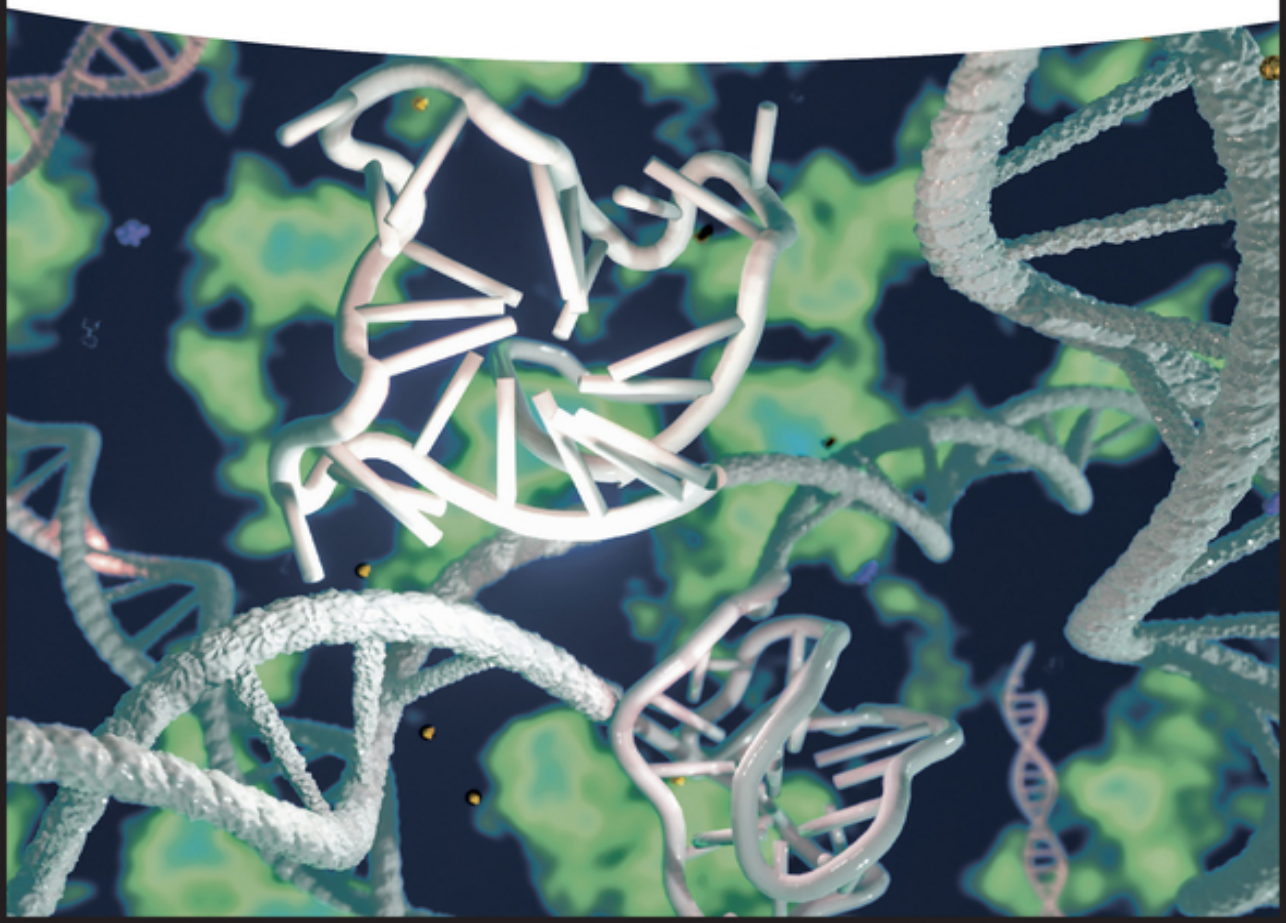


WILEY-VCH

Naoki Sugimoto

Chemistry and Biology of Non-Canonical Nucleic Acids



Chemistry and Biology of Non-Canonical Nucleic Acids

Chemistry and Biology of Non-Canonical Nucleic Acids

Naoki Sugimoto

WILEY-VCH

Author

Prof. Naoki Sugimoto

Konan University
Frontier Institute for Biomolecular Engineering
7-1-20
Minatojima-minamimachi
650-0047 Kobe
Japan

■ 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 Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <<http://dnb.d-nb.de>>.

© 2021 WILEY-VCH GmbH, Boschstr. 12, 69469
Weinheim, Germany

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a 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.

Print ISBN: 978-3-527-34521-2

ePDF ISBN: 978-3-527-81788-7

ePub ISBN: 978-3-527-81786-3

oBook ISBN: 978-3-527-81785-6

Typesetting SPi Global, Chennai, India

Printing and Binding

Printed on acid-free paper

10 9 8 7 6 5 4 3 2 1

Contents

Preface *xi*

- 1 History for Canonical and Non-canonical Structures of Nucleic Acids 1**
 - 1.1 Introduction 1
 - 1.2 History of Duplex 1
 - 1.3 Non-Watson–Crick Base Pair 5
 - 1.4 Nucleic Acid Structures Including Non-Watson–Crick Base Pairs 7
 - 1.5 Perspective of the Research for Non-canonical Nucleic Acid Structures 8
 - 1.6 Conclusion and Perspective 9
 - References 9

- 2 Structures of Nucleic Acids Now 11**
 - 2.1 Introduction 11
 - 2.2 Unusual Base Pairs in a Duplex 11
 - 2.2.1 Hoogsteen Base Pair 13
 - 2.2.2 Purine–Pyrimidine Mismatches 13
 - 2.2.3 Purine–Purine Mismatches 14
 - 2.2.4 Pyrimidine–Pyrimidine Mismatches 16
 - 2.3 Non-canonical Backbone Shapes in DNA Duplex 17
 - 2.4 Branched DNA with Junction 19
 - 2.5 Multi-stranded DNA Helices 20
 - 2.6 Structures in RNA 20
 - 2.6.1 Basic Structure Distinctions of RNA 20
 - 2.6.2 Elements in RNA Secondary Structures 21
 - 2.6.2.1 Hairpin Loop 22
 - 2.6.2.2 Bulge Loop 22
 - 2.6.2.3 Internal Loop 23
 - 2.6.3 Elements in Tertiary Interactions of RNA 24
 - 2.6.3.1 A-Minor Interactions 25
 - 2.6.3.2 Ribose Zipper 25
 - 2.6.3.3 T-Loop Motif 26

2.6.3.4	Kissing-Loop Interaction	26
2.6.3.5	GNRA Tetraloop Receptor Interaction	27
2.6.3.6	Pseudoknot Crosslinking Distant Stem Regions	29
2.7	Conclusion	29
	References	30
3	Stability of Non-canonical Nucleic Acids	33
3.1	Introduction	33
3.2	Factors Influencing Stabilities of the Canonical Duplexes	34
3.2.1	Hydrogen Bond Formations	34
3.2.2	Stacking Interactions	35
3.2.3	Conformational Entropy	35
3.3	Thermodynamic Analysis for the Formation of Duplex	36
3.4	Factors Influencing Stabilities of the Non-canonical Nucleic Acids	39
3.4.1	Factors Influencing Stability of Triplexes	39
3.4.2	Factors Influencing Stability of Quadruplex	42
3.4.2.1	G-Quadruplexes	42
3.4.2.2	i-Motif	44
3.5	Thermodynamic Analysis for the Non-canonical Nucleic Acids	45
3.5.1	Thermodynamic Analysis for the Intramolecular Triplex and Tetraplex	45
3.5.2	Thermodynamic Analysis for the Intermolecular Triplex	45
3.5.3	Thermodynamic Analysis for the Tetraplex	46
3.6	Conclusion	48
	References	49
4	Physicochemical Properties of Non-canonical Nucleic Acids	51
4.1	Introduction	51
4.2	Spectroscopic Properties of Non-canonical Nucleic Acids	51
4.2.1	Effect of Non-canonical Structure on UV Absorption	51
4.2.2	Circular Dichroism of Non-canonical Nucleic Acids	53
4.2.3	NMR Spectroscopy	56
4.2.4	Other Spectroscopic Characteristics of Non-canonical Nucleic Acids	57
4.3	Chemical Interactions on Non-canonical Nucleic Acids	59
4.3.1	Hydration	59
4.3.2	Cation Binding	61
4.3.3	pH Effect	62
4.3.4	Chemical Modification	63
4.4	Chemical Platform on the Non-canonical Structures	64
4.4.1	Specificity of a Ligand to Non-canonical Structures	64
4.4.2	Fluorescence Platform of Non-canonical Structures	67
4.4.3	Interface Between Proteins and Nucleic Acids	68
4.5	Physicochemical Property of Non-canonical Nucleic Acids in Cell	69
4.5.1	Molecular Crowding Condition that Reflects Cellular Environments	69

4.5.2	Effects of Crowding Reagents on Non-canonical Nucleic Acid Structures	70
4.5.3	Quantification of Physical Properties of Non-canonical Structures in Crowding Condition	71
4.5.4	Non-canonical Structures Under Mimicking Organelle Environment	72
4.5.5	Insight for the Formation of Non-canonical Nucleic Acids in Cells	73
4.6	Conclusion	75
	References	75
5	Telomere	79
5.1	Introduction	79
5.2	Structural Properties of Telomere	79
5.2.1	Structures of Telomere	79
5.2.2	Structural Properties of Human Telomeric G4s	81
5.2.3	Structure of Repeats of Human Telomeric G4s	84
5.3	Biological Relevance of Telomere G4	86
5.3.1	Telomerase Activity	86
5.3.2	Telomerase Repeated Amplification Protocol (TRAP) Assay	89
5.3.3	Alternative Lengthening of Telomere (ALT) Mechanism	89
5.4	Other Non-canonical Structures Related to Telomere Region	89
5.4.1	Telomere i-Motif	89
5.4.2	Telomere RNA	90
5.5	Conclusion	92
	References	93
6	Transcription	95
6.1	Introduction	95
6.2	Transcription Process	96
6.2.1	Transcription Initiation	96
6.2.2	Transcription Elongation	98
6.2.3	Transcription Termination	99
6.3	Transcription Process Perturbed by Certain Sequences of DNA and RNA	101
6.4	Transcription Process Perturbed by Non-canonical Structures of DNA and RNA	103
6.5	Conclusion	110
	References	110
7	Translation	113
7.1	Introduction	113
7.2	RNAs Involved in Translation Machinery	113
7.3	General Process of Translation	117
7.3.1	Translation Initiation	117
7.3.2	Translation Elongation	119
7.3.3	Translation Termination	119

7.4	RNA Structures Affecting Translation Reaction	121
7.4.1	Modulation of Translation Initiation in Prokaryotes	121
7.4.2	Modulation of Translation Initiation in Eukaryotes	123
7.4.3	RNA Structures Affecting Translation Elongation	126
7.4.4	RNA Structures Affecting Translation Termination	130
7.5	Conclusion	133
	References	134
8	Replication	137
8.1	Introduction	137
8.2	Replication Machineries	137
8.3	Replication Initiation	138
8.3.1	Mechanism of Activation of Replication Origins	138
8.3.2	Activation Control of Origins by G4s	139
8.3.3	Control of Timing of Replication Initiation by G4s	142
8.4	DNA Strand Elongation	142
8.4.1	Mechanism of DNA Strand Elongation	142
8.4.2	Impact of G4 and i-Motif Formations on DNA Strand Synthesis	144
8.4.3	Relationship Between G4 and Epigenetic Modification	145
8.4.4	Expansion and Contraction of Replicating Strand Induced by Hairpin Structures	147
8.5	Termination of Replication	148
8.6	Chemistry of the Replication and Its Regulation	148
8.6.1	Cellular Environments	148
8.6.2	Control of Replication by Chemical Compounds	150
8.7	Conclusion	151
	References	152
9	Helicase	155
9.1	Introduction	155
9.2	Function and Structure of Helicases	155
9.3	Unwinding of Non-canonical DNA Structures by Helicases	158
9.4	G4 Helicases in Gene Expressions	162
9.5	G4 Helicases in Replication	163
9.6	G4 Helicases in Telomere Maintenance	164
9.7	Relation to Diseases by Loss of G4 Helicases	165
9.8	Insight into Specific Properties of Activities of G4 Helicase Under Cellular Conditions	165
9.9	Conclusion	167
	References	167
10	Dynamic Regulation of Biosystems by Nucleic Acids with Non-canonical Structures	171
10.1	Introduction	171
10.2	Time Scale of Biological Reactions	171
10.2.1	Cell Cycle	172
10.2.2	Central Dogma	172

10.2.3	Dynamic Structures of Nucleic Acids	174
10.3	Processes in the Central Dogma Affected by Dynamics of Nucleic Acid Structures	176
10.3.1	Epigenetic Regulation Caused by Chemical Modification of DNA	176
10.3.2	Co-transcriptional Formation of Metastable RNA Structures	178
10.3.3	Co-transcriptional Translation and Transcription Attenuation	180
10.3.4	Co-transcriptional Ligand Binding and Gene Regulation	181
10.3.5	Translation Elongation and Co-translational Protein Folding	183
10.4	Conclusion	184
	References	185
11	Cancer and Nucleic Acid Structures	189
11.1	Introduction	189
11.2	Detail Mechanism of Cancer	189
11.2.1	Cancer Incidence	189
11.2.2	The Relationship Between Genes and Cancer	192
11.3	Non-canonical Structures of Nucleic Acids in Cancer Cells	192
11.3.1	Structural Characteristics of Nucleic Acids in Cancer Cells	192
11.3.2	Non-canonical Structures Perturb Gene Expression of Cancer-Related Genes	195
11.4	Roles of Non-canonical Structures of Nucleic Acids in Cancer Cells	197
11.4.1	Monitoring of Non-canonical Structures in Cancer Cells	197
11.4.2	Regulation of Gene Expressions by the Non-canonical Structures in Cancer Cells	198
11.5	Conclusion	199
	References	199
12	Neurodegenerative Diseases and Nucleic Acid Structures	203
12.1	Introduction	203
12.2	Protein Aggregation-Induced Neurodegenerative Diseases	203
12.3	DNA Shows Key Role for Neurodegenerative Diseases	205
12.4	RNA Toxic Plays a Key Role for Neurological Diseases	210
12.5	Conclusion	212
	References	212
13	Therapeutic Applications	215
13.1	Introduction	215
13.2	Oligonucleotide Therapeutics	216
13.2.1	Antisense Oligonucleotide	216
13.2.2	Functions of Antisense Oligonucleotide Therapeutics	217
13.2.3	Chemical Modifications in Therapeutic Oligonucleotides	220
13.2.3.1	Backbone Modified Oligonucleotides	220
13.2.3.2	Ribose Modified Oligonucleotides	221
13.2.3.3	Oligonucleotides with Unnatural Backbone	221
13.2.4	Oligonucleotide Therapeutics Other Than Antisense Oligonucleotide	223

13.2.4.1	Oligonucleotide Therapeutics Functioning Through RNA Interference	224
13.2.4.2	Oligonucleotide Therapeutics Functioning Through Binding to Protein	224
13.3	Non-canonical Nucleic Acid Structures as Therapeutic Targets	224
13.3.1	Traditional Antibiotics Targeting Structured Region of RNAs	225
13.3.2	Strategies for Constructing Therapeutic Materials Targeting Structured Nucleic Acids	226
13.4	Non-canonical Nucleic Acid Materials for Inducing Non-canonical Structures	230
13.5	Conclusion	231
	References	232
14	Materials Science and Nanotechnology of Nucleic Acids	235
14.1	Introduction	235
14.2	Non-canonical Structure-Based Nanomaterials Resembling Protein Functions	235
14.2.1	Aptamer	235
14.2.2	DNAzyme	238
14.2.3	Ion Channel	240
14.3	Protein Engineering Using G4-Binding Protein	240
14.4	Regulation of Gene Expression by G4-Inducing Materials	242
14.5	Environmental Sensing	246
14.5.1	Sensing Temperature in Cells	246
14.5.2	Sensing pH in Cells	248
14.5.3	Sensing K ⁺ Ion in Cells	248
14.5.4	Sensing Crowding Condition in Cells	249
14.6	Conclusion	250
	References	250
15	Future Outlook for Chemistry and Biology of Non-canonical Nucleic Acids	253
15.1	Introduction	253
15.2	Exploring Potential: Properties of Non-canonical Structures in Unusual Media	253
15.3	Systemizing Properties: Prediction of the Formation of Non-canonical Nucleic Acids Structures	259
15.4	Advancing Technology: Applications of Non-canonical Structures Taking Concurrent Reactions into Account	262
15.4.1	Co-transcriptional Dynamics of G-Quadruplex	263
15.4.2	Co-transcriptional Functionalization of Riboswitch-Like Sensor	263
15.4.3	Co-transcriptional RNA Capturing for Selection of Functional RNAs	266
15.5	Conclusion	267
	References	268
	Index	271

Preface

In chemistry and biology, one of the most important and interesting research subjects is nucleic acids: DNA and RNA. The nucleic acids consist of very simple materials: phosphate, sugar, and organic bases. Their structures are also very simple as single strands or a double helix, in comparison with another biomolecules such as proteins and carbohydrates; however the nucleic acids have very important genetic information and functions.

As I mentioned in the Introduction in Chapter 1, there is close to 70 years history in nucleic acid research after the discovery of the double helix DNA structure (*B-form*) as the canonical one by James Dewey Watson and Francis Harry Compton Crick in 1953 and chemical biology of nucleic acids are facing to new aspect today, that is, *non-canonical nucleic acids*. Through this book, I expect that readers understand how uncommon structure of nucleic acids became one of the common structures as *non-canonical nucleic acids* that fascinate us now. This new research field for non-canonical nucleic acids will soon big-spark at the interface of chemistry and biology.

This book is comprised of 15 chapters covering various aspects of chemistry and biology of *non-canonical nucleic acids* including not only their history, structures, stabilities, and properties but also their functions on transcription, translation, regulation, telomere, helicases, cancers, neurodegenerative diseases, therapeutic applications, nanotechnology, and future outlook. This book is a valuable resource, not only for graduate students but also researchers in the fields of physical chemistry, organic chemistry, inorganic chemistry, analytical chemistry, biochemistry, biophysics, structural biology, computational biology, molecular medicine, molecular biology, cell biology, and nanotechnology and who would like to learn more about the potential important roles of *non-canonical nucleic acids* as well as canonical ones.

I wish all readers enjoy this book and know the importance of not only Watson–Crick double helical nucleic acids (*B-form*) but also *non-canonical nucleic acids* like triplex and quadruplex. Instead of Hamlet by written by William Shakespeare, please answer the question “To B or not to B, that is the question” in the research field of nucleic acids.

I am deeply grateful to my colleagues in FIBER (Frontier Institute for Biomolecular Engineering Research), Konan University, for their excellent contribution to my

writing as the co-authors at the following each chapter. They are Dr. Shuntaro Takahashi (for Chapters 1, 4, 5, 8, 9, 14, and 15), Dr. Tamaki Endoh (for Chapters 2, 7, 10, 13, and 15), and Dr. Hisae Tateishi-Karimata (for Chapters 3, 6, 11, 12, and 15), whose efforts have immeasurably improved the quality and accuracy of the information. I am also deeply grateful to Ms. Miwa Inada for designing a lot of figures and Ms. Katherine Wong and Dr. Lifan Yang in Wiley for their editing this book and encouraging me.

Naoki Sugimoto
Frontier Institute for Biomolecular Engineering Research (FIBER)
Graduate School of Frontiers of Innovative Research
in Science and Technology (FIRST)
Konan University
Kobe, Japan

1

History for Canonical and Non-canonical Structures of Nucleic Acids

The main points of the learning:

Understand canonical and non-canonical structures of nucleic acids and think of historical scientists in the research field of nucleic acids.

1.1 Introduction

This book is to interpret the non-canonical structures and their stabilities of nucleic acids from the viewpoint of the chemistry and study their biological significances. There is more than 60 years' history after the discovery of the double helix DNA structure by James Dewey Watson and Francis Harry Compton Crick in 1953, and chemical biology of nucleic acids is facing a new aspect today. Through this book, I expect that readers understand how the uncommon structure of nucleic acids became one of the common structures that fascinate us now. In this chapter, I introduce the history of nucleic acid structures and the perspective of research for non-canonical nucleic acid structures (see also Chapter 15).

1.2 History of Duplex

The opening of the history of genetics was mainly done by three researchers. Charles Robert Darwin, who was a scientist of natural science, pioneered genetics. The proposition of genetic concept is indicated in his book *On the Origin of Species* published in 1859. He indicated the theory of biological evolution, which is the basic scientific hypothesis of natural diversity. In other words, he proposed biological evolution, which changed among individuals by adapting to the environment and be passed on to the next generation. However, that was still a primitive idea for the genetic concept. After that, Gregor Johann Mendel, who was a priest in Brno, Czech Republic, confirmed the mechanism of gene evolution by using "factor" inherited from parent to children using pea plant in 1865. This discovery became the concept of genetics. At the almost same time in 1869 as Mendel, Johannes

Friedrich Miescher, who was a biochemist in Swiss, discovered nucleic acids as a chemical substance of the gene identity. He named it “nuclein” (later, it was named “nucleic acid,” which exists acidic substance in nucleus) and made the opportunity to study nucleic acid chemistry. However, it would be doubtful if he realized that nucleic acid is the gene identity. After that, it was needed to take a lot of time to conclude that the gene identity is proved the chemical substance.



Charles Robert Darwin (left), Gregor Johann Mendel (middle), and Johannes Friedrich Miescher (right)

Erwin Rudolf Josef Alexander Schrödinger, who was a great physicist, pioneered to go after the mystery of gene. He published a book titled *What Is Life?* in 1944 [1]. This book invited the study of the gene to many researchers. He mentioned in the book that he believed a gene – or perhaps the whole chromosome fiber – to be an aperiodic solid, although he also mentioned that gene is probably one big protein molecule. After the 1950s, chemistry regarding nucleic acids had been developing. One of the organic chemists was Erwin Chargaff, who was a professor at Columbia University in the United States and born in Austria. He discovered that from the result of paper chromatography targeted to the different types of DNA, the number of guanine units equals the number of cytosine units and the number of adenine units equals the number of thymine units [2]. It is called Chargaff's rules. On the other hand, analysis of the superstructure of nucleic acids was also proceeding. At the beginning of the 1950s, at King's College London, the results of X-ray crystal analysis were accumulated by Maurice Hugh Frederick Wilkins, Rosalind Elsie Franklin, and others. Finally, based on their result, Watson and Crick who worked at Cavendish Laboratory in Cambridge and proposed the model of double helix structure of DNA (Figure 1.1 and see Chapter 2), published as a single-page paper about DNA double helix in *Nature* issued on 25 April 1953 [3]. By discovering DNA double helix structure, Watson, Crick, and Wilkins were awarded the Nobel Prize in Physiology or Medicine in 1962.

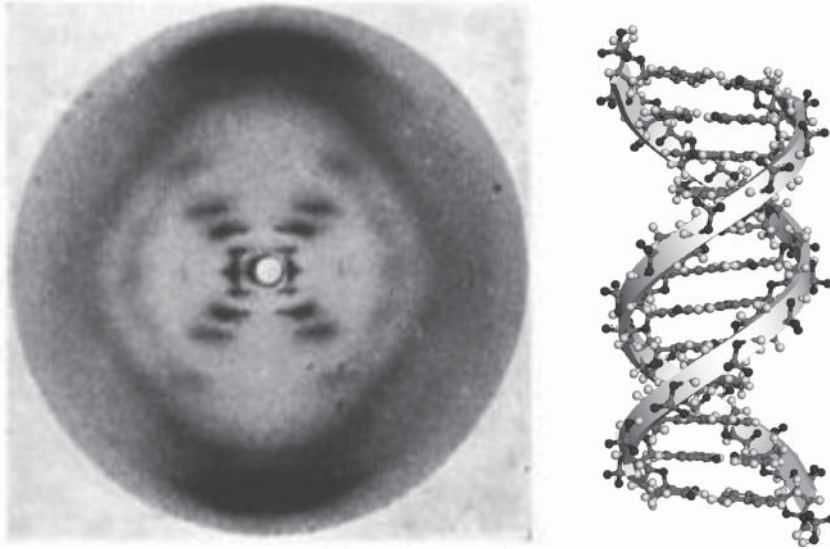


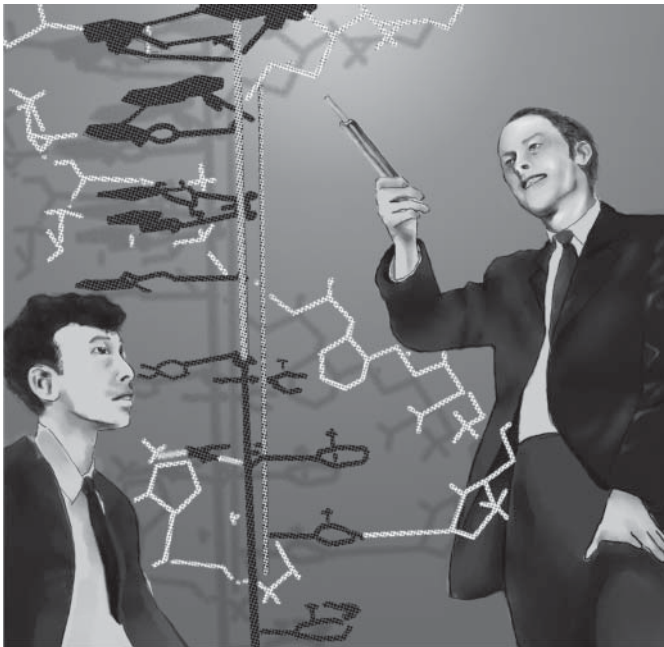
Figure 1.1 The diffraction pattern of the canonical DNA duplex and its chemical structure. Source: Kings College London.



Erwin Rudolf Josef Alexander Schrödinger (left) and Erwin Chargaff (right)



Maurice Hugh Frederick Wilkins (left) and Rosalind Elsie Franklin (right)



James Dewey Watson and Francis Harry Compton Crick

1.3 Non-Watson–Crick Base Pair

Although the discovery of Watson–Crick base pairs is famous, we need to make sure that Watson and Crick initially “proposed” their model. Moreover, Watson and Crick were not the first researchers who proposed the structure of nucleic acids. The physicist Linus Pauling, who earned the Nobel Prize two times in his career, first proposed the helix model of nucleic acids with his associate Robert Corey [4]. However, the structure was faulty: it was a triple helix having negatively charged phosphates located at the core of the helix, which could not exist in nature. After the proposal of Watson–Crick base pairs, the race for determination of the helical structure of DNA had been started using purine and pyrimidine monomers. The first such study was reported in 1959, when Karst Hoogsteen – an associate of Robert Corey at Caltech – used single-crystal X-ray analysis to determine the structures of cocrystals containing 9-methyladenine and 1-methylthymine, where methyl groups were used to block hydrogen bonding to nitrogen atoms otherwise bonded to sugar carbons in DNA [5]. However, the structure was NOT Watson–Crick base pair, in which the adenine base was flipped upside down. The different base pair was later named Hoogsteen base pair (Figure 1.2 and see Chapter 2). After the discovery of

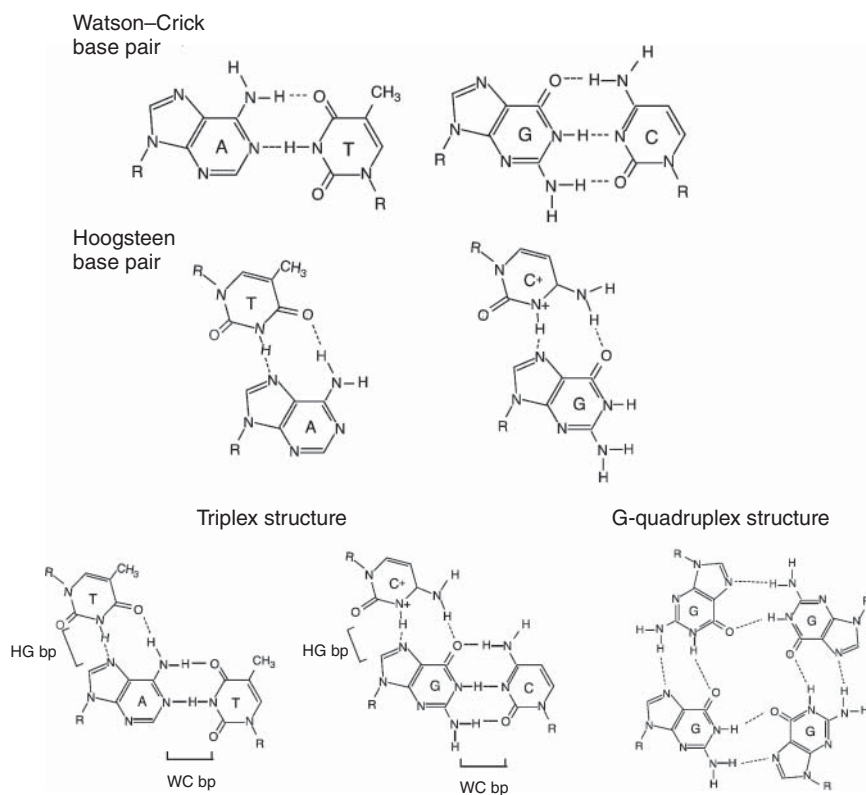


Figure 1.2 Chemical structures of base pairs via Watson–Crick or Hoogsteen types.

Hoogsteen base pairs, many researchers looked for Watson–Crick base pairs. However, only Hoogsteen base pairs were identified. In 1973, Alexander Rich first discovered Watson–Crick base pairs in the cocrystal of the AU and GC dinucleoside phosphate complex [6]. And soon after, Richard E. Dickerson, who took over the Pauling’s lab, first solved the single-crystal structure of a DNA dodecamer using heavy atom X-ray crystallography in 1980 [7]. It takes more than 20 years after the discovery of Watson–Crick base pairs. These results suggest that Watson–Crick base pairs tended to stably form under the constraint of the helical structure of nucleic acids, whereas Hoogsteen base pairs form in other structural conditions. Therefore, there are canonical structures composed by Watson–Crick base pairs in the duplex structures. On the other hand, non-canonical structures include non-Watson–Crick base pairs such as Hoogsteen base pairs.



Linus Pauling (left), Robert Corey (middle), and Karst Hoogsteen (right)



Alexander Rich (left) and Richard E. Dickerson (right)

1.4 Nucleic Acid Structures Including Non-Watson–Crick Base Pairs

Behind the extensive efforts to identify the duplex structure of Watson–Crick base pairs, Hoogsteen base pairs were also found in the structure of nucleic acids in the 1960s. Felsenfeld and Rich explained how poly(rU) strands might associate with poly(rA)-poly(rU) duplexes to form triplexes [8]. From the chemical shift of NMR, they identified evidence for triplex formation via protonated G–C⁺ Hoogsteen base pairs at cytosine N3 in a poly(dG)-poly(dC) complex with dGMP at low pH [9]. In 1962, it was found that short guanine-rich stretches of DNA could assume unusual structures [10]. The diffraction studies of poly(guanilyc acid) gels suggested that if four guanines were close enough together, they could form planar hydrogen-bonded arrangements now called guanine quartets (G-quartets). With a stack of a few G-quartets, a tetraplex structure is formed called as G-quadruplex (see Chapter 2). In the crystal structure, Hoogsteen base pairs of polynucleic acids were first found in tRNA structure [11]. In the structure Watson–Crick base pairs formed the secondary structure of tRNA, whereas Hoogsteen base pairs supported the tertiary structure. Not only Hoogsteen base pairs but also other types of non-Watson–Crick base pairs were found in tRNA structures. The tertiary structure of nucleic acids is important especially for non-coding RNAs, which do not code genetic information. The landmark of research of non-coding RNA is the discovery of ribozyme (ribonucleic acid enzyme) by Thomas Robert Cech in 1982 [12]. Ribozymes catalyze chemical reactions as well as protein enzymes. Later structural studies revealed that there are a lot of non-Watson–Crick base pairs to produce the active core of enzymatic reaction of ribozymes. Therefore, non-canonical Watson–Crick base pairs including Hoogsteen base pairs have been thought of as a tool for the tertiary structure of nucleic acids except for duplexes.



Thomas Robert Cech

With the progress of structural analysis technology in the 1990s, Hoogsteen base pairs are gradually revealed to exist in DNA complexes with low molecular weight compounds and proteins as well as transiently in Watson–Crick-type double helix. Furthermore, another type of tetraplex structures was identified from DNA sequence enriched in cytosine due to the cross intercalations of hemiprotonated cytosine–cytosine (C–C⁺) base pairs under acidic conditions [13]. This tetraplex is called as i-motif (see Chapter 2). Soon after, the roles of the non-canonical structures have been gaining attention. Especially since the 2000s, research on the G-quadruplex structure formed from Hoogsteen base pairs has made remarkable progress. When a G-quadruplex is formed on DNA or RNA, the reactivity of the protein involved in gene expression is affected (see Chapters 6–8). This means that the central dogma proposed by Crick – that genetic information is determined centrally by the flow of replication, transcription, and translation – is highly controlled by the formation of a G-quadruplex structure. In general, it has been thought that the regulation of gene information expression is due to protein functions. However, the specific structure of Hoogsteen base pairs controls gene expression so that the nucleic acid itself can function like a protein. That is, the roles of nucleic acids might be properly used according to base pairs: Watson Crick base pair = information, non-Watson Crick base pair = function. Many sequences that can have a G-quadruplex structure are distributed in the telomere at the end of the chromosome and the promoter region of the oncogene of the gene. Starting with the 2013 report, there have been many reports on the formation of G-quadruplexes and i-motifs in cells. These reports point out that the oncogene may be activated by the formation (or dissociation) of the G-quadruplex to cause cancer (see Chapter 11). Furthermore, it has been suggested that the phase-separated structure formed by the aggregation of RNAs with G-quadruplexes contributes to neurological diseases such as amyotrophic lateral sclerosis (see Chapter 12).

1.5 Perspective of the Research for Non-canonical Nucleic Acid Structures

As the regulation of gene expression by the specific structure of nucleic acids has been clarified, the next important issue is knowing what specific structures are formed where and when in cells. For example, Hoogsteen base pairs are affected by the molecular environments such as ions, pH, and water activity. Cells are in an environment crowded with molecules, so-called molecular crowding (see Chapters 3 and 4), and the molecular environment changes depending on the cell cycle [14]. For example, the nucleolus causes a change in the molecular density in the nucleus by repeating formation and dissociation according to the cell cycle. This regulates the timing of activation of rRNA transcription in each cell cycle, because the transcription of rRNA specifically occurs in nucleolus. In addition, the environment of mitochondria is particularly crowded (up to 500 mg ml⁻¹) but heterogeneous due to locally increased proton concentration by the proton gradient required for ATP synthesis. Therefore, it is desirable to develop a technology that can predict

physicochemical property of specific structures due to Hoogsteen base pairs in each characteristic molecular environment [15]. In addition, there is a possibility to make a new approach of drug development that treats diseases by changing the molecular environments of cells, rather than targeting genes and proteins.

1.6 Conclusion and Perspective

According to Pauling's personal communication revealed by the Nobel Foundation's disclosure, he considered Watson and Crick's Nobel award to be premature. In spite of his opinion, the Nobel Foundation decided to award the Prize to Watson and Crick. This might suggest that Watson–Crick base pairs were very common and meaningful at that time but non-Watson–Crick base pairs were artifact and meaningless. Nowadays, non-Watson–Crick base pairs are becoming common and significant as Pauling perhaps predicted. Now, the day when the essence of nucleic acids becomes beyond the concept of Watson and Crick is coming closer.

References

- 1 Schrödinger, E. (1944). *What Is Life? The Physical Aspect of the Living Cell and Mind*. Cambridge: Cambridge University Press.
- 2 Tamm, C., Hodes, M., and Chargaff, E. (1952). *J. Biol. Chem.* 195: 49–63.
- 3 Watson, J.D. and Crick, F.H. (1953). *Nature* 171: 737–738.
- 4 Pauling, L. and Corey, R.B. (1953). *Nature* 171: 346–346.
- 5 (a) Hoogsteen, K. (1959). *Acta Crystallogr.* 12: 822–823. (b) Hoogsteen, K.R. (1963). *Acta Crystallogr.* 16: 907–916.
- 6 (a) Day, R.O., Seeman, N.C., Rosenberg, J.M., and Rich, A. (1973). *Proc. Natl. Acad. Sci. U. S. A.* 70: 849–853. (b) Rosenberg, J.M., Seeman, N.C., Kim, J.J. et al. (1973). *Nature* 243: 150–154.
- 7 Wing, R., Drew, H., Takano, T. et al. (1980). *Nature* 287: 755–758.
- 8 Felsenfeld, G. and Rich, A. (1957). *Biochim. Biophys. Acta* 26: 457–468.
- 9 Kallenbach, N.R., Daniel, W.E. Jr., and Kaminker, M.A. (1976). *Biochemistry* 15: 1218–1224.
- 10 Gellert, M., Lipsett, M.N., and Davies, D.R. (1962). *Proc. Natl. Acad. Sci. U. S. A.* 48: 2013–2018.
- 11 Robertus, J.D., Ladner, J.E., Finch, J. et al. (1974). *Nature* 250: 546.
- 12 (a) Kruger, K., Grabowski, P.J., Zaug, A.J. et al. (1982). *Cell* 31: 147–157. (b) Zaug, A.J., Grabowski, P.J., and Cech, T.R. (1983). *Nature* 301: 578–583.
- 13 Gehring, K., Leroy, J.L., and Gueron, M. (1993). *Nature* 363: 561–565.
- 14 Nakano, S., Miyoshi, D., and Sugimoto, N. (2014). *Chem. Rev.* 114: 2733–2758.
- 15 (a) Takahashi, S. and Sugimoto, N. (2020). *Chem. Soc. Rev.* 49: 8439–8468. (b) Takahashi, S. and Sugimoto, N. (2021). *Acc. Chem. Res.* 54. In press.

2

Structures of Nucleic Acids Now

The main points of the learning:

- (1) Learn interactions in nucleic acid structures.
- (2) Understand structure polymorphisms of nucleic acids.
- (3) Study differences in conformational properties between DNA and RNA.

2.1 Introduction

Nucleic acids are basically molecules with a high degree of structural flexibility and polymorphic property. Phosphates in nucleic acids are negatively charged and cause electrostatic repulsion in each phosphate moiety. This electrostatic repulsion is disadvantageous for nucleic acids to form a compact and ordered structure. Nucleic acids form the higher-order structures by offsetting unfavorable entropy changes and electrostatic repulsion by internal interactions such as hydrogen bonding and stacking interactions and external factors such as interactions of nucleic acids with cations and cosolutes. In other words, the canonical nucleic acid structure consisting of double helix with Watson–Crick base pairs is a part of the possible structural forms, and nucleic acids form various non-canonical structures depending on the internal and external factors. This chapter shows basic elements that form non-canonical nucleic acid structures including unusual base pairing, whose existence has been revealed by structural analyses, and their properties of thermodynamic stabilities. Detailed analyses of the stabilities of nucleic acid structures and factors that affect them are explained in Chapter 3.

2.2 Unusual Base Pairs in a Duplex

As in Chapter 1, the canonical structure of nucleic acids is double-stranded helix with Watson–Crick base pairs (Figure 2.1), which are almost identical in their

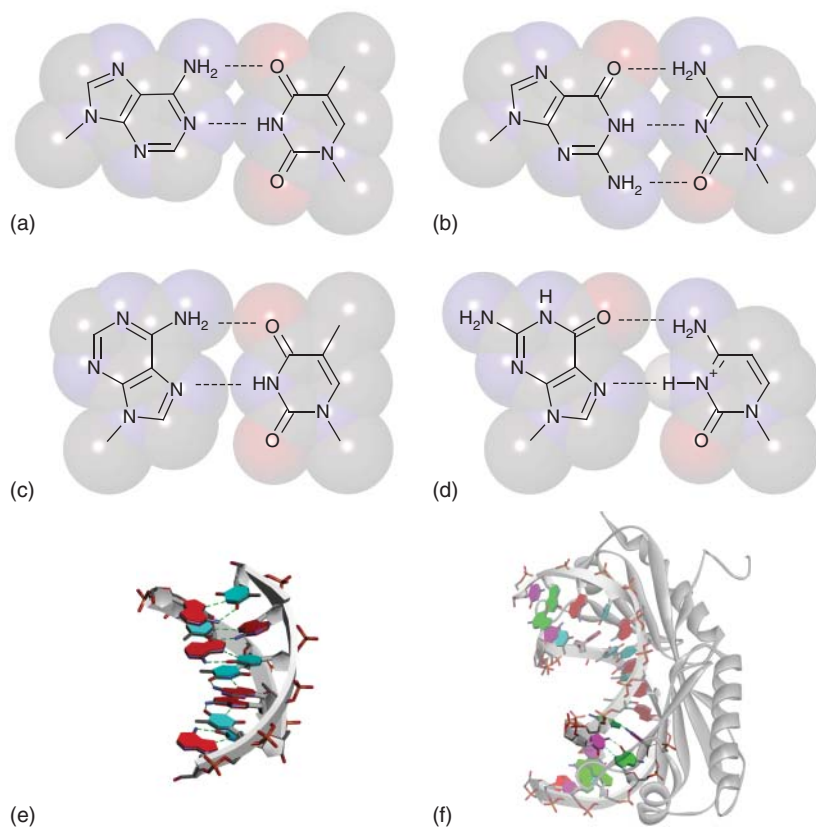


Figure 2.1 Watson–Crick and Hoogsteen base pairs in double helix. Chemical structures of A–T (a) and G–C (b) Watson–Crick base pairs. Chemical structures of A–T (c) and G–C⁺ (d) Hoogsteen base pairs. N3 atom of cytosine nucleobase is protonated. (e) Structure of DNA duplex consisting of all A–T Hoogsteen base pairs (PDBID: 1RSB). (f) Structure of DNA duplex containing two consecutive G–C⁺ Hoogsteen base pairs (PDB ID: 1QN3). The DNA duplex is bending due to interaction of TATA-box binding protein. Nucleobases forming the Hoogsteen base pairs are emphasized dark. In (e) and (f), hydrogen bonds between the Hoogsteen base pairs are shown in dashed lines.

geometric and dimensional arrangement in the helix. The sugar groups are both attached to the bases on the same side of the base pair. The distance between C1' atoms of the sugars on opposite strands is essentially the same. These geometric features enable any sequence of Watson–Crick base pairs fit into the duplex. On the other hand, a large number of other arrangements and hydrogen bonding patterns of base pairs are possible, and many have been observed experimentally such as using X-ray and NMR analyses. X-ray analysis can define the duplex structures, which incorporate the non-Watson–Crick base pairs and provide details of their hydrogen bonding scheme [1]. On the other hand, NMR analysis provides information regarding the dynamics of the nucleic acid conformation such as

mismatched base pairs, in which transient tautomeric and anionic species form Watson–Crick-type hydrogen bonds [2].

2.2.1 Hoogsteen Base Pair

Hoogsteen base pair (Figure 2.1) is one of the major non-Watson–Crick base pairs that can be seen in several crystal structures of duplexes containing A·T base pairs. For example, the crystal structure of AT-rich sequences adopted parallel and antiparallel stranded duplexes with all Hoogsteen-type hydrogen bonding in their A·T base pairs (Figure 2.1) [3]. In the case of antiparallel duplex with the Hoogsteen base pairs, the overall structure features of the duplex such as diameter of the duplex, number of base pairs per turn, and sugar pucker conformation are similar to the canonical B-type DNA duplex. A unique characteristic is that the adenine nucleobases in the duplex have *syn* conformation in their glycosidic bond angles. Although the same pattern of hydrogen bonding is possible, A-type RNA duplexes disfavor the A·U Hoogsteen base pair because the A-form geometry disfavors the *syn* conformation in the adenine nucleobase due to sugar-backbone rearrangements needed to sterically accommodate the adenine [4]. Formation of Hoogsteen-type hydrogen bonding is also possible between guanine (G) and cytosine (C⁺), in which N3 atom is protonated. Formation of transient Hoogsteen base pairs including the G·C⁺ in diverse sequence composition has been demonstrated by relaxation dispersion assay using NMR (Figure 2.1) [5]. It is considered that the Hoogsteen base pairs play roles in modulating interaction of proteins and biological reactions such as induction or repair of DNA damage and replication of DNA by altering the structural and chemical properties of the duplex [5].

2.2.2 Purine–Pyrimidine Mismatches

Mismatched base pairs between purine and pyrimidine nucleobases are known as transition mismatches. G·T and G·U mismatches can form two hydrogen bonds, which are usually known as typical “wobble” base pairs, by shifting the nucleobase geometry from that of Watson–Crick base pairs (Figure 2.2). These mismatches are one of the most stable mismatches in DNA and RNA duplexes (Tables 2.1 and 2.2). G·U base pair is frequently observed in natural RNAs. It is because that both guanine and uracil take *anti* conformation in their glycosidic bond angles that are the same with Watson–Crick base pairs and the distance and geometric arrangement of C1' atoms of the sugars do not largely change compared with the canonical duplex. Thus, there is no significant worsening of base stacking and littler perturbation of the duplex conformation. Adenine and cytosine also form two hydrogen bonds as similar way as the G·T(U) mismatch when the adenine nucleobase is protonated (Figure 2.2). Formation of the A⁺·C mismatched base pair is demonstrated by X-ray diffraction analysis using dodecamer oligonucleotide

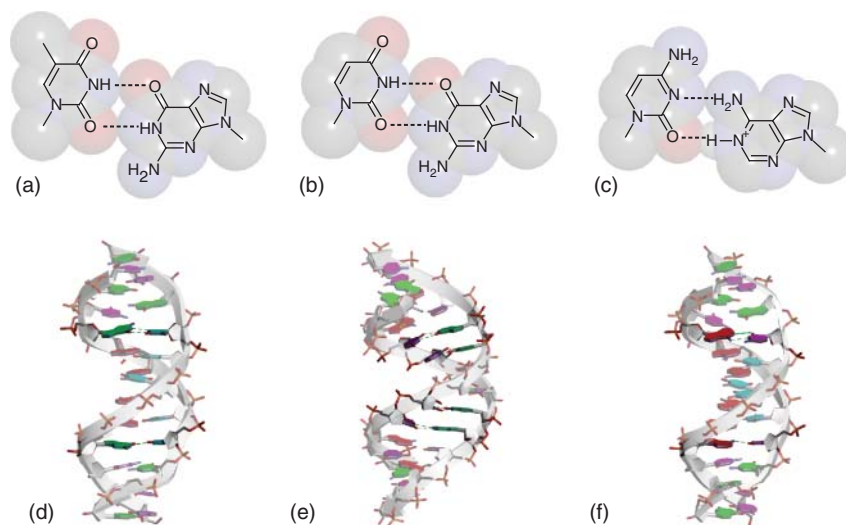


Figure 2.2 Wobble base pairs in duplexes. Chemical structures of G-T (a), G-U (b), and A⁺-C (c) wobble base pairs. N1 atom of adenine nucleobase is protonated. (d) Structure of B-form DNA duplex containing G-T wobble base pairs (PDB ID: 113D). (e) Structure of A-form RNA duplex containing two consecutive G-U wobble base pairs (PDB ID: 433D). (f) Structure of B-form DNA duplex containing A⁺-C wobble base pairs (PDB ID: 1D99). Nucleobases forming the wobble base pairs are emphasized dark. Hydrogen bonds in the wobble base pairs are shown in dashed lines.

strand [7]. However, neutral A-C mismatches are in equilibrium between the wobble and reverse wobble forms, each of which only forms one hydrogen bond. Thus, A-C mismatch is much less stable than G-T(U) mismatch in a physiological condition (Tables 2.1 and 2.2).

2.2.3 Purine–Purine Mismatches

Purine–purine and pyrimidine–pyrimidine mismatches are known as transversion mismatches. There are G-A, G-G, and A-A mismatches in the purine–purine mismatch. Among them, G-A mismatch forms relatively stable unusual base pairs in both DNA and RNA duplexes (Tables 2.1 and 2.2). It is highly polymorphic depending on sequence compositions. In DNA duplexes containing G-A mismatches that form two hydrogen bonds, various combinations of *anti* and *syn* were observed in their glycosidic bond angles (Figure 2.3). G-G mismatch potentially adopts a base pairing with two hydrogen bonds, in which two guanines are symmetrically or asymmetrically oriented with *anti* and *syn* conformation in their glycosidic bond angles (Figure 2.3). When the asymmetric G-G base pairs face each other in rotation, four guanines form a symmetric quartet as described later (Chapter 3). Recent X-ray diffraction analyses also demonstrated polymorphic feature of the G-G mismatch by showing that with *syn*–*syn* combination in the glycosidic bond angles in

Table 2.1 Thermodynamic parameters for duplex formations in 1M NaCl by DNA oligonucleotides containing mismatches^a.

Sequence	XY	$-\Delta H^\circ$	$-\Delta S^\circ$	$-\Delta G_{37}^\circ$	T_m
		(kcal mol ⁻¹)	(cal mol ⁻¹ K ⁻¹)	(kcal mol ⁻¹)	(°C at 10 ⁻⁴ M)
5'CAAA X AAAG	CG	64.5	183	7.7	42.9
3'GTTT Y TTTC	GC	62.8	179	7.3	40.8
	AT	68.0	196	7.2	40.1
	TA	58.6	168	6.5	36.8
	GG	53.5	158	4.5	25.6
	TG	55.6	165	4.4	25.7
	GA	52.6	156	4.2	23.9
	GT	46.7	137	4.2	22.3
	AG	39.9	116	3.9	18.0
	AA	36.9	107	3.7	15.0
	CT	53.2	161	3.3	19.1
	TC	50.0	151	3.2	17.5
	CA	(40.3) ^b	(120) ^b	(3.1) ^b	(13) ^b
	TT	(54.6) ^b	(167) ^b	(2.8) ^b	(17) ^b
	AC	(35.8) ^b	(106) ^b	(2.9) ^b	(9) ^b
	CC	(55.3) ^b	(171) ^b	(2.3) ^b	(15) ^b
5'CAACTTGATATTAATA +	Mismatch				
3'GTTGAACTATAATTAT	-	102.1	289	12.4	55.8
3'GTTGAGCTATAATTAT	TG	92.6	266	10.1	49.4
3'GTTGAACTATAGTTAT	TG	95.5	274	10.5	50.5
3'GTTGAACTCTAATTAT	TC	98.4	286	9.7	47.3
3'GTTGAATTATAATTAT	GT	91.3	264	9.4	47.1
3'GTTGAACCATAATTAT	AC	90.9	265	8.7	44.6
3'GTTGAACAATAATTAT	AA	92.0	267	9.26	46.2

a) Values are summarized ones in a reference [6].

b) Values in parenthesis are significantly less accurate estimated using flat lower baselines in their melting analyses.

the presence of chromomycin A3, which binds minor groove of the mismatched place and supports the structure analysis [8]. Detailed structure of A·A mismatch is rarely determined by X-ray diffraction analysis. It is considered that the mismatch is dynamically fluctuated and not able to be a particular structural state.