Thomas Schrader, Andrew D. Hamilton (Eds.)

Functional Synthetic Receptors

WILEY-VCH Verlag GmbH & Co. KGaA

T. Schrader, A. D. Hamilton (Eds.)

Functional Synthetic Receptors

Further Titles of interest:

W. R. Briggs, J. L. Spudich (Eds.)

Handbook of Photosensory Receptors

2005

ISBN 3-527-31019-3

S. Frings, J. Bradley (Eds.)

Transduction Channels in Sensory Cells

2004

ISBN 3-527-30836-9

I. Willner, E. Katz (Eds.)

Bioelectronics

From Theory to Applications

2005

ISBN 3-527-30690-0

C. M. Niemeyer, C. A. Mirkin (Eds.)

Nanobiotechnology

Concepts, Applications and Perspectives

2004

ISBN 3-527-30658-7

G. Krauss

Biochemistry of Signal Transduction and Regulation

Third, Completely Revised Edition

2003

ISBN 3-527-30591-2

Thomas Schrader, Andrew D. Hamilton (Eds.)

Functional Synthetic Receptors

WILEY-VCH Verlag GmbH & Co. KGaA

Editors

Prof. Dr. Thomas Schrader
Philipps-Universität Marburg
Fachbereich Chemie
Hans-Meerwein-Strasse
35032 Marburg
Germany

Prof. Dr. Andrew D. Hamilton
Yale University
Department of Chemistry
Sterling Chemistry Laboratory
225 Prospect Street
New Haven, CT 06520
USA

All books published by Wiley-VCH are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.: Applied for

British Library Cataloging-in-Publication Data:
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>.

© 2005 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – nor transmitted or translated into machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Printed in the Federal Republic of Germany

Printed on acid-free paper

Typesetting TypoDesign Hecker GmbH, Leimen
Printing betz-druck gmbH, Darmstadt
Bookbinding Litges & Dopf Buchbinderei GmbH, Heppenheim

ISBN-10 3-527-30655-2
ISBN-13 978-3-527-30655-8

Table of Contents

Preface *IX*

List of Contributors *XI*

1 Artificial (Pseudo)peptides for Molecular Recognition and Catalysis *1*
L. J. Prins, P. Scrimin
1.1 Introduction *1*
1.2 Recognition of Biological Targets by Pseudo-peptides *2*
1.2.1 Introduction *2*
1.2.2 Polyamides as Sequence-specific DNA-minor-groove Binders *2*
1.2.3 Peptide Nucleic Acids *12*
1.2.4 Protein Recognition by (Pseudo)peptides *17*
1.3 Synthetic (Pseudo)peptide-based Supermolecules: From Structure to Function *20*
1.3.1 Catalytic (Pseudo)peptides *21*
1.3.2 (Pseudo)peptides Altering Membrane Permeability *27*
1.3.3 Nanoparticle- and Dendrimer-based Functional (Pseudo)peptides *30*
1.4 Combinatorial Selection of Functional (Pseudo)peptides *35*
1.5 Conclusions *38*
References *39*

2 Carbohydrate Receptors *45*
A. P. Davis, T. D. James
2.1 Introduction *45*
2.2 Carbohydrate Receptors Employing Noncovalent Interactions *47*
2.2.1 Recognition in Organic Solvents *48*
2.2.2 Recognition in Two-phase Systems *62*
2.2.3 Carbohydrate Recognition in Water *69*
2.3 Receptors Employing B–O Bond Formation *79*
2.3.1 Carbohydrate Recognition in Water *80*
2.3.2 Carbohydrate Recognition in Water *84*
References *104*

3 Ammonium, Amidinium, Guanidinium, and Pyridinium Cations *111*
T. Schrader, M. Maue
3.1 Introduction *111*
3.2 Ammonium Cations *112*
3.2.1 New Receptor Structures *112*
3.2.2 Theoretical Investigations *119*
3.2.3 New Functions *122*
3.2.4 Peptide and Protein Recognition *125*
3.2.5 Conclusion and Outlook *131*
3.3 Amidinium Cations *132*
3.3.1 Introduction *132*
3.3.2 Artificial Receptors *133*
3.3.3 Conclusion *138*
3.4 Guanidinium Cations *138*
3.4.1 Introduction *138*
3.4.2 Artificial Receptors *139*
3.4.3 Conclusion *148*
3.5 Pyridinium Cations *149*
3.5.1 Introduction *149*
3.5.2 Artificial Receptors *149*
3.5.3 Conclusion *157*
3.6 Conclusions and Outlook *157*
References *159*

4 Artificial Pyrrole-based Anion Receptors *165*
W.-S. Cho, J. L. Sessler
4.1 Introduction *165*
4.2 Anions in Biological Systems *166*
4.3 Cationic Pyrrole-based Receptors *170*
4.3.1 Cyclic Receptors *170*
4.3.2 Linear Receptors *180*
4.4 Neutral Pyrrole-based Anion Receptors *183*
4.4.1 Cyclic Receptors *183*
4.4.2 Linear Receptors *200*
4.5 Anion Carriers in Transport Applications *203*
4.6 Anion Sensing *212*
4.7 Guanidinium-based Anion Receptors *228*
4.8 Amide-based Anion Receptors *234*
4.9 Urea-based Anion Receptors *245*
4.10 Conclusions *249*
Acknowledgment *250*
References *250*

5 Molecular Containers in Action *257*
D. M. Rudkevich
5.1 Introduction *257*
5.2 Variety of Molecular Containers *258*
5.3 Chemistry Inside Capsules *272*
5.3.1 Observing Unusual Species Through Encapsulation *272*
5.3.2 Changing Reaction Rates by Encapsulation *275*
5.3.3 Encapsulated Reagents *277*
5.4 Storage of Information Inside Capsules *280*
5.5 Materials and Sensors by Encapsulation *283*
5.5.1 Molecular Containers as Sensors and Sensing Materials *283*
5.5.2 Supramolecular Polymers *286*
5.6 Biologically Relevant Encapsulation *290*
5.6.1 Entrapment of Biologically Active Guests *290*
5.6.2 Encapsulation of Gases *292*
5.7 Concluding Remarks *294*
Acknowledgment *295*
References *295*

6 Formation and Recognition Properties of Dynamic Combinatorial Libraries *299*
A. D. Hamilton, D. M. Tagore, K. I. Sprinz
6.1 Introduction *299*
6.2 Covalent Interactions Used in DCC Design *302*
6.2.1 Acyl Hydrazone and Imine Exchange *302*
6.2.2 Transesterification *310*
6.2.3 Disulfides *311*
6.2.4 Olefin Metathesis *313*
6.3 Noncovalent Interactions Used in DCC Design *315*
6.3.1 Metal Ligand Coordination *315*
6.3.2 Hydrogen Bonding *321*
6.4 Conformational/Configurational Isomerization *326*
6.5 Receptor-based Screening, Selection, and Amplification *328*
6.6 Conclusions *331*
Acknowledgment *331*
References *331*

7 Synthetic Molecular Machines *333*
E. R. Kay, D. A. Leigh
7.1 Introduction *333*
7.1.1 Molecular-level Machines and the Language Used to Describe Them *334*
7.1.2 Principles of Motion at the Molecular Level – the Effects of Scale *335*
7.2 Controlling Conformational Changes *336*
7.3 Controlling Configurational Changes *340*

7.4 Controlling Motion in Supramolecular Systems *346*
7.4.1 Switchable Host–Guest Systems *347*
7.4.2 Intramolecular Ion Translocation *348*
7.5 Controlling Motion in Interlocked Systems *350*
7.5.1 Basic Features *350*
7.5.2 Inherent Dynamics: Ring Pirouetting in Rotaxanes *351*
7.5.3 Inherent Dynamics: Ring Pirouetting in Catenanes *353*
7.5.4 Inherent Dynamics: Shuttling in Rotaxanes *358*
7.5.5 Controlling Translational Motion: Molecular Shuttles *361*
7.5.6 Controlling the Motion: Ring Pirouetting in Rotaxanes *376*
7.5.7 Controlling the Motion: Switchable Catenanes – the Issue of Directionality *377*
7.6 From Laboratory to Technology: Toward Useful Molecular Machines *383*
7.6.1 Current Challenges *383*
7.6.2 Reporting Motion: Switches and Memories *383*
7.6.3 The Interface with Real-world Technology *393*
7.7 Summary and Outlook *395*
References *397*

8 Replicable Nanoscaffolded Multifunctionality – A Chemical Perspective *407*
W.-M. Pankau, S. Antsypovich, L. Eckardt, J. Stankiewicz, S. Mönninghoff, J. Zimmermann, M. Radeva, G. von Kiedrowski
Abstract *407*
8.1 Introduction *407*
8.2 A Manifesto for Nanorobot Implementation *408*
8.2.1 Noncovalent Informational Nanoscaffolding *409*
8.2.2 Self-assembly From Synthetic Three-arm Junctions *409*
8.2.3 Tensegrity and Maximum Instruction as the Keys for Nanoarchitecture Control *411*
8.2.4 Chemical Copying of Connectivity Information (CCC) as the Key for Nanomachine Replication *412*
8.2.5 Cloning and Copying on Surfaces Using eSPREAD *413*
8.2.6 Linear Conjugates as Building Blocks for Junction eSPREADing and Nanoscaffolded Multifunctionality *414*
8.2.7 Directed Evolution of Replicable Nanomachines *416*
8.2.8 Probing the Existence of Nanoepitopes on the Surface of Biological Cells *417*
8.2.9 External Control of the Operation of Such Nanomachines by GHz Radio-frequency Magnetic Field Inductive Heating of Metal Clusters Attached to Such Constructs *417*
8.3 Conclusion *419*
Acknowledgment *419*
References *420*

Index *421*

Preface

Although numerous books appeared on the general topic of Supramolecular Chemistry, the combination of an overview about specific receptor structures for important compound classes with a broad description of their potential functions and applications is not found today. Either the supramolecular principles are emphasized (noncovalent interactions, enthalpy-entropy calculations, selectivity etc.) or the (often narrow) range of applications is detailed from a certain view point, e.g. in the field of sensors, electrochemical devices etc. In addition, several reviews or books which were already written on certain chapters of this book, appeared in the 90's, so that a fresh compilation of recent advances in the field during the past 4-5 years seems necessary.

This book comprises a timely overview about receptor molecules for the most important classes of compounds with a concentration on recent literature (1998-2004). Special emphasis is placed on potential applications. As a consequence, a strong interdisciplinary touch reaches out to other fields like Materials Science (devices), Bioorganic Chemistry (enzyme mechanisms, model compounds), Medicinal Chemistry (prevention of pathological processes with synthetic receptors) and Organic Synthesis (catalysts, mechanistic elucidations), to name just a few. Especially in the biological context challenging solvents are preferred: which receptor systems work well in water or even better under physiological conditions? Any overlap with other books in the neighbourhood of supramolecular chemistry is avoided, unless the latest review is already more than 5-6 years old.

The level of presentation aims at the advanced readership, i. e., graduate students and specialists in the field. Introductory remarks are restricted to a minimum. The topics dealt within will be interesting for pharmaceutical companies dealing with drug design as well as chemical companies with a polymer branch or a nanotechnology group. Firms working in the field of molecular biology and biotechnology will also benefit from the chapters with biologic content. Finally analytical companies working with or producing the advanced analytical equipment mentioned in this book will find interesting new fields of applications of their technologies.

The editors have asked leading authors in their fields to summarize the most important developments of the recent past, and place their emphasis on the most fascinating and promising applications of these new receptor molecules. Unfortunately, this compilation must be a (subjective) selection, leaving out many excellent contributions – we apologize to all those at the outset.

We are indebted to all authors for their invaluable contributions to this book. Dr. Gudrun Walter and Dr. Rainer Münz of Wiley-VCH did a great job in shaping the concept and assembling the individual items into a complete opus.

Marburg and New Haven, *Thomas Schrader and Andrew D. Hamilton*
January 2005

List of Contributors

Sergey Antsypovich
Chair of Organic Chemistry I
Ruhr-University Bochum
Universitätsstraße 150
44780 Bochum (Germany)

Won-Seob Cho
Department of Chemistry
and Biochemistry
Institute for Cellular and
Molecular Biology
1 University Station – A5300
The University of Texas at Austin
Austin, TX 78712-0165 (USA)

Anthony P. Davis
School of Chemistry
University of Bristol
Cantock´s Close
Bristol BS8 1TS (UK)

Lars Eckardt
Chair of Organic Chemistry I
Ruhr-University Bochum NC2/133
Universitätsstraße 150
44780 Bochum (Germany)

Andrew D. Hamilton
Department of Chemistry
Yale University
225 Prospect Street
P.O. Box 208 107
New Haven, CT 06520 (USA)

Tony D. James
Department of Chemistry
University of Bath
Bath BA2 7AY (UK)

Euan R. Kay
School of Chemistry
University of Edinburgh
The King’s Buildings
West Mains Road
Edinburgh EH9 3JJ (UK)

Günter von Kiedrowski
Chair of Organic Chemistry I
Ruhr-University Bochum NC2/171
Universitätsstraße 150
44780 Bochum (Germany)

David A. Leigh
School of Chemistry
University of Edinburgh
The King’s Buildings
West Mains Road
Edinburgh EH9 3JJ (UK)

Michael Maue
Department of Chemistry
Marburg University
Hans-Meerwein-Straße
35032 Marburg (Germany)

Sven Mönninghoff
Chair of Organic Chemistry I
Ruhr-University Bochum NC2/132
Universitätsstraße 150
44780 Bochum (Germany)

Wolf-Matthias Pankau
Chair of Organic Chemistry I
Ruhr-University Bochum NC2/132
Universitätsstraße 150
44780 Bochum (Germany)

Leonard J. Prins
Department of Chemical Sciences
University of Padova
Via F. Marzolo 1
35131 Padova (Italy)

Maya Radeva
Chair of Organic Chemistry I
Ruhr-University Bochum NC2/126
Universitätsstraße 150
44780 Bochum (Germany)

Dmitry M. Rudkevich
Department of Chemistry
and Biochemistry
University of Texas at Arlington
Arlington, TX 76019-0065 (USA)

Thomas Schrader
Department of Chemistry
Marburg University
Hans-Meerwein-Straße
35032 Marburg (Germany)

Paolo Scrimin
Department of Chemical Sciences
University of Padova
Via Marzolo 1
35131 Padova (Italy)

Jonathan L. Sessler
Department of Chemistry
and Biochemistry
Institute for Cellular and
Molecular Biology
1 University Station – A5300
The University of Texas at Austin
Austin, TX 78712-0165 (USA)

K. Ingrid Sprinz
Department of Chemistry
Yale University
PO Box 208107
New Haven, CT 06520 (USA)

Johanna Stankiewicz
Chair of Organic Chemistry I
Ruhr-University Bochum NC2/164
Universitätsstraße 150
44780 Bochum (Germany)

Debarati M. Tagore
Department of Chemistry
Yale University
PO Box 208107
New Haven, CT 06520 (USA)

Jan Zimmermann
Chair of Organic Chemistry I
Ruhr-University Bochum NC2/133
Universitätsstraße 150
44780 Bochum (Germany)

1
Artificial (Pseudo)peptides for Molecular Recognition and Catalysis

Leonard J. Prins and Paolo Scrimin

1.1
Introduction

This chapter focuses on recognition and catalytic processes in which artificial (pseudo)peptide sequences, which can be very short, play a decisive role. The enormous amount of literature related to this topic is far beyond the scope of a single chapter, and, therefore, we intend to emphasize concepts and breakthroughs by using representative examples. Obviously, the reason for the interest in the role of (pseudo)peptides in molecular recognition and catalysis is the fact that polypeptides, e.g. proteins, play a crucial role in practically all biologically relevant processes. An incredible number of recognition events is of key importance for the occurrence of life. The origin of biological recognition is the tertiary structure of proteins, which is marvelously determined by conformationally well-defined secondary structures such as α-helices, β-sheets, coiled coils, etc. These locally structured units give order to the overall system, positioning functional groups precisely in three-dimensional space, thus creating an active site where recognition takes place. Molecular recognition is especially crucial in the functioning of enzymes. To accomplish its powerful tasks an enzyme first needs to recognize the substrate and, subsequently, in the course of its chemical transformation, also the intermediate transition state that lies on the reaction pathway toward the product. These impressive results in Nature form an almost infinite source of inspiration for the chemist, not only to mimic natural functions but also to modify them and apply them in unnatural situations.

In this chapter we will discuss advances that have been made in our learning process from Nature and, more specifically, show how chemists are able to mimic natural functions using artificial synthetic molecules. First, we will focus on the biomolecular recognition of oligonucleotides (DNA/RNA) and protein surfaces by artificial oligopeptides. Next, we will show that chemists have learned to control the secondary structure of (pseudo)peptides and that specific (catalytic) functions can be introduced at will. Finally, we will conclude with a brief overview of the selection of (pseudo)peptide catalysts by a combinatorial approach.

1.2
Recognition of Biological Targets by Pseudo-peptides

1.2.1
Introduction

In all organisms, nucleic acids are responsible for the storage and transfer of genetic information. With the aim of curing gene-originated diseases, artificial molecules that can interact with DNA and RNA are of utmost interest. In this section we will discuss the current state of two major classes of pseudo-peptides that are currently under intense investigation – polyamides that bind in the minor groove of DNA and peptide nucleic acids (PNA). Both classes of compounds are inspired by naturally occurring analogs. The high synthetic accessibility and the ease with which chemical functionality can be introduced illustrate the high potential of artificial pseudo-peptides. In addition, their high biostability has enabled successful applications in both *in-vitro* and *in-vivo* studies. The limiting properties of these compounds will also be addressed.

Another way of interfering with biological processes is to obstruct the activity of proteins themselves. Pseudo-peptides that inhibit the formation of protein–protein complexes via competitive binding to the dimerization interface will be discussed. Selected examples will be given that clearly illustrate the strong increase in activity when amino acids present in a wild-type peptide sequence are replaced by artificial amino acids.

1.2.2
Polyamides as Sequence-specific DNA-minor-groove Binders

The discovery of the mode of interaction between the natural compounds distamycin and netropsin and the minor groove of DNA has been the impetus for the development of a set of chemical rules that determine how the minor groove of DNA can be addressed sequence-specifically [1]. NMR and X-ray spectroscopy showed that distamycin binds to A,T-tracts 4 to 5 base pairs in length either in a 1:1 or 2:1 fashion, depending on the concentration (Fig. 1.1) [2, 3]. It was then immediately realized by the groups of Dickerson, Lown, and Dervan that the minor groove of DNA is chemically addressable and, importantly, that chemical modifications of the natural compounds should, in theory, provide an entry to complementary molecules for each desirable sequence [4, 5].

1.2.2.1 Pairing Rules

The minor groove of DNA is chemically characterized by several properties. First, the specific positions of hydrogen-bond donor and acceptor sites on each Watson–Crick base pair, as depicted schematically in Fig. 1.2. Next, the molecular shape of the minor groove in terms of specific steric size, such as the exocyclic NH_2 guanine. Finally, an important property is the curvature of the double stranded DNA helix. Having

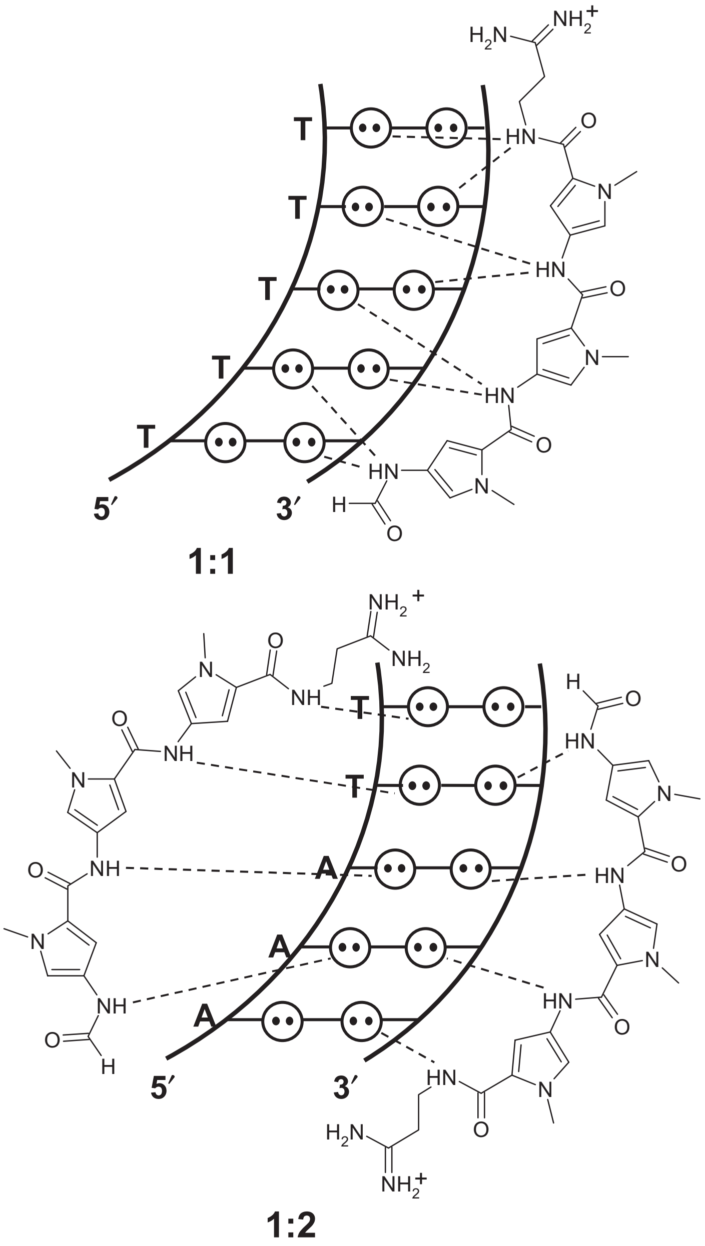
H_2N
NH_2^+
T
T
T
T
T
HN
O
N
HN
O
N
HN
O
N
HN
H
O
5′
3′
1:1
NH_2^+
NH_2
O
N
NH
O
NH
N
O
NH
N
NH
O
H
T
T
A
A
A
5′
3′
H
O
HN
N
HN
O
N
HN
O
HN
N
O
H_2N
NH_2^+
1:2

Figure 1.1 Observed binding modes of distamycin to DNA (1:1 and 1:2 complexes). The dotted circles represent the lone pairs of N(3) of purines and O(2) of pyrimidines in the minor groove. The dotted lines represent hydrogen bonds between distamycin and DNA.

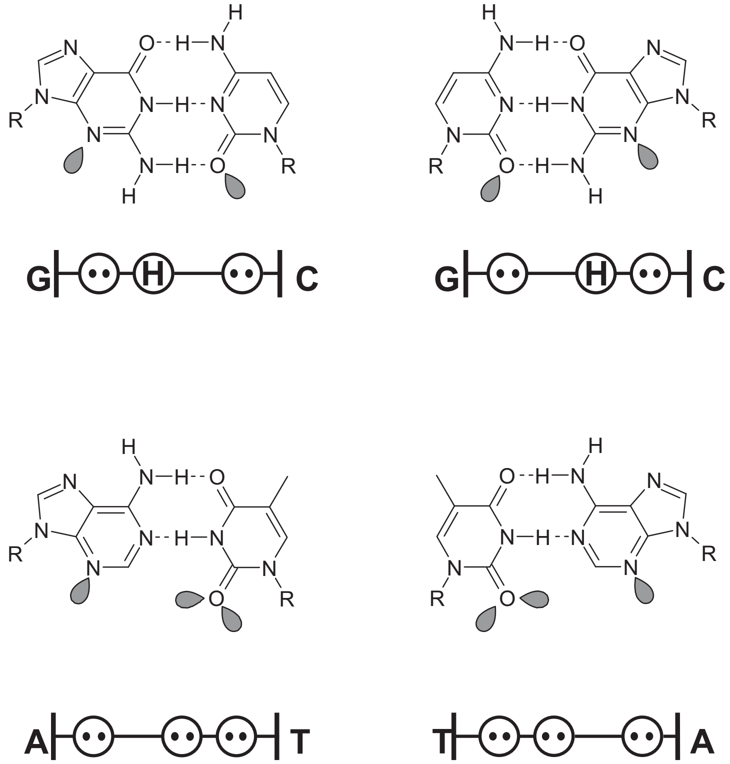

Figure 1.2 Hydrogen-bond donors and acceptors present in the minor groove of ds DNA for each of the four Watson–Crick base pairs. Lone pairs are indicated by shaded orbitals and R represents the sugar–phosphate backbone. In the schematic representations circles with dots are hydrogen-bond acceptors and circles with H are hydrogen-bond donors.

these properties as a guideline, Dervan and coworkers have developed a series of five-membered heterocycles that pairwise can recognize each of the four base pairs [6, 7]. These couples and their binding modes are schematically depicted in Fig. 1.3. To gain selectivity for a G,C over an A,T base pair, the pyrrole ring (Py) was substituted by an imidazole (Im), which forms an additional hydrogen-bond with the exocyclic NH_2 of guanine, as confirmed by crystal structure analysis. In addition, replacement of the pyrrole CH for an N eliminates the steric clash of pyrrole and the exocyclic NH_2 of guanine. The presence of an additional hydrogen-bond acceptor on thymine residues stimulated the synthesis of the *N*-methyl-3-hydroxypyrrole (Hp) monomer, which contains an additional hydrogen-bond donor. Also, in this case, the complementary molecular shape between the cleft imposed by the thymine-O2 and the adenine-C2 and the bumpy –OH are important.

The selective binding to T,A over A,T base pairs (and, similarly, G,C over C,G) originates from the *antiparallel* binding of *two* polyamide strands in the minor groove of DNA. A key NMR spectroscopy study confirmed that an ImPyPy polyamide bound

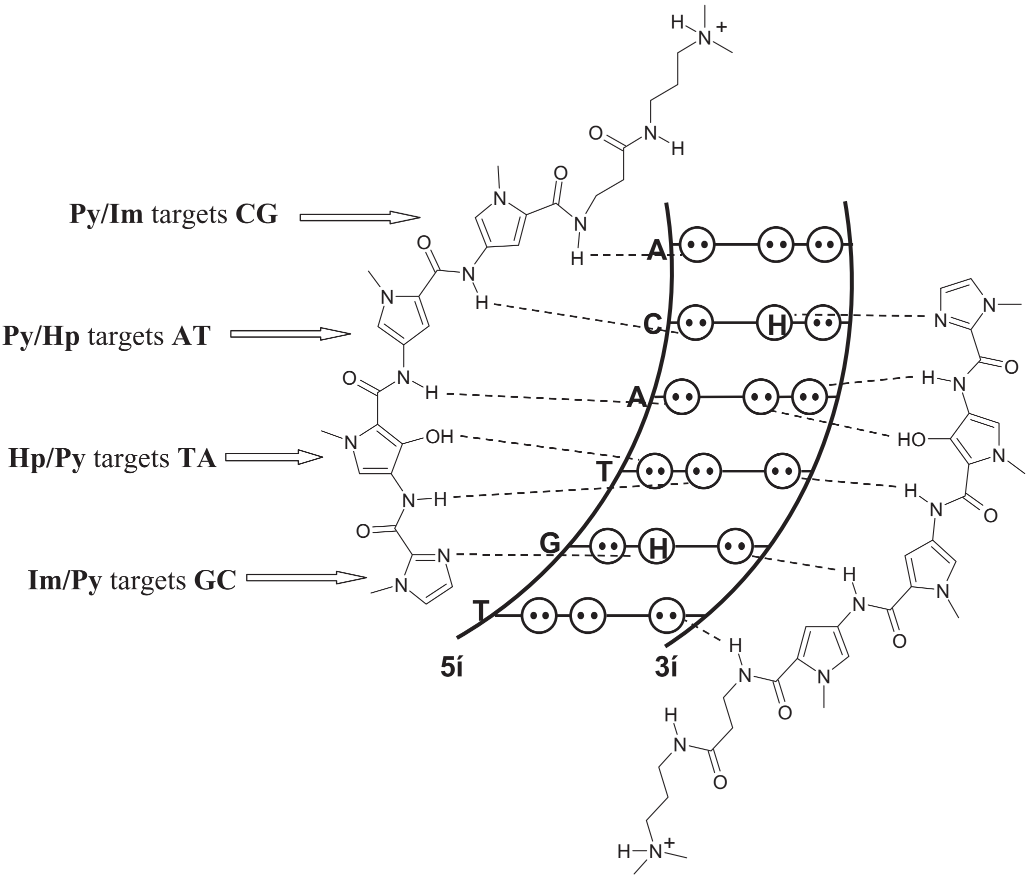

Figure 1.3 Pairing rules for dsDNA recognition by polyamides. Py: pyrrole, Im: imidazole, Hp: hydroxypyrrole.

antiparallel in a 2:1 fashion to a 5’-WGWCW-3’ sequence (W = A or T) with the polyamide oriented N→C with respect to the 5’→3’ direction of the adjacent DNA strand.

Recently, the repertoire of the heterocycles used (Py, Im, and Hp) has been expanded to novel structures based on pyrazole, thiophene, and furan, to increase binding specificity and stability (the Hp monomer has limited stability in the presence of free acid or radicals) [8]. In addition, benzimidazole-based monomers (Ip and Hz) were incorporated in polyamides as alternatives for the dimeric subunits PyIm and PyHp, respectively [9, 10]. DNase I footprinting revealed functionally similar behavior with regard to the parent compounds containing exclusively Py, Im, and Hp monomers. An important advantage is the chemical robustness of the benzimidazole monomer Hz relative to Hp.

1.2.2.2 Binding Affinity and Selectivity

The ternary complex composed of two three-ring structures, such as distamycin, and DNA is rather modest, for entropic reasons and because of the low number of hydrogen bonds involved. In an important step forward towards artificial DNA binders that can effectively compete with DNA-binding proteins, the carboxyl and amino termini of two polyamide chains were covalently connected via a γ-aminobutyric acid linker (Fig. 1.4a) [11]. A so-called hairpin polyamide composed of eight heterocycles was shown to bind to the complementary six-base-pair DNA sequence with an affinity constant of the order of 10^{10} M^{-1}. A single base-pair mismatch site induced a 10–100-fold drop in affinity. Importantly, the N→C orientation with respect to the 5'→3' direction of DNA is generally retained for these compounds. An additional tenfold increase in affinity was observed for a cyclic polyamide in which the two strands were covalently connected at both termini (Fig. 1.4b) [12]. The γ-turn has a preference for an A,T over a G,C base pair, presumably because of a steric clash between the aliphatic turn and the exocyclic amine of guanine. New polyamide structures that are covalently bridged via the heterocycle nitrogen atoms, either at the center (H-pin, Fig. 1.4c) or terminus (U-pin, Fig. 1.4d) have recently been prepared [13, 14]. The U-pins resulted in a loss in affinity, because of the removal of two hydrogen bond donors, but were insensitive to the base pair adjacent to the turn. Cleverly, the H-pin polyamides were synthesized on a solid support using the Ru-catalyzed alkene metathesis reaction to connect the different polyamides. This approach enabled the rapid synthesis and screening of a series of polyamides with alkyl bridges differing in size ($(CH_2)_n$, with n ranging from 4 to 8); the optimum affinity and specificity was obtained for $n = 6$.

In the gigabase-sized DNA database it is desirable to address sequences of 10–16 base pairs, because these occur much less frequently. Increasing the number of heterocycles in polyamides increases the sequence size that can be targeted, but only up to a certain limit. Studies revealed that the binding affinity is maximized at a contiguous ring number of 5. For longer systems affinity drops because the different curvatures of polyamides and B-DNA starts to give energetically strongly unfavorable interactions. These problems can be partially overcome by replacing one (or more) of the pyrrole units by a more flexible β-alanine unit. In this way polyamides have been prepared that bind sequences as long as 11 base pairs with subnanomolar affinities [15]. Alternatively, two hairpin polyamides have been covalently connected either turn-to-turn or turn-to-tail and were shown to bind ten-base-pair sequences with impressive affinities in the order of 10^{12} M^{-1} [16, 17]. It should be noted, however, that these high affinities come with rather low selectivity.

The potential of these molecules in controlling gene expression is extremely important; here are examples that illustrate this point. Because excellent reviews have appeared that cover in great detail all results obtained, we will limit ourselves to recent examples that illustrate well the different concepts.

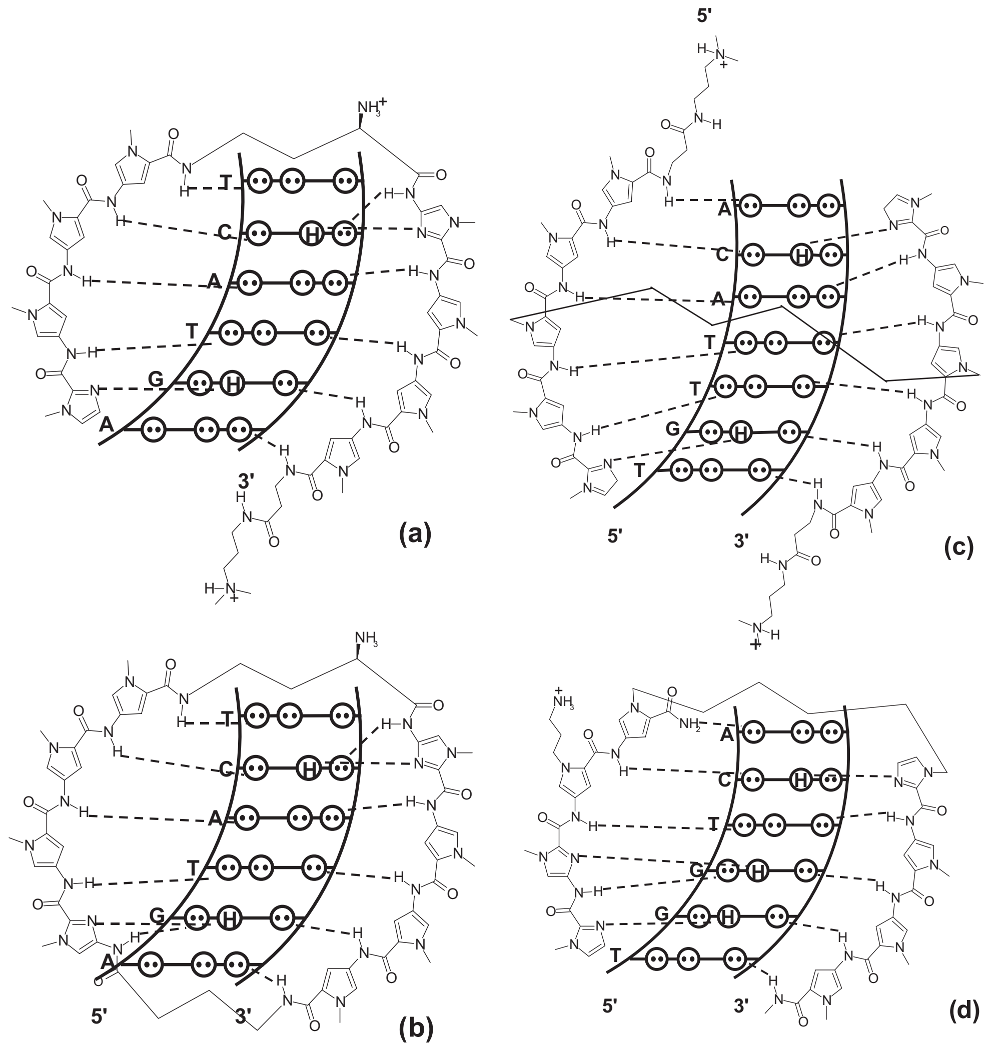

Figure 1.4 Strategies in polyamide design: (a) hairpin ($K_a \approx 10^{11}$ M^{-1}); (b) cycle ($K_a \approx 10^{11}$ M^{-1}); (c) H-pin ($K_a \approx 10^{10}$ M^{-1}); (d) U-pin ($K_a \approx 10^{9}$ M^{-1}).

1.2.2.3 **DNA Detection**

The ability to detect double stranded DNA sequences, *and* single base-pair mismatches, is an extremely useful tool in the field of genetics. Most methods involve hybridization of single-stranded DNA by a complementary oligonucleotide probe, which carries a signaling moiety. These techniques, however, require denaturation of DNA. On the other hand, double-stranded DNA can be detected by dyes such as ethidium bromide and thiazole orange, but binding is unspecific, making these dyes useful solely as quantitative tools for DNA detection. Dervan and coworkers prepared

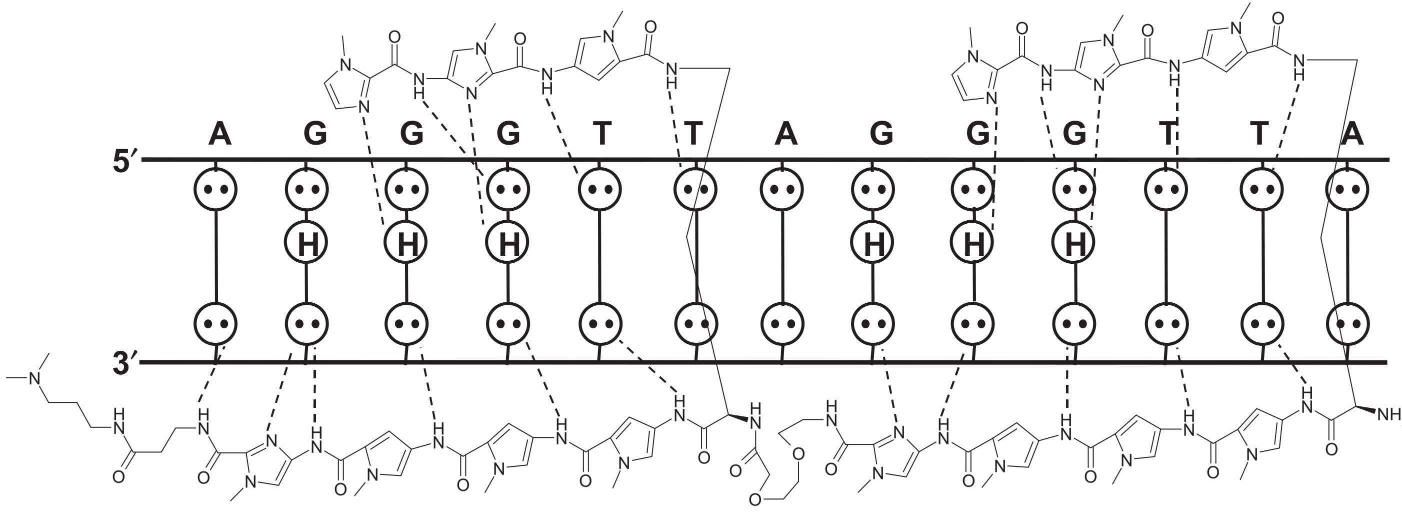
A G G G T T A G G G T T A
5′
3′
NH2

Figure 1.5 Tandem polyamide designed to target the repetitive sequence TTAGG present in insect telomers.

a series of eight-ring hairpin polyamides with tetramethyl rhodamine (TMR) attached to internal pyrrole rings and studied the fluorescence in the presence and absence of 17-mer duplex DNA [18]. In the absence of DNA, the fluorescence of the conjugates was strongly diminished compared with that of the free dye. This was hypothesized to result from the short linker separating the polyamide and the dye, which enables nonradiative decay of the excited state. The addition of increasing amounts of duplex DNA with a match sequence resulted in an increase in fluorescence until 1:1 DNA:conjugate stoichiometry was reached. Binding of the polyamide fragment to the minor groove of DNA results in forced spacing between polyamide and dye, thus diminishing any quenching effect.

In an impressive study by Laemmli and coworkers the ability of polyamide–dye conjugates to function in a genomic context was demonstrated [19]. A series of tandem polyamides was synthesized that interact specifically with two consecutive insect-type telomeric repeat sequences (TTAGG) (Fig. 1.5). The dissociation constant for the best polyamide was 0.5 nM, as determined by DNase I footprinting. Epifluorescence microscopy studies using Texas Red-conjugated analogs of these polyamides showed a very strong staining of both insect and vertebrate telomeres of chromosomes and nuclei. Convincingly, the telomere-specific polyamide signals of HeLa chromosomes colocalize with the immunofluorescence signals of the telomere-binding protein TRF1. Studies in live Sf9 cells seem to suggest rapid uptake of the conjugates, thus enlarging the potential of these compounds as human medicine. These results should be interpreted with caution, however, because the fluorescence studies were performed after fixation of the cells, which is known to dramatically increase the membrane permeability of cells [20].

In a related approach, Trask et al. targeted the TTCCA motif repeated in the heterochromatic regions of human chromosomes 9, Y, and 1, using polyamides tagged with fluorescein [21]. Staining of the targeted regions was similar to that with the conventional technique (FISH), which employs hybridization of fluorescent complementary sequences. In sharp contrast, however, polyamide–dye conjugates do not require denaturation of the chromosomes.

1.2.2.4 **Gene Inhibition**

Gene expression requires recruitment of the transcription machinery to the promoter region, after which transcription of the coding region into mRNA can start. Inhibition of this process by polyamides can occur in either the promoter or coding region of a gene. The latter is more difficult to achieve, because any molecule noncovalently bound to the double helix will be expelled by RNA polymerases during the transcription of DNA. To address this issue polyamides have been tagged with alkylating agents such as chlorambucil and *seco*-CBI [22]. Indeed it was observed that alkylation occurs specifically at base pairs flanking the binding site of the polyamide. Whether this strategy enables effective inhibition of RNA polymerases has not yet been reported.

Most attention has been paid toward polyamides that act in the promoter region of a gene as competitors for the binding of transcription factors to DNA. Inhibition by polyamides can occur for a variety of reasons. A minor-groove-binding protein can be inhibited by a minor-groove polyamide because of steric hindrance. Binding of a protein with major-groove/minor-groove contacts can be similarly inhibited when the polyamide is crucially located in the minor groove. Alternatively, polyamides can also function as allosteric effectors that rigidify the shape of B-DNA, thus competing with a major-groove binding protein that requires helical distortion. Examples of protein–DNA complexes that have been inhibited by polyamides are TBP, LEF-1, Ets-1, and Zif268. In a key study, the viral HIV-1 gene was targeted (Fig. 1.6) [23]. The HIV-1 enhancer/promoter region contains binding sites for multiple transcription factors, among them Ets-1, TBP, and LEF-1. Two different polyamides were designed to target DNA sequences immediately adjacent to the binding sites of these transcription factors. Cell-free assays showed that these ligands specifically inhibited binding of the transcription factors to DNA and consequently repressed HIV-1 transcription. In isolated human peripheral blood cells, incubation with a combination of these two polyamides resulted in 99% inhibition of viral replication, with no obvious decrease in cell viability. RNase protection assays indicated that the transcript levels of some other genes were not affected, suggesting that the polyamides indeed affect transcription directly.

However, despite the success in inhibiting binding of a large variety of proteins to DNA, problems remain with the class of major-groove-binding proteins that are not affected by the presence of ligands in the minor groove. Recent studies have been aimed at a generic solution that would inhibit binding of any sort of transcription factor [24]. Very promisingly, it was observed that attachment of an acridine intercalator to a polyamide locally extended and unwound the double helix and thus acted as an allosteric inhibitor for the major-groove binding of the GCN4 bZip protein.

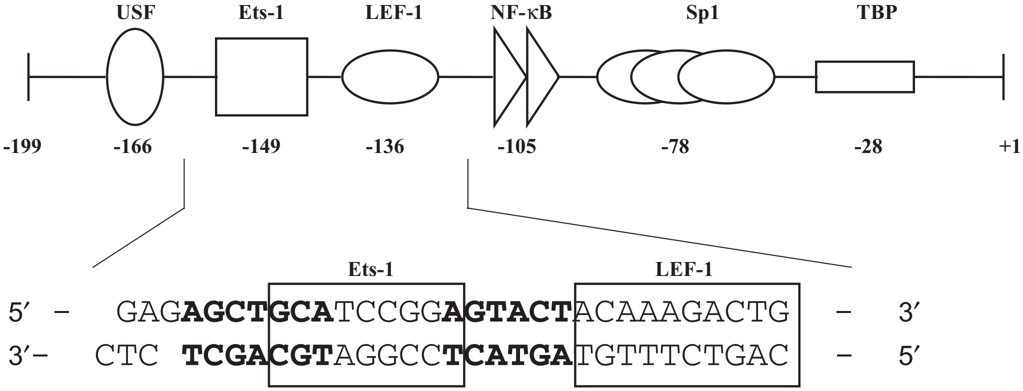
USF
Ets-1
LEF-1
NF-κB
Sp1
TBP
-199
-166
-149
-136
-105
-78
-28
+1
Ets-1
LEF-1
5′ – GAGAGCTGCATCCGGAGTACTACAAAGACTG – 3′
3′– CTC TCGACGTAGGCCTCATGATGTTTCTGAC – 5′

Figure 1.6 Schematic representation of the enhancer/promoter-region of the HIV-1 gene, with indications of the binding sites of the various transcription factors. In the enlargement, the binding sites of Ets-1 and LEF-1 are shown together with the binding sites of the polyamides (bold).

1.2.2.5 **Gene Activation**

Gene activation can occur either by inhibiting the binding of a repressor protein or via recruitment of the transcription machinery. The first method is conceptually identical to the examples given in the previous section. For instance, polyamides have been successfully applied as an upregulator for transcription of the human cytomegalovirus MIEP [25]. The second method requires an entirely different role of the polyamide. Eukaryotic transcription factors are minimally composed of a DNA-binding and an activation domain. Most activator proteins also contain a dimerization element. Generally, gene transcription starts with binding of a transcription factor to the promoter-region, which induces recruitment of a series of other transcription factors to nearby promoter sites, and finally transcription is initiated. Dervan and coworkers showed that it is possible to replace an activator protein by an artificial minimum system composed only of short oligopeptides (Fig. 1.7) [26, 27]. In the mimic of the Gal4 yeast activator the DNA-binding domain was replaced with an eight-ring hairpin polyamide and the activation domain with VP2, a 16-amino-acid residue oligopeptide derived from the viral activator VP16. The two modules were linked with an eight-atom spacer giving a polyamide–peptide conjugate with a size not exceeding 3.2 kDa. Cell-free activation assays revealed an upregulation of almost 15 fold at concentrations that caused full occupancy of the binding sites. In a recent study, the influence of the linker on activation activity was investigated by examining a series of polyamide–peptide conjugates linked via rigid oligoproline sequences varying in size between 18 and 45 Å [28]. Optimum activity was observed for a Pro_{12} linker, about 36 Å in length.

The Hox proteins belong to a family of transcriptional regulators that bear the "home domain" – a trihelical DNA-binding domain that is conserved across vast evo-

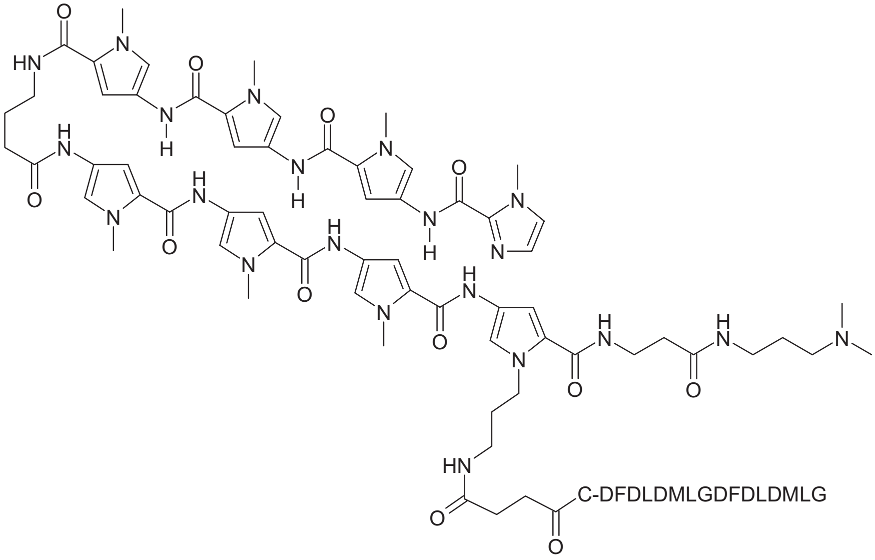

Figure 1.7 A polyamide-based artificial transcription factor for Gal4.

lutionary distances. Generally, Hox proteins by themselves bind DNA with low affinity and selectivity, but recently it has been suggested that heterodimer formation with the TALE class of homeoprotein causes high affinity and selectivity. Interaction between a *Drosophila* Hox (Ubx) and TALE (Exd) protein occurs via a short YPWM peptide. This conserved peptide is a feature of all Hox proteins. A polyamide was conjugated to this small oligopeptide to serve as a Hox-mimic [29]. Electrophoretic mobility shift assays revealed positive cooperative interactions between Exd and the conjugate in the formation of a ternary complex with DNA. It was, in fact, shown that the artificial activator was more effective than its naturally occurring equivalent, clearly illustrating the strong potential of this class of conjugates.

1.2.2.6 **Future Perspective**

The examples discussed in this section demonstrate the enormous achievements made in using polyamides, and their large potential as medicine for genetic diseases. However, some important hurdles still have to be passed to reach this ultimate goal. The main problem, as for many man-made synthetic structures, is the problem of delivery of the compounds to where they are needed – in the cellular nuclei in an organism. Evolution has decided, and for good reasons, that cells should not be a social room in which everybody can enter and leave as desired. Most polyamide studies have been performed on cell-free assays, clearly showing that the principles work. Also some encouraging results have been obtained in a cellular context, most notably inhibition of HIV-1 expression in human blood lymphocytes and also the induction of specific gain- and loss-of-function phenotypes in *Drosophila* embryos [30]. Recent studies, however, have revealed that polyamide uptake is largely cell-dependent and that in many cell lines polyamides are excluded from the nucleus [20, 31]. The next research phase for these compounds will reveal whether a generic solution to the cellular uptake problem is feasible, or if each target will require different chemical modifications of the polyamides.

1.2.3
Peptide Nucleic Acids

The second large class of (pseudo)peptides renowned for their ability to interact specifically with nucleic acids (both RNA and DNA) are the peptide nucleic acids (PNA). This research area also has been covered extensively in reviews [32–35] and we will therefore limit ourselves to illustrative and, where possible, recent examples.

1.2.3.1 **Chemistry and Interaction with DNA/RNA**

PNA is a DNA structural mimic in which the DNA backbone is replaced by a (pseudo)peptide, to which the nucleobases are connected by methylenecarbonyl linkages (Fig. 1.8) [36]. The structural simplicity and high accessibility of PNA has stimulated many scientists to study its properties in DNA/RNA recognition and to devise alternative structures [37, 38]. The strong binding properties of PNA to complementary ss DNA and RNA strands are best illustrated by the much higher melting points of the PNA–DNA and PNA–RNA dimers compared with the corresponding ds DNA and

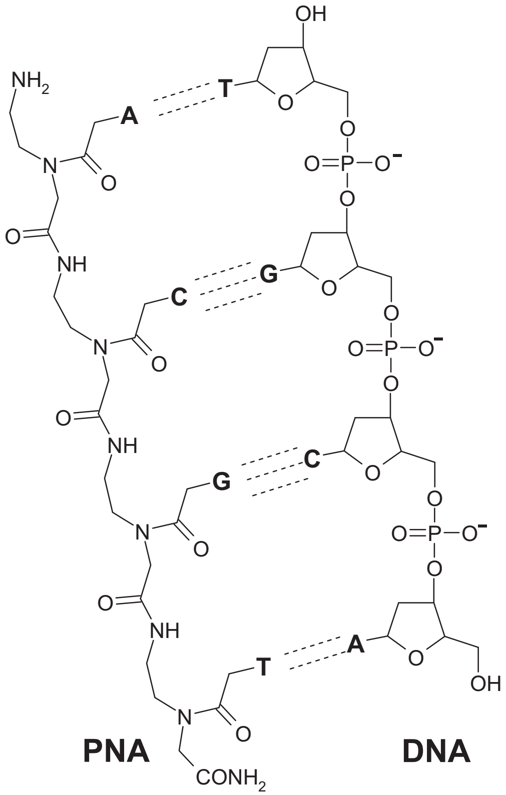

Figure 1.8. Chemical structures of PNA and DNA.

DNA–RNA duplexes (69.5 and 72.3 compared with 53.3 and 50.6 °C, respectively) [39]. In addition, the PNA–DNA dimer proved much less tolerant of single base pair mismatches than the native DNA dimer. The increased stability of the PNA-containing dimers is primarily ascribed to the lack of repulsive electrostatic interactions between the two strands.

Originally designed in the early 1990s to serve as the third strand in triple helix DNA, the original *N*-(2-aminoethylglycine) PNA turned out to behave differently than expected. Addition of a 10mer homothymine PNA to complementary ds DNA resulted in helix invasion rather than triple-helix formation (Fig. 1.9) [36, 40]. In the resulting complex one PNA strand binds via Watson–Crick base-pairing antiparallel to the complementary DNA strands. A second PNA binds in a parallel fashion via Hoogsteen base-pairing. Triplex invasion is limited to homopurine tracks, similar to the behavior observed for triple-helix-forming oligonucleotides. Interestingly, the PNA_2–DNA complexes are kinetically very stable with half-lives in the order of hours [41]. The stability is highly dependent on the ionic strength of the solution, however, with even physiologically relevant levels having a detrimental effect on binding. On the other hand, binding is greatly facilitated in cases where DNA is (tran-

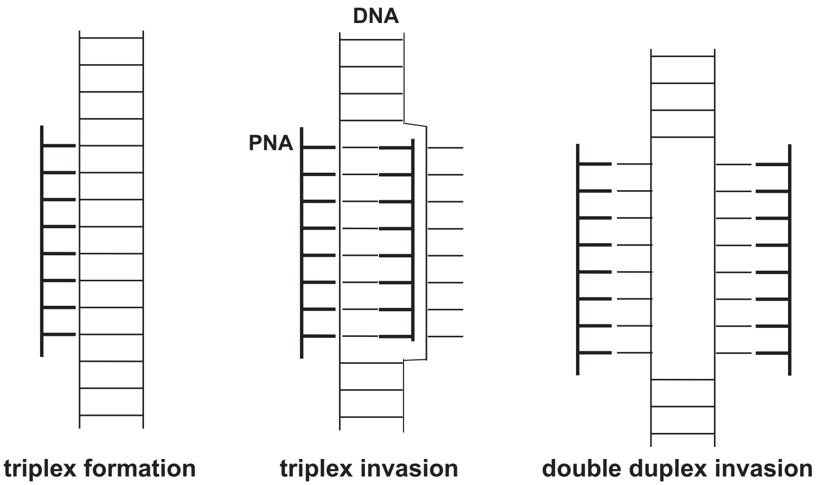

Figure 1.9 Different modes of PNA-interaction with dsDNA. PNA strands are depicted in bold.

siently) unwound, such as negative DNA supercoiling or by a passing RNA polymerase [42, 43]. Other factors beneficial for complex stability are a covalent connection of the two PNA strands and replacement of the cytosines in the Hoogsteen strand for pseudoisocytosines, which do not require a low pH for protonation of the N3 position. Alternatively, it was recently shown that conjugation of a DNA-intercalator such as 9-aminoacridine to a PNA significantly increases binding affinity and enables helix invasion even at physiologically relevant ionic strength [44].

Also, sequences other than homopurine tracks can be targeted when pseudo-complementary PNA are used. Pseudo-complementarity means that the AT nucleobases in the PNA strands are replaced by a 2,6-diaminopurine/2-thiouracil pair [45]. This artificial base-pair is very unstable, because of steric hindrance. Ds DNA recognition in this case occurs via so-called double-duplex invasion, in which each PNA binds to the complementary DNA strand, thus relieving the steric repulsion present in the PNA duplex (Fig. 1.9).

Other attractive features of PNA that increase its potential as a therapeutic agent include a high chemical and enzymatic stability and low toxicity. A limiting factor, on the other hand, is the poor cellular (nuclear) delivery of this class of compounds [46]. These issues will be addressed in more detail in the selective examples discussed below.

1.2.3.2 **PNA as a Regulator of Gene Expression**

In-vitro studies showed that formation of the triple helix complex $(PNA)_2$–DNA can inhibit gene transcription by RNA polymerase [47]. Using an *in-vitro* replication run-off assay under physiological conditions, 14-mer PNA inhibited replication of mitochondrial DNA mutant templates by more than 80%, whereas no inhibition of wild-type template replication was observed. From a medicinal perspective this is an important observation, because numerous genetic diseases result from single nucleotide polymorphism (SNP). Similarly, binding of either a 10-mer homothymidine PNA or a 15-mer PNA with mixed sequence to a plasmid (pBSA10) containing a complementary target sequence caused 90 to 100 % site-specific termination of pol II transcription elongation [48].

Interestingly, transcription activation using PNA has also been achieved via two conceptually different approaches. Triplex invasion causes the expelled DNA strand to form a stable D-loop structure at the DNA-binding site, which is known to induce transcription. A series of C,T-PNA with different lengths (8mer – 20mer) was designed to bind to a 20 base-pair DNA homopurine sequence cloned into two promoter reporter vectors – pGL3-Basic, which carries a promoterless luciferase gene, and pEGFP-1, which carries a promoterless green fluorescent protein (GFP) gene [49]. In this setup, fluorescence readout is a direct reporter of the extent of transcription activation. Gel-mobility shift assays showed the strongest binding for the 16mer PNA, even at 1×10^{-7} M concentrations. Transcription activation was performed both in HeLa nuclear extracts and in human NF cells. Transfection of the cells in the *in-vivo* experiment was performed by preloading the plasmid with PNA and subsequent transfection using cationic liposomes. Both experiments resulted in high expression levels of the fluorescent protein, with an optimum for the 16- and 18-mer PNA.

The second approach toward gene activation is conceptually identical to the strategy employed with the minor-groove-binding polyamides already discussed [50]. A synthetic activator was formed by connecting a twenty residue Gal80-binding peptide sequence to a bis-PNA, i.e. two homopyrimidine PNA "clamped" via a poly(ethylene glycol) linker. The ability of this artificial activator to recruit Gal80 was assessed by binding of the PNA–DNA complex to Gal80 attached to agarose beads. Large amounts of the complex were retained for the DNA duplex containing five potential PNA binding sites. DNA retention was practically absent in a series of control experiments. The results of an *in-vitro* study on HeLa nuclear extracts were recently reported [51]. It was shown that promoter-targeted PNA alone acts as a strong inhibitor of basal transcription, for reasons already discussed, but that conjugation of the Gal80-binding peptide reactivates transcription.

Finally, in a very exciting and controversial study Richelson and coworkers have claimed that antigene and antisense PNA can pass the blood–brain barrier and can bind *in-vivo* to the neurotensin receptor (NTR1) in rats [52, 53]. Specifically, intraperitoneal injection of PNA inhibited the hypothermic and antinociceptive activities of neurotensin microinjected in the brain. Controversy has arisen because of the lack of evidence that the unmodified PNA used can indeed enter cells, not to mention nuclei. A second, more fundamental, problem lies in the design of the PNA used – all current information known about this class of compound suggests that it is unlikely that short hetero purine/pyrimidine PNA sequences can effectively bind to their complementary DNA strands.

1.2.3.3 **Antisense Properties of PNA**

Principally, oligodeoxynucleotides (ODN) and analogs can inhibit translation of mRNA in two ways. The first is activation of ribonuclease H (RNase H) toward cleaving of the mRNA strand in the ODN–RNA complex. Although this property has been observed for several ODN analogs, for example phosphorothioates, PNA–RNA complexes do not seem to be substrates for RNase H. It has, however, been shown that antisense properties of PNA can arise from the second mode, in which either formation of the initial ribosomal translation machinery is inhibited or ribosome elongation in the coding region of mRNA is arrested. The efficiency depends on the type of complex formed [54]. Triplex-forming 10mer PNA were effective both when the AUG start codon was targeted, but also when bound to a downstream sequence in the coding region [55]. Duplex-forming PNA, however, seems effective in inhibiting formation of the translation machinery only; except in some special cases it is unable to stop translation once started. This difference is ascribed to the higher stability of the $(PNA)_2/RNA$ triplex and its tighter binding. Presumably, the "looser" binding of the duplex at the termini makes it easier for the elongating ribosome to expel it. Similar behavior has also been observed for minor-groove-binding polyamides. Also processes involving RNA, other than translation, have seemed to be sensitive to interactions with PNA. Examples include reverse transcription, telomerase activity, and RNA splicing.

Although most studies have been performed under cell-free *in-vitro* conditions, results *in vivo* have also been obtained. Most notably, Kole and coworkers reported on the antisense activity of PNA in a transgenic mouse containing the gene EGFP-654 encoding for the enhanced green fluorescent protein (EGFP) [56]. This gene was, however, interrupt by an aberrantly spliced mutated intron of the human β-globin gene. Consequently, in this model EGFP is only expressed in tissues in which an antisense oligomer (such as PNA) has restored the correct splicing. PNA was administered for one to four days via intraperitoneal injection, after which various organs were examined for the presence of EGFP. High activity was observed in the kidney, liver, and small intestine. Importantly, PNA activity was only observed for PNA that had four lysines connected to the C-terminus. A PNA containing only one lysine residue was completely inactive.

1.2.3.4 **Future Perspective**

Minor-groove binding polyamides and peptide nucleic acids are impressive examples of the progress made starting from a chemical design towards biologically active compounds. *In-vitro* studies have revealed the chemical robustness (and limitations) of these compounds as antigene and antisense agents, and it is striking that both classes are now more or less at the same stage of development, with the hurdle of cellular uptake preventing a wide applicability *in vivo*. In both instances, isolated examples have illustrated their potential *in vivo*, but a general solution has not yet been found. It should be remembered, however, that development of these classes of compound was started no more than approximately 15 years ago. The impressive progress made in this relatively short time and their wide use throughout the (bio)chemical community should evoke optimism for their future use as gene medicine.