe these Δn values to the corresponding pK_a values determined for paromomycin sulfate. However, the following illustrative comparison suggests that such is not the case. The pK_a values for paromomycin sulfate predict that 1.09 out of 5.00 amino groups are deprotonated at pH 7.5, with the corresponding pK_a values for the free base form of paromomycin predicting that 1.59 out of 5.00 amino groups are deprotonated. The observed Δn value of 1.47 at pH 7.5 is indicative of *at least* that number of amino groups being deprotonated in the unbound (RNA-free) state of the drug, a deprotonation state that is inconsistent with that predicted by the pK_a values for paromomycin sulfate. Thus, the observed values of Δn for paromomycin binding to the host RNA reflect the pK_a values of the free base and not the sulfate salt form of the drug.

3.3 Identification of the Specific Amino Groups Whose Protonation Is Linked to RNA Complex Formation

A comparison of the Δn values determined at pH 7.5 and 9.0 (Table 2) with the corresponding deprotonation states of paromomycin predicted by the pK_a values for the free base form of the drug at these values of pH (Table 3) allows one to identify the specific drug amino groups whose protonation is linked to RNA complex formation. Recall that the Δn value for paromomycin binding to the host RNA at pH 9.0 is 3.25 (Table 2). The sum of the predicted deprotonated (NH_2) fractions for the 3-, 2′-, 2‴-, and 6‴-amino groups is 3.24, with the sums of the predicted fraction NH_2 values from only two other combinations of amino groups being similar in magnitude (3.21 for the 1-, 3-, 2‴-, and 6‴-amino groups and 3.18 for the 1-, 3-, 2′-, and 6‴-amino groups). Thus, the Δn and pK_a data at pH 9.0 narrow the potential identities of drug amino groups that participate in RNA binding-linked protonation reactions to three different combinations of four amino groups. Two of these three combinations can be ruled out upon comparison of the Δn and pK_a data at pH 7.5. Specifically, the sums of the predicted fraction NH_2 values at pH 7.5 are 1.38 for the 1-, 3-, 2‴-, and 6‴-amino groups, 1.31 for the 1-, 3-, 2′-, and 6‴-amino groups, and 1.43 for the 3-, 2′-, 2‴-, and 6‴-amino groups. Note that the sum of the predicted fraction NH_2 values from only the latter combination of amino groups is consistent with the observed Δn value of 1.47 (Table 2). In the aggregate, the comparisons described above reveal that the binding of paromomycin to the host RNA is coupled to the protonation of the 3-, 2′-, 2‴-, and 6‴-amino groups, but not the 1-amino group. In other words, these amino groups must be in their fully protonated NH_3^+ states when the drugs is bound to the host RNA. Note that this

requirement is not evident in the previously reported NMR-derived and crystal structures of paromomycin in complex with 16S rRNA A site model oligomers or with the 30S ribosomal subunit [15, 18, 22]. Significantly, in each of these structures, all drug amino groups are in their deprotonated NH_2 rather than their protonated NH_3^+ states. Despite this inconsistency, the crystal structure of paromomycin in complex with a 16S rRNA A site model oligomer previously reported by the Westhof group [22] provides insight into the molecular basis for the linkage between protonation of the 3-, 2′-, 2‴-, and 6‴-amino groups and drug–RNA complex formation. Inspection of this structure, which is depicted in Fig. 8, reveals that the 3-, 2′-, 2‴-, and 6‴-amino groups form electrostatic contact with phosphate functionalities on the RNA backbone. By contrast, the 1-amino group, whose protonation is not linked to RNA binding, is directed away from the phosphate backbone into the floor of the major groove.

It is of interest to note that a similar type of analysis to that described above, but using the pK_a values for the sulfate salt instead of the free base form of paromomycin, yields a different set of four amino groups that participate in RNA binding-linked protonation reactions, namely, the 1-, 3-, 2′-, and 2‴-amino groups [7]. This set of amino groups has three groups (the 3-, 2′-, and 2‴-amino groups) in common with the set of amino groups defined above using the pK_a values of the free base form of the drug. The difference between the two sets of amino groups lies in the identity of the fourth group, which is the 6‴-amino group when using the free base pK_a values and the 1-amino group when using the sulfate salt pK_a values.

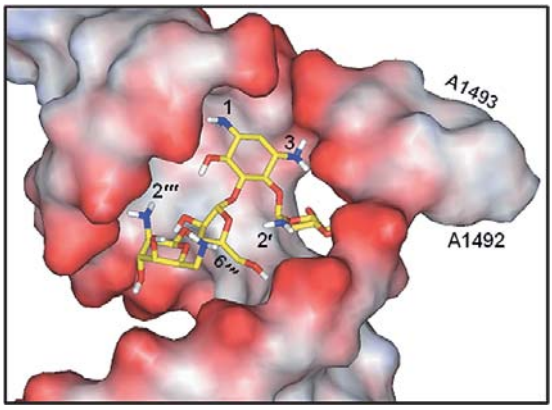


Fig. 8 Crystal structure of a 16S rRNA A site model oligonucleotide in complex with paromomycin (PDB code 1J7T) [22]. The RNA is depicted in its solvent accessible surface (probe radius 1.4 Å), and is colored according to its electrostatic potential (*red*=negative, *white*=neutral, and *blue*=positive). The paromomycin molecule is depicted in stick format, and is colored according to the following scheme: carbon in *yellow*, oxygen in *red*, nitrogen in *blue*, and hydrogen in *white*. The five amino groups of the drug are indicated, as are the adenine residues at positions 1492 and 1493

3.4
Estimation of the Average Enthalpy and Free Energy Change That Accompanies Protonation of a Paromomycin Amino Group

Inspection of the data in Table 2 reveals ΔH_{int} values of –7.5 and –17.4 kcal mol^{-1} for the RNA binding of paromomycin at pH 6.3 and 7.5, respectively. Although these ΔH_{int} values are independent of buffer ionization effects, they include contributions from *both* intrinsic drug–RNA interactions as well as binding-linked protonation reactions. Thus, the values of ΔH_{int}, coupled with the corresponding values of Δn, allow one to estimate the average enthalpy change (ΔH_{prot}) for protonation of a drug amino group whose protonation is linked to RNA complex formation using the following relationship:

$$\Delta H_{prot} = \frac{\Delta H_{int-pH_1} - \Delta H_{int-pH_2}}{\Delta n_{pH_1} - \Delta n_{pH_2}} \quad (2)$$

Such a calculation yields a ΔH_{prot} value of –10.5 kcal mol^{-1}. This heat of protonation is in excellent agreement with those previously reported [29] for the α-amino groups of amino acids, whose protonation heats typically range from –9.2 to –10.9 kcal mol^{-1}, as well as for 2-amino- and 3-aminoglucose, whose heats of protonation are –10.2 and –10.1 kcal mol^{-1}, respectively.

One can employ a similar approach to estimate the average free energy change (ΔG_{prot}) for protonation of a drug amino group using the following relationship:

$$\Delta G_{prot} = \frac{\Delta G_{pH_1} - \Delta G_{pH_2}}{\Delta n_{pH_1} - \Delta n_{pH_2}} \quad (3)$$

Recall that, at pH 6.3, paromomycin binds to the host RNA with an affinity (K_a) of $(7.1\pm0.7)\times10^6$ M^{-1} (see Table 1). The corresponding binding constant at pH 7.5 (and an identical Na^+ concentration of 150 mM) is $(8.1\pm0.4)\times10^5$ M^{-1}. These binding constants can be converted to binding free energies using the standard Gibbs relationship $\Delta G = -RT \ln K_a$. The binding free energies so calculated are –9.3 and –8.1 kcal mol^{-1} at pH 6.3 and 7.5, respectively. Substitution of these ΔG values into Eq. 3 yields a ΔG_{prot} value of +1.3 kcal mol^{-1}. This value of ΔG_{prot} represents the average energetic penalty associated with the RNA binding-linked protonation of a drug amino group, and translates approximately into an order of magnitude reduction in binding constant. Taken together, these results clearly identify drug protonation reactions as important thermodynamic participants in the specific binding of paromomycin to the A site of 16S rRNA. The results described above also have implications with regard to drug design. In this connection, the results suggest that enhanced drug affinity for the 16S rRNA A site can be achieved through the incorporation of drug modifications that raise the pK_a values of amino groups. Such modifications might include dehydroxylation of positions (e.g., the 3″-, 3′-, and 6-positions) adjacent to

amino groups or conversion of specific amino groups (e.g., the 2′- and 3-amino groups) to alkylamines, both of which would serve to raise the pH range over which the amino functionalities would exist in their NH_3^+ states.

4 Salt Dependence of Paromomycin Binding to the 16S rRNA A Site

The RNA binding-linked protonation studies described in the preceding sections suggest that electrostatic interactions between the drug and the RNA play an important role in paromomycin recognition of the 16S rRNA A site. Such electrostatic contributions to binding can be quantitatively assessed by monitoring the salt dependence of the drug–RNA binding constant (K_a). The salt dependence of K_a can be analyzed according to the polyelectrolyte theories of Manning [30] and Record [31], which invoke the linkage between the nucleic acid binding of a positively charged drug molecule and the release of counterions (e.g., Na^+ ions) from a condensed state surrounding the nucleic acid to a free state in solution. This binding-induced release of counterions provides an entropically favorable contribution to the binding free energy.

Figure 9 shows plots of log K_a versus log $[Na^+]$ for the binding of paromomycin to *Ec*WT at pH 5.5 and pH 9.0. At pH 5.5, the free base form of paromomycin is 98% protonated, while being 82% deprotonated at pH 9.0. Thus, the log K_a versus log $[Na^+]$ plots in Fig. 9 reflect the salt dependence of paromomycin-RNA binding under conditions where the free drug is either significantly deprotonated (pH 9.0) or almost fully protonated (pH 5.5). Note the linearity of these plots, the slope from each of which yields the following quantity [31]:

$$\frac{\partial \log K_a}{\partial \log [M^+]} = -Z\varphi \tag{4}$$

In this relationship, $[M^+]$ is monovalent cation concentration, Z denotes the apparent charge on the bound drug, and φ is the fraction of M^+ bound per nucleic acid phosphate. The value of φ for the A–form poly(rA)•poly(rU) duplex is 0.89, while that for single-stranded poly(rA) is 0.78 [31]. The φ value for the internal loop region of the 16S rRNA A site may lie between these two values. Linear regression analyses of the plots in Fig. 9 yield slopes ($-Z\varphi$) of -4.6 ± 0.1 and -3.6 ± 0.1 at pH 5.5 and 9.0, respectively. These $-Z\varphi$ values indicate that, at pH 9.0, four drug amino groups participate in electrostatic interactions with the host RNA, but all five amino groups do so at pH 5.5. In other words, under conditions where the five amino groups of the unbound drug exist in their fully protonated NH_3^+ states (i.e., pH 5.5), they contribute electrostatically to the formation of the drug–RNA complex. By contrast, under conditions where the five amino groups of the unbound drug exist predominantly in their deprotonated NH_2 states (i.e., pH 9.0), only four of the five amino groups contribute electro-

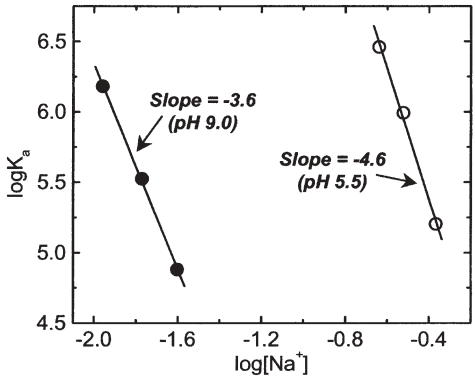


Fig. 9 Salt dependence of the association constants at 25°C for the binding of paromomycin to *Ec*WT at pH 5.5 and 9.0. The experimental data points were fit by linear regression (with the resulting fits depicted as *solid lines*) to yield slopes of –4.6 at pH 5.5 and –3.6 at pH 9.0

statically to the formation of the drug–RNA complex. This latter observation is consistent with the pH-dependent ITC and NMR studies discussed above, which revealed that the RNA binding of paromomycin requires four of the five drug amino groups to adopt their protonated NH_3^+ states.

The determination of $Z\varphi$ as described above allows one to calculate the polyelectrolyte contribution to the RNA binding free energy (ΔG_{pe}) using the following relationship [31]:

$$\Delta G_{pe} = Z\varphi \, RT \ln [M^+] \tag{5}$$

This calculation yields ΔG_{pe} values of –5.2 and –4.0 kcal mol^{-1} for the RNA binding of paromomycin at a Na^+ concentration of 150 mM and a pH of 5.5 and 9.0, respectively. The corresponding binding free energies at 150 mM Na^+ (as derived using K_a values determined from the linear fits of the log K_a versus log [Na^+] plots) are –10.0 kcal mol^{-1} at pH 5.5 and –2.8 kcal mol^{-1} at pH 9.0. A comparison of these binding free energies with the corresponding values of ΔG_{pe} reveals significant (≥50%) contributions to drug-RNA binding from polyelectrolyte effects. Thus, with respect to the targeting of the 16S rRNA A site, the positively charged functionalities of paromomycin comprise an essential part of the drug pharmacophore.

5
Hydration and Paromomycin Recognition of the 16S rRNA A Site

The osmotic stress method has been used extensively to evaluate the participation of water molecules in a wide variety of biochemical reactions, including drug-DNA interactions [32–36]. One version of this method involves the

addition of neutral solutes or cosolvents (osmolytes) to solutions containing the macromolecules and the ligands being studied, thereby altering the water activity in the solution. This osmotic stress method can be used to investigate the hydration changes, if any, that accompany the binding of paromomycin to *Ec*WT. An important consideration when taking such an approach is the selection of appropriate osmolytes that will alter water activity without introducing volume exclusion effects.

Table 4 summarizes the paromomycin–RNA association constants obtained in the presence (K_{a-Osm}) and absence (K_a) of the osmolytes ethylene glycol and glycerol at the indicated osmolalities (Osm). Note that the presence of either osmolyte at osmolalities up to 6.5 does not significantly alter the drug–RNA binding constant, with any observed differences between K_{a-Osm} and K_a being essentially within the experimental uncertainty. These observations suggest that the RNA binding of paromomycin is accompanied by little or no net change in hydration. In other words, we find no evidence for a net uptake or release of water molecules upon complex formation. This result is somewhat surprising in light of the recent crystallographic study by the Westhof group [22], which revealed several water-mediated contacts in the paromomycin complex with a 16S rRNA A site model oligomer. It is possible that such interactions make use of water molecule that preexist in the hydration shells of the free RNA and/or drug, a concept that has been previously proposed by the Westhof group based on an analysis of the hydration patterns in the crystal structures of a multitude of nucleic acids as well as of protein-nucleic acid and drug-nucleic

Table 4 Osmolyte dependence of the association constant at 5°C for the binding of paromomycin to *Ec*WT at pH 5.5 and Na^+ concentration of 230 mM

Osmolyte	Osmolality (OsM)	K_a (M^{-1})	K_{a-Osm} (M^{-1})
None	0	$(2.9\pm0.3)\times10^6$	–
Ethylene glycol	2.1	–	$(2.8\pm0.1)\times10^6$
Ethylene glycol	4.2	–	$(3.5\pm0.1)\times10^6$
Ethylene glycol	6.2	–	$(3.7\pm0.1)\times10^6$
Glycerol	2.2	–	$(2.5\pm0.1)\times10^6$
Glycerol	4.3	–	$(3.8\pm0.2)\times10^6$
Glycerol	6.5	–	$(3.4\pm0.1)\times10^6$

Solutions contained 10 mM sodium cacodylate, 0.1 mM EDTA, and sufficient NaCl to bring the total Na^+ concentration to 230 mM. The osmolalities of the experimental solutions were measured using an Osmette A Model 5002 freezing point depression osmometer (Precision Systems). Control UV melting experiments were conducted (aspreviously described [26]) to ensure that the RNA remained in its folded duplex form in the presence of the osmolytes. The presence of the osmolytes decreased the T_m of the RNA. However, at the concentrations of osmolytes used here, the T_m of the RNA remained well above the temperature used in our binding studies.

acid complexes [37–39]. Regarded as a whole, our osmotic stress studies suggest that water does not provide a major driving force for the binding of paromomycin to the A site of 16S rRNA. The generality of this observation with regard to other aminoglycosides remains to be assessed.

6 Comparing the Energetics of Paromomycin Binding to Prokaryotic Versus Eukaryotic rRNA A Sites

Eukaryotic ribosomes (including human ribosomes) are known to be resistant to the deleterious effects of 2-DOS aminoglycosides [40]. One of the key differences between the sequences of prokaryotic and eukaryotic rRNA A sites is the identity of the base at position 1408 (by *E. coli* numbering), with this base being an adenine in prokaryotes and a guanine in eukaryotes. Significantly, prokaryotic ribosomes whose 16S rRNA carries the A1408G mutation also are resistant to aminoglycosides [41]. Differential RNA binding energetics may play a role in dictating the differential sensitivities of prokaryotic and A1408G-containing ribosomes to aminoglycosides.

Table 5 shows the ITC-derived thermodynamic profiles for the binding of paromomycin to *Ec*WT and the corresponding A1408G mutant. These thermodynamic profiles were acquired at pH 5.5, thereby eliminating significant contributions to the observed binding energetics from binding-linked drug protonation reactions. Inspection of the data in Table 5 reveals that the equilibrium constant for paromomycin-*Ec*WT binding, $(3.7\pm0.7)\times10^7$ M^{-1}, is approximately 31-fold greater than the corresponding value, $(1.2\pm0.3)\times10^6$ M^{-1}, for paromomycin binding to the A1408G mutant. Thus, the A1408G substitution reduces paromomycin affinity for the rRNA A site by more than an order of magnitude. The ~31-fold higher affinity with which paromomycin binds the *Ec*WT duplex relative to the A1408G mutant translates to a difference in binding free energy of 2.0 kcal mol^{-1}, with this differential binding free energy being enthalpic in origin (Table 5 and Fig. 10). In fact, the enhanced binding affinity of paromomycin for the *Ec*WT duplex occurs despite a less favorable entropic contribution to binding (+7.5 versus +9.4 kcal mol^{-1}), which is overcompensated by a more

Table 5 Thermodynamic profiles for the binding of paromomycin to *Ec*WT and its A1408G mutant at pH 5.5, Na^+ concentration of 150 mM, and temperature of 25°C

RNA	K (M^{-1})	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	TΔS (kcal mol^{-1})
*Ec*WT	$(3.7\pm0.7)\times10^7$	−10.3±0.1	−2.8±0.1	+7.5±0.2
A1408G	$(1.2\pm0.3)\times10^6$	−8.3±0.2	+1.1±0.1	+9.4±0.3

Data were obtained and analyzed as described in [47].

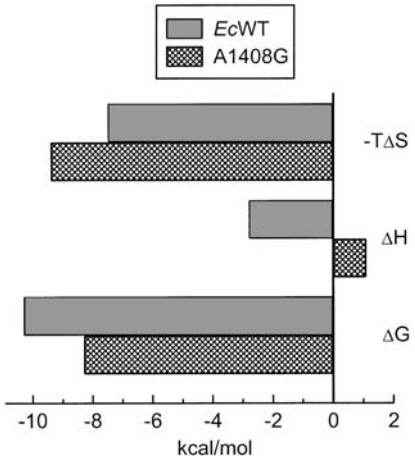


Fig.10 Comparison of the thermodynamic profiles for the binding of paromomycin to *Ec*WT and its A1408G mutant (A1408G) at a temperature of 25°C, a Na^+ concentration of 150 mM, and pH 5.5. The data were obtained as detailed in [47]

favorable binding enthalpy (–2.8 versus +1.1 kcal mol^{-1}). Thus, the 3.9 kcal mol^{-1} enthalpic penalty conferred by the A1408G substitution results in a 2.0 kcal mol^{-1} reduced affinity for the host RNA. In other words, paromomycin enthalpically distinguishes between the prokaryotic and eukaryotic rRNA sequence.

The Puglisi group has used NMR to determine the solution structures of paromomycin in complex with *Ec*WT and its A1408G mutant [18, 42]. The structure of the paromomycin-*Ec*WT complex is shown in Fig. 3B, with the corresponding structure of paromomycin in complex with the A1408G mutant being shown in Fig. 11. A comparison of these two structures suggests that the mol-

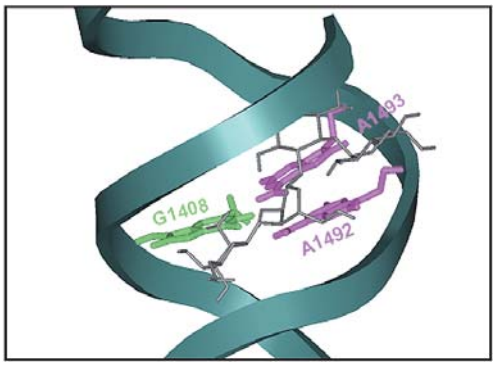


Fig.11 Solution structure of the A1408G mutant form of *Ec*WT in complex with paromomycin (PDB code 1FYP) [42]. The RNA backbone is depicted in ribbon format, with the color coding of the bases at positions 1408, 1492, and 1493 being as described in the legend to Fig. 3

ecular origins of the enthalpically driven 2.0 kcal mol^{-1} enhanced affinity of paromomycin for the prokaryotic versus the eukaryotic rRNA sequence may be derived from the differing geometries of the two binding sites. In particular, A1492 and A1493 are not destacked in the paromomycin–A1408G mutant complex the way they are in the corresponding paromomycin–*Ec*WT complex. These differing binding site geometries are dictated by the identity of the base at the 1408 position. The A1408 base that is present in the prokaryotic sequence forms a noncanonical A•A pair with A1493 [19], with this pair being G1408•A1493 in the eukaryotic sequence [42]. From an energetic standpoint, the G•A pair is known to be more stable than the A•A pair [43]. This enhanced stability may confer too great an energetic barrier for drug-induced destacking of A1493 (and, in turn, A1492) in the eukaryotic sequence.

7
Potential Role for Paromomycin-Induced rRNA Base Destacking in Determining Its Mechanism and Specificity of Action

The reduced affinity of paromomycin for the eukaryotic versus prokaryotic rRNA sequence revealed by the calorimetric studies described above may be a determinant in the specificity of paromomycin for prokaryotic ribosomes. However, the affinity exhibited by paromomycin for the eukaryotic rRNA sequence, $(1.2\pm0.3)\times10^6$ M^{-1}, is comparable to or even greater than those exhibited by other 2-DOS aminoglycoside antibiotics (e.g., tobramycin, kanamycins A and B, ribostamycin, and neamine) for the prokaryotic rRNA sequence [7, 26, 44, 45]. Thus, RNA binding affinity alone does not appear sufficient to dictate the specificities and activities of aminoglycosides. It is likely that aminoglycoside-induced conformational changes in the host rRNA also modulate these biological endpoints, perhaps to an even greater extent than RNA binding affinity. As described in Sects. 2 and 6, a hallmark of such aminoglycoside-induced conformational changes is the displacement of the two conserved adenine residues at positions 1492 and 1493 (by *E. coli* numbering) out of the helical stack toward the minor groove (compare Fig. 3A with Fig. 3B, Fig. 4, and Fig. 8). One model for the molecular mechanism by which the specific aminoglycoside-rRNA recognition event results in aberrant translation invokes the destacking of these two adenines as a key component [15, 46]. According to this model, the destacked bases interact with and stabilize the minihelix that is formed by the mRNA codon and the tRNA anticodon during the translation process. It is thought that even minihelices formed between *noncognate* anticodons and codons are also stabilized in this manner, thereby giving rise to observed mistranslation effects of the aminoglycosides.

A fluorescence-based approach for detecting and characterizing the aminoglycoside-induced conformational change in the A site of rRNA has been developed in our laboratory [47]. This approach makes use of the fluorescent base analog 2-aminopurine (2AP), which was inserted into both *Ec*WT and its A1408G mutant in place of A1492. Hereafter, the 2AP-substituted forms of the

*Ec*WT and A1408G duplexes will be referred to as *Ec*WT(2AP) and A1408G(2AP), respectively. A1492 was selected as the base for 2AP substitution, since it is not engaged in any base pairing interactions (canonical or noncanonical) and its substitution does not perturb the structure or stability of the host RNA duplex [47]. In a seminal set of solvent-dependent fluorescence studies, Ross and coworkers have shown that the fluorescence of 2AP is modulated by hydrophobic stacking interactions and dynamic collisions with other bases, but not by hydrogen bonding interactions [48]. Thus, aminoglycoside-induced changes in the fluorescence of 2AP1492 provide a means for monitoring and characterizing the destacking of the base at position 1492 that accompanies aminoglycoside-RNA complex formation.

Figure 12 presents the steady-state fluorescence emission spectra of *Ec*WT(2AP) and A1408G(2AP) in the presence and absence of paromomycin or

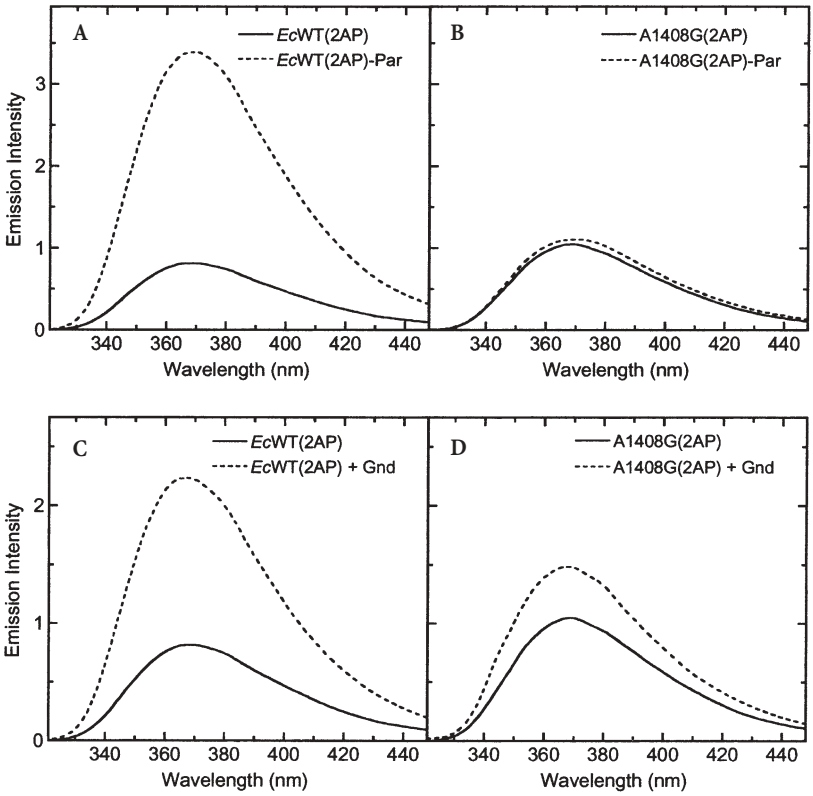


Fig.12 Steady-state fluorescence emission spectra at 25°C of the *Ec*WT(2AP) (**A** and **C**) and A1408G(2AP) (**B** and **D**) duplexes in the presence (*dashed lines*) and absence (*solid lines*) of paromomycin (Par) (**A** and **B**) or 6 M guanidine•HCl (Gnd) (**C** and **D**). Solution conditions were 10 mM sodium cacodylate (pH 5.5), 0.1 mM EDTA, and sufficient NaCl to bring the total sodium ion concentration to 60 mM. The data were obtained as detailed in [47]

guanidine•HCl. The binding of paromomycin to *Ec*WT(2AP) induces a marked increase in the fluorescence intensity of the RNA (Fig. 12A). By contrast, paromomycin binding does not significantly alter the fluorescence of A1408G(2AP) (Fig. 12B). These paromomycin-induced changes in the steady-state fluorescence of *Ec*WT(2AP) and A1408G(2AP) can be quantified in the form of quantum yields (Φ), which are listed in Table 6. Paromomycin binding to *Ec*WT(2AP) results in an increase of Φ from 0.030±0.001 to 0.077±0.001. By contrast, paromomycin binding to A1408G(2AP) results in a decrease of Φ from 0.076±0.001 to 0.063±0.001. These observations agree with the structural database (Figs. 3, 4, 8, and 11) in being consistent with the binding-induced destacking of the base at position 1492 only when *Ec*WT(2AP), but not when A1408G(2AP), serves as the host duplex [15, 18, 19, 22, 42]. Note that the fluorescence intensities of both *Ec*WT(2AP) and A1408G(2AP) also increase upon denaturation with guanidine•HCl (Fig. 12C,D), supporting our correlation of paromomycin-induced changes in RNA fluorescence with corresponding changes in base stacking.

Combining fluorescence quantum yield (Φ) and lifetime (τ) information allows one to quantify the extent of base stacking by calculating the fraction of intrahelical stacked base ($f_{stacked}$) using the following relationship [49]:

$$f_{stacked} = 1 - \Phi_{rel}/\overline{\tau}_{rel} \quad (6)$$

In this relationship, Φ_{rel} ($=\Phi/\Phi_{r2AP}$) and $\overline{\tau}_{rel}$ ($=\overline{\tau}/\overline{\tau}_{r2AP}$) are the quantum yield and amplitude-weighted fluorescence lifetime of 2AP in the RNA or drug–RNA complex relative to the free riboside (r2AP). Barkley and coworkers have previously reported Φ and $\overline{\tau}$ values for r2AP of 0.74 and 10.7 ns, respectively [50]. By definition, the value of $f_{stacked}$ for r2AP is zero. Table 6 summarizes the steady-state and time-resolved fluorescence parameters for *Ec*WT(2AP), A1408G(2AP), and their paromomycin complexes. Note that, when *Ec*WT(2AP) serves as the host duplex, paromomycin binding decreases $f_{stacked}$ from 0.62±0.04 to 0.47±0.04. By contrast, paromomycin binding to A1408G(2AP) does not significantly

Table 6 Steady-state and time-resolved fluorescence parameters for *Ec*WT(2AP), A1408G(2AP), and their paromomycin complexes at 25°C

Sample	Φ (±0.001)	$\overline{\tau}$ (ns)	$f_{stacked}$
r2AP	0.74	10.7	0
*Ec*WT(2AP)	0.030	1.13±0.07	0.62±0.04
*Ec*WT(2AP)-Par	0.077	2.09±0.10	0.47±0.04
A1408G(2AP)	0.076	3.01±0.15	0.64±0.03
A1408G(2AP)-Par	0.063	2.68±0.14	0.66±0.02

Values of Φ and $\overline{\tau}$ for r2AP were taken from [50].The remaining data were obtained and analyzed as described in [47].

alter $f_{stacked}$, with the drug-induced change in $f_{stacked}$ being within the experimental uncertainty. In other words, paromomycin binding increases the average number of *Ec*WT(2AP) molecules in which the 2AP at position 1492 is destacked by 15%, while exerting little or no influence on the stacking of 2AP1492 in A1408G(2AP) molecules. Thus, $f_{stacked}$ can provide a quantitative measure for the detection and characterization of the conformational change in the host rRNA induced by aminoglycoside binding.

The fluorescence-based approach described above has the advantage of not requiring the drug to be conjugated to a fluorescent molecule, which can alter RNA binding properties relative to the parent compound. Puglisi and coworkers have reported an elegant NMR-based technique for detecting antibiotic-induced conformational changes in rRNA [51]. The fluorescence-based technique described here provides an alternative approach that is particularly useful when material and solubility are limiting factors. Furthermore, the fluorescence-based approach can be easily incorporated into a high-throughput regimen for the screening of both existing and novel compounds, with no chromophoric requirements on the part of these compounds. Finally, incorporation of anisotropy measurements into the methodology described here offers the potential of obtaining dynamic information (on a ns timescale) about the drug-induced conformational change that is difficult to derive from the structural database alone.

8 Summary

The number of infections caused by microorganisms that do not respond well to currently available drugs has risen markedly in recent years. This alarming trend has provided impetus for the detailed study of antibiotic agents, and, ultimately, expansion of the current clinical arsenal. The aminoglycosides represent a paradigm for drug–ribosome interactions, and information gleaned from their study has relevance to other ribosome-directed antibiotics of acute clinical importance. The structural and thermodynamic data summarized here enhance our understanding of the molecular forces that govern aminoglycoside recognition of the rRNA A site, and underlie the mechanism and specificity of action of these drugs. Such information is critical to the formation of a rational basis for future drug design strategies.

Acknowledgements Supported by grants S10 RR15959–01 and CA 097123 from the NIH and RSG-99–153–04-CDD from the American Cancer Society. CMB was supported by an NIH training grant (5T32 GM 08319) in molecular biophysics.

References

1. Martin AR (1998) In: Delgado JN, Remers WA (eds) Wilson and Gisvold's textbook of organic medicinal and pharmaceutical chemistry. Lippincott-Raven, Philadelphia, p 253
2. Kotra LP, Haddad J, Mobashery S (2000) Antimicrob Agents Chemother 44:3249
3. Dorman DE, Paschal JW, Merkel KE (1976) J Am Chem Soc 98:6885
4. Botto RE, Coxon B (1983) J Am Chem Soc 105:1021
5. Gaggelli E, Gaggelli N, Maccotta A, Valensin G, Marini D, Di Cocco ME, Delfini M (1995) Spectrochim Acta Part A 51:1959
6. Kane RS, Glink PT, Chapman RG, McDonald JC, Jensen PK, Gao H, Pasa-Tolíc L, Smith RD, Whiteside GM (2001) Anal Chem 73:4028
7. Kaul M, Barbieri CM, Kerrigan JE, Pilch DS (2003) J Mol Biol 326:1373
8. Gutell RR (1994) Nucleic Acids Res 22:3502
9. Puglisi JD, Blanchard SC, Dahlquist KD, Eason RG, Fourmy D, Lynch SR, Recht MI, Yoshizawa S (2000) In: Garrett RA, Douthwaite SR, Liljas A, Matheson AT, Moore PB, Noller HF (eds) The ribosome: structure, function, antibiotics, and cellular interactions. ASM, Washington, DC, p 419
10. Green R, Noller HF (1997) Annu Rev Biochem 66:679
11. Davies J, Gorini L, Davis BD (1965) Mol Pharmacol 1:93
12. Davies J, Davis BD (1968) J Biol Chem 243:3312
13. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235
14. Powers T, Noller HF (1995) RNA 1:194
15. Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V (2000) Nature 407:340
16. Wimberly BT, Brodersen DE, Clemons WMJ, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V (2000) Nature 407:327
17. Recht MI, Fourmy D, Blanchard SC, Dahlquist KD, Puglisi JD (1996) J Mol Biol 262:421
18. Fourmy D, Recht MI, Blanchard SC, Puglisi JD (1996) Science 274:1367
19. Fourmy D, Yoshizawa S, Puglisi JD (1998) J Mol Biol 277:333
20. Yoshizawa S, Fourmy D, Puglisi JD (1998) EMBO J 17:6437
21. Lynch SR, Gonzalez Jr RL, Puglisi JD (2003) Structure 11:43
22. Vicens Q, Westhof E (2001) Structure 9:647
23. Vicens Q, Westhof E (2002) Chem Biol 9:747
24. Vicens Q, Westhof E (2003) J Mol Biol 326:1175
25. Jin E, Katritich V, Olson WK, Kharatisvili M, Abagyan R, Pilch DS (2000) J Mol Biol 298:95
26. Kaul M, Pilch DS (2002) Biochemistry 41:7695
27. Doyle ML, Louie G, Dal Monte PR, Sokoloski TD (1995) Methods Enzymol 259:183
28. Lesniak W, McLaren J, Harris WR, Pecoraro VL, Schacht J (2003) Carbohydr Res 338:2853
29. Christensen JJ, Hansen LD, Izatt RM (1976) Handbook of proton ionization heats. Wiley, New York
30. Manning GS (1978) Q Rev Biophys 11:179
31. Record MTJ, Anderson CF, Lohman TM (1978) Q Rev Biophys 11:103
32. Robinson CR, Sligar SG (1995) Methods Enzymol 259:395
33. Parsegian VA, Rand RP, Rau DC (1995) Methods Enzymol 259:43
34. Garner MM, Rau DC (1995) EMBO J 14:1257
35. Spink CH, Chaires JB (1999) Biochemistry 38:496
36. Qu X, Chaires JB (2001) J Am Chem Soc 123:1
37. Westhof E (1987) Int J Biol Macromol 9:186
38. Westhof E, Dumas P, Moras D (1988) Biochimie 70:145

39. Auffinger P, Westhof E (2000) J Mol Biol 300:1113
40. Wilhelm JM, Pettitt SE, Jessop JJ (1978) Biochemistry 17:1143
41. Recht MI, Puglisi JD (2001) Antimicrob Agents Chemother 45:2414
42. Lynch SR, Puglisi JD (2001) J Mol Biol 306:1037
43. Turner DH (2000) In: Bloomfield VA, Crothers DM, Tinoco I Jr (eds) Nucleic acids: structures, properties, and functions University Science, Sausalito, p 259
44. Wong C-H, Hendrix M, Priestley ES, Greenberg WA (1998) Chem Biol 5:397
45. Pilch DS, Kaul M, Barbieri CM, Kerrigan JE (2003) Biopolymers 70:58
46. Yoshizawa S, Fourmy D, Puglisi JD (1999) Science 285:1722
47. Kaul M, Barbieri CM, Pilch DS (2004) J Am Chem Soc 126:3447
48. Rachofsky EL, Osman R, Ross JBA (2001) Biochemistry 40:946
49. Rachofsky EL, Seibert E, Stivers JT, Osman R, Ross JBA (2001) Biochemistry 40:957
50. Tsujikawa L, Strainic MG, Watrob H, Barkley MD, deHaseth PL (2002) Biochemistry 41:15334
51. Lynch SR, Puglisi JD (2000) J Am Chem Soc 122:7853
52. Fukada H, Takahashi K (1998) Proteins: Struct Funct Genet 33:159
53. Fourmy D, Recht MI, Puglisi JD (1998) J Mol Biol 277:347
54. Marky LA, Breslauer KJ (1987) Biopolymers 26:1601

Author Index Volumes 251 – 253

Author Index Vols. 26–50 see Vol. 50
Author Index Vols. 51–100 see Vol. 100
Author Index Vols. 101–150 see Vol. 150
Author Index Vols. 151–200 see Vol. 200
Author Index Vols. 201–250 see Vol. 250

The volume numbers are printed in italics

Subject Index