# Small Molecule DNA and RNA Binders

From Synthesis to Nucleic Acid Complexes

*M. Demeunynck, C. Bailly, W. D. Wilson (Eds.)*

WILEY-VCH

**Small Molecule DNA and RNA Binders**

*M. Demeunynck, C. Bailly, W. D. Wilson (Eds.)*

## *Related Titles from WILEY-VCH*

Anneliese E. Barron (Ed.)

**DNA Sequencing**

2002, ca. 500 pages.
Hardcover. ISBN 3-527-30599-8

Susanne Brakmann and Kai Johnsson (Eds.)

**Directed Molecular Evolution of Proteins**

2002, 368 pages.
Hardcover. ISBN 3-527-30423-1

O. Zerbe, R. Mannhold, H. Kubinyi and G. Folkers (Eds.)

**BioNMR in Drug Research**

2002, ca. 350 pages.
Hardcover. ISBN 3-527-30465-7

Douglas T. Gjerde, Christopher P. Hanna and David D. Hornby

**DNA Chromatography**

2002, 243 pages.
Hardcover. ISBN 3-527-30244-1

# Small Molecule DNA and RNA Binders

From Synthesis to Nucleic Acid Complexes

*M. Demeunynck, C. Bailly, W. D. Wilson (Eds.)*

WILEY-VCH

***Dr. Martine Demeunynck***
Université Joseph Fourier
BP53
38041 Grenoble cedex
France

***Dr. Christian Bailly***
Institut de Recherches sur le Cancer
INSERM Unité 124
Place de Verdun
59045 Lille cedex
France

***Prof. Dr. W. David Wilson***
Department of Chemistry
Georgia State University
University Plaza
Atlanta GA 30303-3083
USA

***Cover design***
Christian Coulombeau



**Library of Congress Card No.: applied for**
A catalogue record for this book is available from the British Library.
**Bibliographic information published by Die Deutsche Bibliothek**
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at http://dnb.ddb.de.



Printed in the Federal Republic of Germany.
Printed on acid-free paper.

**Typesetting** Asco Typesetters, Hong Kong
**Printing** betz-druck gmbH, Darmstadt
**Bookbinding** Litgas & Dopf Buchbinderei GmbH, Heppenheim

**ISBN** 3-527-30595-5

## Contents

# Preface

The ultimate goal of most organic-medicinal chemists is to see the small molecule that they have synthesized become a useful drug for the treatment of human diseases. Unfortunately, even with modern technology this is an extremely rare event. In most cases, the compounds designed and synthesized (generally with pain and passion) have a brief existence that does not exceed the first biological activity assay. The valley between chemistry and therapeutics is deep and difficult to cross but nevertheless the two disciplines are intimately associated. It is our goal to help construct a bridge between the makers of the small molecules and the users. Over the past two decades, a relatively large number of useful anticancer and antiparasitic drugs have been discovered or rationally designed based on the principle of nucleic acids recognition. A better understanding of the molecular rules that govern interactions between small molecules and the many sequences and structures of DNA and RNA is pivotal to the development of novel drug candidates. How does the drug adapt to the nucleic acid target (and *vice versa*)? How do nucleic acid structures affect ligand binding? How do small molecules read the genetic information? These types of questions continue to excite our scientific curiosity and the quest for better DNA/RNA binders drives modern researchers much as the search for the Holy Grail did the ancients.

Design and development of nucleic acid targeted drugs is a challenging enterprise but real breakthroughs have been made in recent years and many are reported here. This volume is intended to give the reader an up-to-date view of the current status and expected developments in research involving ligand-nucleic acid recognition. This book was built on a discussion among the three of us on how chemistry, biophysical chemistry and pharmacology serve our field to help design new drugs. Our different but complementary view angles on the subject prompted us to edit this volume focussed on DNA/RNA recognition by a variety of small molecules: peptides, intercalators, groove binders, metal complexes. The various DNA structures that can be targeted by drugs are also considered and the field of natural products is partially covered. Altogether, the 25 chapters of this volume survey most of the drug categories that bind, bond or cleave nucleic acids. The reader will notice the diversity of small molecules mentioned here, from marine products to platinum complexes, from G4-binders to RNA cleaving agents, from abasic site selective agents to aptamers, as well as the panel of biophysical and

biochemical approaches routinely used to investigate the structures and dynamics of drug-nucleic acids complexes. The portraits of specific drug families (anthracyclines, indolocarbazoles, bleomycins, ...) are also thoroughly presented. The amalgam was deliberately chosen to cross ideas of organic chemists and biophysicists and those more interested in the therapeutic end point of the research.

The volume starts with a general introduction (magisterially presented in a British style) and then it flies over the world, from several countries in Europe (Spain, Italy, France, Czech Republic, UK) to the USA, via Japan and New Zealand, illustrating the essential international character of the research (and the friendly atmosphere of the edition). Inevitably we have neglected (mostly for consideration of space) a number of interesting areas that should have been cited here, such as clinical applications. But the gallery of molecules presented in this volume must be considered as a live exhibit to explore and to use for further drug design. Come on in, and like us, become fascinated by the "small molecules" that bind or bite the genetic material in its many forms. We hope you will also find examining this volume an enriching experience.

The enterprise was very exciting and proceeded smoothly (with no delay!) thanks to the enthusiastic contribution of all the authors. We are grateful to everyone for delivering their manuscript on time (and in some cases even well before the deadline!) and for making our task as editors such a memorable one. We also thank our "artist" Christian Coulombeau who kindly drew the front cover, sort of a railway to the future. Finally, we shall dedicate this volume to our colleagues who left the world too quickly to contribute (Marc Leng, David S. Sigman, and Peter A. Kollman in particular).

M.D., C.B., W.D.W.
Grenoble, Lille and Atlanta
September 2002

## Contributors

Dr. Christian Bailly
Institut de Recherches sur le Cancer de Lille
INSERM U-524
Place de Verdun
59045 Lille Cedex
France

Prof. Dr Jacqueline K. Barton
Division of Chemistry and Chemical Engineering
California Institute of Technology
Pasadena, CA 91125
USA

Dr. David W. Boykin
Department of Chemistry
Georgia State University
Atlanta, Georgia 30303
USA

Dr. Viktor Brabec
Institute of Biophysics
Academy of Sciences of the Czech Republic
Kralovopolska 135
61265 Brno
Czech Republic

Dr. Jonathan B. Chaires
Department of Biochemistry
University of Mississippi Medical Center
2500 North State Street
Jackson, Mississippi 39216-4505
USA

Dr. Jean-François Constant
LEDSS
Université Joseph Fourier
BP 53
38041 Grenoble cedex 9
France

Dr. Peter C. Dedon
Division of Bioengineering and Environmental Health
Massachusetts Institute of Technology
Cambridge, MA 02139
USA

Dr. William A. Denny
Auckland Cancer Society Research Centre
Faculty of Medicine and Health Science
The University of Auckland, Private Bag 92019
Auckland

Prof. Dr. Keith R. Fox
Division of Biochemistry & Molecular Biology
School of Biological Sciences
University of Southampton
Bassett Crescent East
Southampton SO16 7PX
UK

Prof. Dr. Sidney M. Hecht
Department of Chemistry
University of Virginia
Charlottesville, Virginia 22901
USA

Dr. Laurence H. Hurley
Howard J. Schaeffer Endowed Chair in Pharmaceutical Sciences
Arizona Cancer Center
1515 N. Campbell Ave.
Room 4949
Tucson, AZ 85724
USA

Dr. Moses Lee
Department of Chemistry
Furman University
Greenville, SC 29613
USA

Dr. Eric Long
Department of Chemistry
Indiana University/Purdue University
Indianapolis, IN 46202-3274
USA

Dr. Jean-Louis Mergny
Laboratoire de Biophysique
Museum National d'Histoire Naturelle
INSERM U 201 CNRS UMR 8646
43 rue Cuvier
75231 Paris cedex 05
France

Prof. Dr. Stephen Neidle
Cancer Research UK Biomolecular Structure Group
The School of Pharmacy
University of London, 29–39
Brunswick Square
London WCIN 1AX
UK

Dr. Manlio Palumbo
Professor of Medicinal Chemistry
Department of Pharmaceutical Sciences
Via Marzolo, 5
35131 Padova
Italy

Dr. Laurence H. Patterson
Department of Pharmaceutical and Biological Chemistry
School of Pharmacy
University of London
Brunswick Square
London, WC1N 1AX
UK

Dr. Daniel Pilch
Department of Pharmacology
University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School
Piscataway, NJ 08854
USA

Dr. Tariq M. Rana
Department of Pharmacology
Robert Wood Johnson Medical School
675 Hoes Lane, Piscataway, NJ 08854
USA

Dr. Steve Rokita
Department of Chemistry and Biochemistry
University of Maryland
College Park, Maryland 20742
USA

Dr. Mark Searcey
Department of Pharmaceutical and Biological Chemistry
The School of Pharmacy
University of London
29/39 Brunswick Square
London WC1N 1AX
UK

Dr. Shigeori Takenaka
Department of Applied Chemistry
Faculty of Engineering
Kyushu University
Fukuoka 812-8581
Japan

Dr. David E. Thurston
Professor of Anticancer Drug Design
The School of Pharmacy
University of London
Brunswick Square
London
UK

Prof. Dr. Yitzhak Tor
Department of Chemistry & Biochemistry
University of California, San Diego
9500 Gilman Drive
San Diego, CA 92093-0358
USA

Dr. Jean-Pierre Vigneron
Collège de France, Chimie des interactions moléculaires
11 pl Marcelin Berthelot
75231 Paris cedex 5
France

Prof. M.J. Waring
Department of Pharmacology
University of Cambridge
Tennis Court Road
Cambridge CB2 1QJ
UK

Prof. Dr. W. David Wilson
Department of Chemistry
Georgia State University
University Plaza
Atlanta, GA 30303-3083
USA

# 1
# Forty Years On

*Michael J. Waring and L. P. G. Wakelin*

## 1.1
## Early Experiments Prior to Molecular Modeling

The quest to understand specific interactions between drugs and nucleic acids dates back a long time – more than 40 years. Even though the concept of gene targeting could not be explicitly formulated until much later, there were early realizations that DNA could provide a fine receptor for drugs. A major turning point in the history of drug binding to DNA, the publication in 1961 of the intercalation hypothesis by Leonard Lerman [1], in many people's estimation represents the true birth of the subject, but it would be wrong to neglect mention of the contributions of earlier workers. These workers knew they were dealing with drug–nucleic acid interactions and must have had some inkling of the future importance of the topic. Among them were the histologists who employed dyes such as aminoacridines to stain cells and tissue sections, particularly the fluorescent dye acridine orange, whose capacity to cause nuclei to fluoresce bright green while the cytoplasm fluoresced red was a valuable tool in histology and cell biology. Indeed in the researches of these pioneers can be found the first evidence that particular dyes can react differently with different kinds of nucleic acid-containing structures and therefore that the small molecules must be capable of some form of discrimination based upon what we would today call molecular recognition. From the variable and sometimes capricious performance of substances such as acridines employed as stains it could also be surmised that depending upon the solvent conditions a given dye might react in more than one way with its 'receptor,' foreshadowing the concept of heterogeneity in binding that was later to occupy the attention of biophysicists.

At the same time, thanks to the seminal work of Paul Ehrlich half a century earlier, the usefulness of dyes – particularly aminoacridines – as antiseptics and antimalarials was widely recognized so that the connection between cell staining and useful biological activity was more than implicit. Thus it happened at a critical moment that the potency of proflavine as a mutagen was recognized. This led to the brilliant experimental work of Crick, Brenner and colleagues [2] showing that exposure of bacteriophage-infected bacteria to proflavine produced frameshift mutations – a phenomenon that enabled them to deduce the triplet nature of the

genetic code. Meanwhile, the careful experiments of Peacocke and Skerrett [3] on the interaction of proflavine with purified DNA were under way and the first truly quantitative measurements of a reversible drug–DNA binding reaction became available, complete with a proper description of the metachromatic shift in the absorption spectrum, application of spectrophotometry and equilibrium dialysis to determine genuine binding constants, and clear evidence of the occurrence of secondary binding after saturation of the strong primary binding sites had been accomplished.

Now all the elements were in place for Lerman, at that time working in the Cambridge MRC Laboratory of Molecular Biology with Crick and Brenner, to get to work on the intercalation hypothesis. Stone and Bradley [4] disposed of the secondary interaction of acridine orange with nucleic acids by attributing it to the formation of stacked aggregates of dye bound externally to the polyanion.

Two other pre-intercalation areas of endeavor must be mentioned, the first of which is the action of the antibiotic actinomycin D. Actinomycin had been discovered in the 1940s and was the first antibiotic found to be highly active against certain tumors – indeed, through the 1950s and early 1960s it was reckoned to be the most potent anticancer agent available in the arsenal of chemotherapy. The antibiotic was known to be capable of inhibiting nucleic acid synthesis in susceptible cells, a process that was consequently identified as a prime target for anticancer chemotherapy. The discovery of mRNA and the process of gene transcription owes much to the earnest work of early cell biologists who showed that actinomycin was an exquisitely selective inhibitor of transcription by virtue of its specific inhibitory action on the newly discovered enzyme RNA polymerase; that in turn was attributable to tight but reversible binding of actinomycin to the double-helical DNA template [5]. These discoveries firmly established the business of ligand–DNA interaction as a matter of concern to biologists, clinicians, and a breed of pharmacologists who later emerged as key players in founding what was to become the illustrious discipline of molecular interactions.

The second area of endeavor, though it had little influence on the development of ideas about reversible ligand–nucleic acid interactions, is the remarkable work of people like Kohn, Brooks, and Lawley on nitrogen mustards and comparable alkylating agents used for cancer chemotherapy [6, 7]. We should recall that nitrogen mustards were the very first chemicals used to treat cancer, prompted by unhappy events that occurred during the Second World War; it is indeed salutary that so evidently worthy a purpose as the alleviation of suffering from one of humanity's most dreaded diseases should have come about in such an inauspicious manner. The determined attentions of a few medically minded individuals capable of grappling with rather complicated and messy chemistry did a lot to clarify the mechanisms of action of these highly reactive substances, and again the critical target turned out to be DNA. Painstaking analysis of the products formed *in vitro* and *in vivo* when cells were exposed to mustards eventually identified the N7 position of the guanine ring as the most reactive (i.e. nucleophilic) site for alkylation of DNA, and the perceived correlation of anticancer activity with the possession of two alkylating centers spaced some five atoms apart led to the concept that bifunctional reactivity must be crucial for therapeutic effect.

## 1.2 Formulation of Molecular Models and Mechanisms of Binding to DNA

Before we return to the historic turning point at which the intercalation hypothesis was born, it is logical to finish consideration of the early alkylating agents by referring to their identification as crosslinking agents capable of covalently linking the complementary strands of the DNA double helix. At first it was thought that this action would adequately explain their cytotoxic activity through inhibiting the progress of the replication fork, but more recently the possible contribution of *intra*-strand crosslinks and DNA–protein crosslinks has brought this assumption into question [8, 9]. Meanwhile other types of powerful alkylating agents have been discovered that do not necessarily form interstrand crosslinks but are endowed with excellent biological activity. Moreover, an early twist to the tale of covalent reaction with DNA came with the finding that the antibiotic mitomycin C must be activated by reduction prior to forming inter-strand DNA crosslinks; this discovery added impetus to the idea of bioreductive activation of pro-drugs, particularly for cancer treatment, which has become an important focus for the efforts of several groups of drug designers (see Chapter 9). There is also a complex relation between bond-forming and bond-breaking interactions with nucleic acids that can be seen with several DNA-binding compounds described elsewhere in this volume (see Chapters 3 and 23).

A unifying theme that runs through these lines of work, and indeed throughout the volumes of this publication, is the extraordinarily sophisticated chemistry that attends the reaction of many compounds with nucleic acids, not to mention the amazing biosynthetic capabilities of the organisms that produce those substances that are of natural origin. Neither should we belittle the remarkable inventiveness and achievements of the organic chemists who increasingly are succeeding in their efforts to design strategies to come up with novel DNA-reactive compounds for chemotherapy as well as other purposes.

Returning to the historical thread, we go back to the year 1961, which was when the first reasonably explicit model for binding of a drug to the double helix – the intercalation hypothesis – was proposed for the interaction of aminoacridines like proflavine with DNA [1]. It is no secret, though not often appreciated outside laboratories of molecular biology, that the notion of frameshift mutation furnished a degree of inspiration for the model. However, the idea of intercalation was not universally acclaimed: indeed it was greeted with profound skepticism in certain quarters. Lerman's original evidence, drawn from observations of changes in viscosity, sedimentation coefficient, and X-ray diffraction from oriented fibers of DNA, was perfectly reasonable so far as it went. But that was not far enough to satisfy many of the "real" structure solvers, who made it clear that they were not going to believe the postulate unless and until it had been verified by their own favorite "direct" technique as opposed to the admittedly rather indirect evidence adduced by Lerman. One of the present authors remembers conversations including such phrases as "do you believe in intercalation?", as if it were an article of faith akin to religion. In due course, experiments were devised to verify or disprove the hypothesis, eventually to the satisfaction (or conversion) of the most hardened skeptics.

One early experiment was the circular DNA unwinding test, based upon the generally (but not quite universally) agreed expectation that intercalation must locally unwind the double helix [10]. It worked, and confirmed ethidium together with aminoacridines and several other interesting ligands, including actinomycin, as intercalators [11]. By the same token, antibiotics like netropsin and distamycin were identified as something else: minor groove binders as we now know [11]. It also became clear that different drugs unwind the helix by different angles when they intercalate, and the test even unearthed certain ligands that could unwind the double helix somewhat without apparently intercalating in the usual sense: steroidal diamines and triphenylmethane dyes [12, 13]. Questions still remain to be answered about these ligands. Of course a legacy of this early work is the detailed understanding of higher order structure, especially circularity, of DNA which studies on drug interactions have helped to elucidate. Thirty-five years after it was first shown to unwind circular DNA, ethidium is still routinely used to isolate plasmids.

Perhaps because of the seminal contributions of physical (bio)chemists during the early years of probing mechanisms of drug–nucleic acid interaction, the study of reaction kinetics soon emerged as a powerful tool for throwing light on the forces involved [14]. Don Crothers, an influential advocate of the kinetic approach, used to remark that the study of kinetics was uniquely valuable, if only because it added a new dimension – time – to the analysis of the phenomena. He was absolutely right. A highlight was the discovery that some ligands which bound well but not outrageously tightly to DNA could be characterized by on-rates and off-rates many orders of magnitude slower than ostensibly comparable substances. The anthracycline antibiotic nogalamycin is a good case in point; its slow association and dissociation kinetics are attributable to the disruption of base pairing needed to ''thread'' its bulky sugar substituents through the double helix [15]. Slow dissociation kinetics have been correlated with improved biological activity, and underlie the success of Phillips' relatively recent assay for transcription termination at particular drug-binding sites on DNA [16]. Direct ligand transfer between binding sites on DNA without involving complete dissociation from the polymer was evidenced many years ago and has given rise to the ''shuffling'' concept whereby ligands are supposed to migrate one-dimensionally along a DNA molecule in search of better (tighter) binding sites [17, 18].

## 1.3 Specificity of Nucleotide Sequence Recognition

Although the value of DNA and, to a lesser extent, RNA as a target for selective drug action had been evident from the outset, it also quickly became apparent that few known drugs showed much, if any, selectivity for binding to particular nucleotide sequences. Yet the holy grail of selectively suppressing gene expression was conceived early on, together with the realization that to attain this end it would be necessary to recognize moderately long stretches of base pairs. Eventually it was

calculated that one might need to recognize a sequence composed of a number of base pairs in the high teens in order to identify a single targeted site in the human genome. The first experiments aimed at examining drug-binding preferences were crude and laborious to say the least, consisting of little more than attempts to detect different levels of binding to nucleic acids from different sources. Scatchard plots were employed to determine affinity constants, together with the frequency of binding sites, initially by simple and inappropriate means that were eventually much improved by better theoretical treatments like those of McGhee and von Hippel [19]. Sometimes the available methods (spectroscopy, equilibrium dialysis, etc.) were simply inapplicable because of the poor aqueous solubility of the ligands under investigation and alternative techniques had to be devised, such as solvent partition analysis used for the quinoxaline antibiotics [20]. Much effort was required just to establish a preference for, say, GC-rich DNA. Then the steady development of chemical methods for polynucleotide synthesis began to extend the range of synthetic, defined sequences available to the investigator and furnished substrates that could be used to examine whether or not a particular sequence would support interaction with a drug of interest.

A quantum leap occurred in the early 1980s with the invention of footprinting methodology in several laboratories at much the same time, using enzymes or Dervan's cleverly designed synthetic reagent MPE-Fe(II) to cut a cloned radiolabeled DNA fragment [21–23]. At a stroke it became possible to identify exactly where the preferred binding sites for a ligand were on a substrate that amounted to a real gene or a chosen fragment of a known gene. Although only semiquantitative at first, methods were quickly developed to adapt the technology to provide passable binding constants so that a true thermodynamic comparison of ligand affinity and capacity to discriminate between different sites could be gained in a single experiment or series of experiments. The power of the footprinting technique can be gauged from the reports of sequence-selectivity to be found in several chapters of this book. With its application, a substantial database of binding affinities for different sequences has been amassed, so that it is now becoming possible to enquire about general mechanisms that underlie the recognition of particular base-pair sequences, such as whether binding occurs predominantly in the major groove, the minor groove (much the most common with small molecules), or occasionally both.

Some workers have focused attention on the distinction between "digital" and "analog" readouts of sequence information, based on the notion that microstructural variation in the exact parameters of the double helix (groove width, for example) can sensitively reflect nucleotide sequence heterogeneity and therefore afford a means of sequence recognition that is independent of direct, specific contacts with the base pairs themselves. One of the techniques that can throw light on such questions involves looking at the behavior of DNA molecules containing unnatural nucleotide substitutions, which have the effect of shifting, removing, or adding specific base substituents. Such experiments have amply confirmed the dominant role of the 2-amino group of guanine in directing many ligands to their preferred binding sites, and have also thrown light upon related questions like the

role of base pair substituents in modulating groove width, reactivity towards alkylating agents, helix curvature or flexibility, and the sequence-dependent winding of DNA around the histone octamer in nucleosome core particles [24].

While footprinting and related gel methodology continues to play a major role in studies of this sort it has recently been joined by the elegant but simple method of competition dialysis, whereby the relative binding of a test ligand to many different types of nucleic acids can be assessed at the same time [25]. This method is of particular interest for investigating structure-specific binding of drugs to nucleic acids or indeed other polymers, whether natural or synthetic.

## 1.4 Details at the Atomic and Molecular Levels

Insight into the structure and dynamics of intercalation complexes has progressed by a close synergy between theoretical and experimental approaches, which today has developed to the point at which it is now possible to give a complete molecular description of what a drug–DNA complex looks like at the atomic level, and how its constituent atoms move in solution. The techniques that have proved invaluable in this quest are X-ray crystallography, NMR spectroscopy, quantum chemistry, molecular mechanics, and molecular dynamics. Lerman himself used X-ray fiber diffraction data from proflavine–DNA complexes as part of the initial evidence he marshaled for the intercalation hypothesis [1], and Fuller and Waring adopted the technique to produce the first molecular model of the ethidium–DNA complex [26].

These early attempts at model building took an important step forward at the beginning of the 1970s when Sobell and colleagues solved the crystal structure of a 2:1 actinomycin–deoxyguanosine complex, which enabled them to construct a fairly precise intercalation model based upon purely geometrical constraints [27]. Later in the 1970s the commercial availability of DNA and RNA dinucleoside monophosphates made possible crystallographic and NMR studies of intercalated mini-duplexes of ethidium and aminoacridines such as 9-aminoacridine, proflavine, and acridine orange. The crystallographic studies by Sobell, Neidle, Rich, and their colleagues provided the first truly atomic description of intercalation complexes, and unequivocally proved that the DNA duplex could indeed stretch so as to sandwich acridine and phenanthridine chromophores between two base pairs [28–30].

These were seminal studies that not only provided insight into the fine details of individual drug–DNA complexes, but also revealed modifications to the geometry of the sugar–phosphate backbone generally required to open the intercalation cavity. Armed with the latter information, Neidle and others were able to construct molecular models of suitably modified B- and A-DNA duplexes containing stereochemically sound intercalation cavities [31, 32]. This provided the means for many investigators, Neidle, Pullman, and Hopfinger prominent amongst them, to use