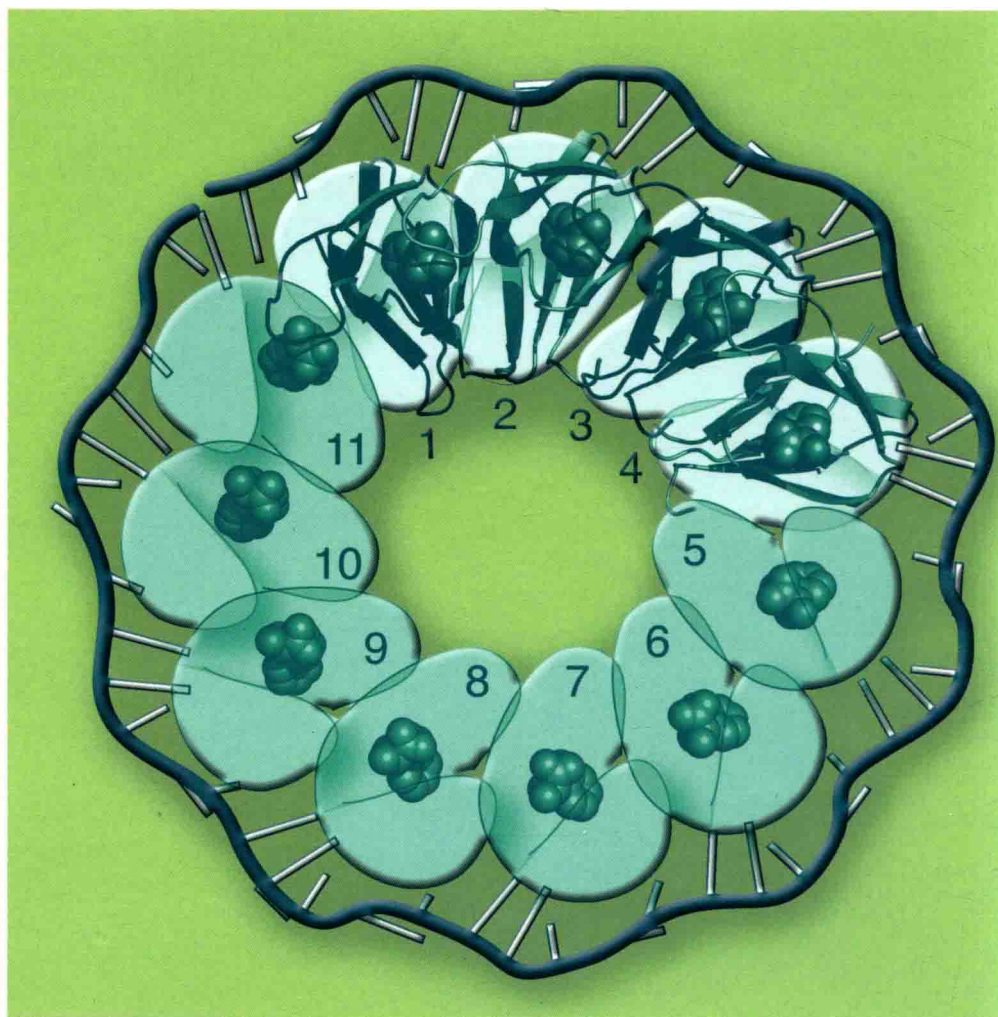


RSC Biomolecular Sciences

Edited by Phoebe A Rice and Carl C Correll

# Protein–Nucleic Acid Interactions

Structural Biology



RSC Publishing

# ***Protein-Nucleic Acid Interactions***

## ***Structural Biology***

Edited by

**Phoebe A. Rice**

*Department of Biochemistry & Molecular Biology, The University of Chicago, Chicago, IL, USA*

**Carl C. Correll**

*Department of Biochemistry and Molecular Biology, Rosalind Franklin University of Medicine and Science, North Chicago, IL, USA*

**RSC Publishing**

ISBN: 978-0-85404-272-2

A catalogue record for this book is available from the British Library

© Royal Society of Chemistry 2008

*All rights reserved*

*Apart from fair dealing for the purposes of research for non-commercial purposes or for private study, criticism or review, as permitted under the Copyright, Designs and Patents Act 1988 and the Copyright and Related Rights Regulations 2003, this publication may not be reproduced, stored or transmitted, in any form or by any means, without the prior permission in writing of The Royal Society of Chemistry or the copyright owner, or in the case of reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency in the UK, or in accordance with the terms of the licences issued by the appropriate Reproduction Rights Organization outside the UK. Enquiries concerning reproduction outside the terms stated here should be sent to The Royal Society of Chemistry at the address printed on this page.*

Published by The Royal Society of Chemistry,  
Thomas Graham House, Science Park, Milton Road,  
Cambridge CB4 0WF, UK

Registered Charity Number 207890

For further information see our web site at [www.rsc.org](http://www.rsc.org)

# Protein-Nucleic Acid Interactions

## Structural Biology

## RSC Biomolecular Sciences

### Editorial Board:

Professor Stephen Neidle (Chairman), *The School of Pharmacy, University of London, UK*

Dr Simon F. Campbell CBE, FRS

Dr Marius Clore, *National Institutes of Health, USA*

Professor David M.J. Lilley FRS, *University of Dundee, UK*

This Series is devoted to coverage of the interface between the chemical and biological sciences, especially structural biology, chemical biology, bio- and chemo-informatics, drug discovery and development, chemical enzymology and biophysical chemistry. Ideal as reference and state-of-the-art guides at the graduate and post-graduate level.

### Titles in the Series:

#### **Biophysical and Structural Aspects of Bioenergetics**

Edited by Mårten Wikström, *University of Helsinki, Finland*

#### **Computational and Structural Approaches to Drug Discovery: Ligand-Protein Interactions**

Edited by Robert M. Stroud and Janet Finer-Moore, *University of California in San Francisco, San Francisco, CA, USA*

#### **Exploiting Chemical Diversity for Drug Discovery**

Edited by Paul A. Bartlett, *Department of Chemistry, University of California, Berkeley* and Michael Entzeroth, *S\*Bio Pte Ltd, Singapore*

#### **Metabolomics, Metabonomics and Metabolite Profiling**

Edited by William J. Griffiths, *University of London, The School of Pharmacy, University of London, London, UK*

#### **Protein–Carbohydrate Interactions in Infectious Disease**

Edited by Carole A. Bewley, *National Institutes of Health, Bethesda, Maryland, USA*

#### **Protein-Nucleic Acid Interactions: Structural Biology**

Edited by Phoebe A. Rice, *Department of Biochemistry & Molecular Biology, The University of Chicago, Chicago IL, USA* and Carl C. Correll, *Dept of Biochemistry and Molecular Biology, Rosalind Franklin University, North Chicago, IL, USA*

#### **Quadruplex Nucleic Acids**

Edited by Stephen Neidle, *The School of Pharmacy, University of London, London, UK* and Shankar Balasubramanian, *Department of Chemistry, University of Cambridge, Cambridge, UK*

#### **Ribozymes and RNA Catalysis**

Edited by David M.J. Lilley FRS, *University of Dundee, Dundee, UK* and Fritz Eckstein, *Max-Planck-Institut für Experimentelle Medizin, Goettingen, Germany*

#### **Sequence-specific DNA Binding Agents**

Edited by Michael Waring, *Department of Pharmacology, University of Cambridge, Cambridge, UK*

#### **Structural Biology of Membrane Proteins**

Edited by Reinhard Grieshammer and Susan K. Buchanan, *Laboratory of Molecular Biology, National Institutes of Health, Bethesda, Maryland, USA*

#### **Structure-based Drug Discovery: An Overview**

Edited by Roderick E. Hubbard, *University of York, UK* and Vernalis (R&D) Ltd, *Cambridge, UK*

Visit our website on [www.rsc.org/biomolecularsciences](http://www.rsc.org/biomolecularsciences)

For further information please contact:

Sales and Customer Care, Royal Society of Chemistry, Thomas Graham House,  
Science Park, Milton Road, Cambridge, CB4 0WF, UK

Telephone: +44 (0)1223 432360, Fax: +44 (0)1223 426017, Email: [sales@rsc.org](mailto:sales@rsc.org)

# *Preface*

The structural biology of protein-nucleic acid interactions is in some ways a mature field and in others in its infancy. High-resolution structures of protein-DNA complexes have been studied since the mid 1980s. A vast array of such structures has now been determined, but surprising and novel structures still appear quite frequently.

High-resolution structures of protein-RNA complexes were relatively rare until the last decade. Propelled by advances in technology as well as a blossoming realization of RNA's importance to biology, the number of example structures has ballooned in recent years. As with many other fields, the deeper one digs, the more questions surface. New insights are now being gained from comparative studies only recently made possible due to the size of the database, as well as from careful biochemical and biophysical studies.

The field has, in some ways, been a victim of its own success: it is no longer possible to write a comprehensive review. Instead, current review articles tend to focus on particular subtopics of interest. This makes it difficult for newcomers to the field to attain a solid understanding of the basics. One goal of this book is therefore to provide in-depth discussions of the fundamental principles of protein-nucleic acid interactions as well as to illustrate those fundamentals with up-to-date and fascinating examples for those who already possess some familiarity with the field.

This book also aims to bridge the gap between the DNA- and the RNA-centric views of nucleic acid-protein recognition, which are often treated as separate fields. However, this is a false dichotomy because protein-DNA and protein-RNA interactions share many general principles. This book therefore includes relevant examples from both sides, and frames discussions of the fundamentals in terms that are relevant to both. History supports this approach: despite the amazing conformational versatility discovered for RNA, many of the lessons learned from early studies of protein-DNA complexes

could be applied directly to understanding newer protein-RNA complexes. Conversely, recent structures of proteins bound to noncanonical DNA structures reveal recognition strategies more commonly associated with RNA-binding proteins.

We have assembled a team of experts to write the individual chapters of this book. The beginning chapters (1–8) focus on more fundamental aspects of protein-nucleic acid interactions, such as thermodynamics and recognition strategies, while later chapters (9–14) highlight more specialized topics. Since it is impossible to cover all aspects of this rapidly expanding field, we have chosen to highlight a few topics that are fascinating in their own right while also providing a broad range of examples to underscore the basic principles laid out earlier. We hope that readers at all levels will find this an interesting guide and a useful reference.

Phoebe A. Rice  
Carl C. Correll

# Contents

## Chapter 1 Introduction

*Carl C. Correll and Phoebe A. Rice*

1.1	Overview	1
1.2	Fundamentals of DNA and RNA Structure	1
1.2.1	Stabilizing Forces	1
1.2.2	Chemical Differences between DNA and RNA	3
1.2.3	Canonical A- and B-form Helices	4
1.2.4	Deviation is the Norm	6
1.2.5	Bending and Supercoiling DNA	6
1.2.6	Folded RNA and Noncanonical DNA	7
1.3	Principles of Recognition	7
1.3.1	Forces that Contribute to Complex Formation	8
1.3.2	Site Recognition Overview	8
1.3.3	Recognizing Duplex DNA <i>via</i> Direct and Indirect Readout	8
1.3.4	Recognizing Single-stranded Nucleic Acids	9
1.3.5	Recognizing Folded RNAs	9
1.3.6	Recognizing Noncanonical DNA Structures	10
1.3.7	Conformational Rearrangements	10
1.4	Future Directions	11
	References	11



## **Chapter 2 Role of Water and Effects of Small Ions in Site-specific Protein-DNA Interactions**

*Linda Jen-Jacobson and Lewis A. Jacobson*

2.1	Introduction	13
2.2	Affinity and Specificity	14
2.3	Macromolecular Hydration Influences $\Delta H^\circ$ , $\Delta S^\circ$ and $\Delta C_p^\circ$	15
2.4	Water Release Attending Protein-DNA Association	17
2.5	Retained Water Molecules Contribute to Affinity and Specificity	26
2.6	Thermodynamic Effects of Retained Water	27
2.7	Overview of Small Ion Effects on Protein-DNA Interactions	28
2.8	Multiple Physical Phenomena Associated with Salt Dependence	28
2.9	Cation Release Favors Protein-DNA Association	30
2.10	Selective Effects of Anions on Protein-DNA Binding	32
2.11	Divalent Cation Binding at Active Sites Relieves Electrostatic Strain	34
2.12	Ion Effects and Cosolute Effects are Mechanistically Independent	36
2.13	Comparison with Nonspecific Binding: How Water and Ions Affect Specificity	36
2.14	Conclusions	39
	Acknowledgements	40
	References	40

## **Chapter 3 Structural Basis for Sequence-specific DNA Recognition by Transcription Factors and their Complexes**

*Manqing Hong and Ronen Marmorstein*

3.1	Introduction	47
3.2	Transcriptional Regulators that Bind Core DNA Elements	48
3.2.1	Helix-turn-helix and Winged Helix-turn-helix	48
3.2.2	Basic Leucine-zipper and Basic Helix-loop-helix	50
3.2.3	Zinc-binding Domains that Bind as Monomeric Units	52
3.2.4	DNA Recognition by $\beta$ -Ribbons	53
3.2.5	Immunoglobulin Fold	53
3.2.6	HMG Domain	54

3.3	Transcriptional Regulators that Bind as Dimers to two DNA Half Sites with Different Spacing and Polarity	54
3.3.1	Zn <sub>2</sub> Cys <sub>6</sub> Binuclear Cluster	54
3.3.2	Nuclear Receptors	57
3.4	Transcription Regulatory Complexes that use a Combination of Different DNA-binding Motifs	59
3.4.1	Combinatorial DNA Interactions	59
3.4.2	ETS Family Ternary Complexes	60
3.4.3	NFAT/Fos-Jun/DNA Quaternary Complex	62
3.5	Conclusions	62
	References	63

## Chapter 4 Indirect Readout of DNA Sequence by Proteins

*Catherine L. Lawson and Helen M. Berman*

4.1	Introduction	66
4.1.1	DNA Sequence Recognition: A Historical Perspective	66
4.2	Indirect Readout	68
4.2.1	Direct vs. Indirect Readout	68
4.2.2	Language of Indirect Readout: DNA Geometry	69
4.2.3	Sequence-dependent Polymorphisms of B-DNA	69
4.2.3.1	Base Stacking	70
4.2.3.2	Hydrogen Bonding	72
4.2.3.3	Steric Repulsion	72
4.2.3.4	DNA Bending	72
4.2.4	Indirect Readout: A Universal Feature of Protein-DNA Interactions	72
4.3	DNA Sequence Recognition by CAP	73
4.3.1	Direct Readout by CAP	76
4.3.2	Indirect Readout by CAP	79
4.3.2.1	Conformation and Flexibility of the DNA Site for CAP	79
4.3.2.2	Indirect Readout at Positions 1-2	80
4.3.2.3	Indirect Readout at Position 6	80
4.3.2.4	Comparison with Other Protein-induced Positive Roll Deformations	81
4.3.3	DNA Bending vs. DNA Kinking – A Dynamic Duo?	83
4.4	Conclusions	86
	Acknowledgements	86
	References	86

## Chapter 5 Single-stranded Nucleic Acid (SSNA)-binding Proteins

*Martin P. Horvath*

5.1	Introduction	91
5.2	Basic Elements	93
5.2.1	Interaction Types	93
5.2.1.1	Salt Bridges and Electrostatics	93
5.2.1.2	Stacking Interactions	95
5.2.1.3	Steric Packing and van der Waals Interactions	95
5.2.1.4	Hydrogen Bonding	96
5.2.2	Folds, Evolution and Function	98
5.2.2.1	OB-fold	98
5.2.2.2	Sm-fold	103
5.2.2.3	RRM	106
5.2.2.4	KH	106
5.2.2.5	Others: Pumilio, TRAP and Whirly	108
5.3	Emergent Properties	110
5.3.1	Molecular Recognition: Specificity, Adaptability and Degeneracy	111
5.3.1.1	Specific yet Adaptable Recognition by Modular Puf Proteins	112
5.3.1.2	A “Hot-spot” for Recognition of Telomere DNA by Cdc13	113
5.3.1.3	“Nucleotide Shuffling” and TEBP- $\alpha/\beta$	113
5.3.1.4	Degeneracy in Splicing Branch Site Identification	114
5.3.2	Cooperativity	115
5.3.2.1	SSB and Multiple Cooperativity Modes	115
5.3.2.2	Anti-cooperativity and TEBP- $\alpha$	116
5.3.2.3	Positive Heterotypic Cooperativity at Telomere Ends	117
5.3.3	Allostery	117
5.3.3.1	Small Molecule Effectors and SSNA-binding	118
5.3.3.2	Proteins as Allosteric Effectors for Binding and Release of SSNA	119
5.4	Conclusion and Perspective	120
	Acknowledgements	120
	References	121

**Chapter 6 DNA Junctions and their Interaction with Resolving Enzymes***David M.J. Lilley*

6.1	The Four-way Junction in Genetic Recombination	129
6.2	Structure and Dynamics of DNA Junctions	129
6.2.1	Dynamics of the Four-way Junction	131
6.2.2	Metal Ions and the Electrostatics of the Four-way Junction	131
6.2.3	Branch Migration	132
6.2.4	Comparison with Four-way RNA Junctions	134
6.3	Proteins that Interact with DNA Junctions	134
6.4	Junction-resolving Enzymes	134
6.4.1	Occurrence of the Junction-resolving Enzymes	135
6.4.2	Phylogeny	135
6.4.3	Junction-resolving Enzymes are Dimeric	135
6.4.4	Structures of the Junction-resolving Enzymes	135
6.5	Molecular Recognition and Distortion of the Structure of DNA Junctions by Resolving Enzymes	137
6.5.1	Sequence Specificity of the Junction-resolving Enzymes	138
6.5.2	Structural Distortion of DNA Junctions by the Junction-resolving Enzymes	138
6.5.3	Coordination of the Resolution Process	139
6.6	T7 Endonuclease I	139
6.6.1	Biochemistry of Endonuclease I	139
6.6.2	Structure of Endonuclease I	140
6.6.3	The Active Site	141
6.6.4	Catalysis of Phosphodiester Bond Hydrolysis	141
6.6.5	Interaction between Endonuclease I and DNA Junctions	142
6.7	In Conclusion	144
	Acknowledgements	145
	References	145

**Chapter 7 RNA-protein Interactions in Ribonucleoprotein Particles and Ribonucleases***Hong Li*

7.1	Introduction	150
7.2	Experimental Methods used to Determine RNA-protein Complex Structures	151

7.3	RNA-protein Interactions in Ribonucleoprotein Particles	152
7.3.1	Ribosome	153
7.3.2	RNAi Complexes	156
7.3.3	Signal Recognition Particle	159
7.3.4	s(no)RNPs	160
7.3.5	RNA Editing Complexes	164
7.4	RNA-protein Interactions in Ribonucleases	165
7.4.1	RNase E	165
7.4.2	RNase II	166
7.4.3	RNase III	167
7.4.4	Restrictocin	168
7.4.5	RNA Splicing Endonucleases	169
7.4.6	tRNase Z	170
7.5	Concluding Remarks	170
	Acknowledgements	171
	References	171

## **Chapter 8 Bending and Compaction of DNA by Proteins**

*Reid C. Johnson, Stefano Stella*

*and John K. Heiss*

8.1	Introduction	176
8.2	Forces Controlling DNA Rigidity	178
8.2.1	DNA Elasticity and the Influence of DNA Sequence	178
8.2.2	Base Stacking Primarily Controls Helix Rigidity	179
8.2.3	Electrostatic Forces Modulate DNA Bending	180
8.3	Bending of DNA at High Resolution	183
8.3.1	Helix Parameters Controlling DNA Structure	184
8.3.1.1	Roll and Tilt	184
8.3.1.2	Twist	185
8.3.1.3	Propeller Twist, Slide, and Shift	186
8.3.1.4	Changes in DNA Groove Width	186
8.3.2	Influence of Exocyclic Groups on Base Stacking	188
8.3.3	Flexibility of Dinucleotide Steps	189
8.3.3.1	Pyrimidine-purine (Y-R) Steps	189
8.3.3.2	Purine-purine (R-R) or Pyrimidine-pyrimidine (Y-Y) Steps	189
8.3.3.3	Purine-pyrimidine (R-Y) Steps	190

8.4	Examples of DNA Bending Proteins	191
8.4.1	Histone Binding to DNA	191
8.4.2	Phage $\lambda$ Xis Protein	194
8.4.3	Papillomavirus E2 Protein	194
8.4.4	<i>Escherichia coli</i> Fis Protein	197
8.4.4.1	Long-range DNA Condensation by Fis	198
8.4.5	<i>Escherichia coli</i> CAP Protein	201
8.4.6	Prokaryotic HU/IHF Protein Family	203
8.4.6.1	Single-DNA Molecule Analysis of HU/IHF Protein Binding	207
8.4.7	HMGB Protein Family	208
8.4.7.1	Single DNA Molecule Analyses of HMGB Protein Binding	211
8.4.7.2	DNA Binding by HMGB Shares Features with TBP	212
8.5	Concluding Remarks	212
	References	213

## Chapter 9 Mode of Action of Proteins with RNA Chaperone Activity

*Sabine Stampfl, Lukas Rajkowitsch, Katharina Semrad and Renée Schroeder*

9.1	Introduction	221
9.1.1	RNA Folding	221
9.1.2	Proteins with RNA Chaperone Activity (RCA)	222
9.1.3	Measuring RCA	223
9.2	Mode of Action of Proteins with RCA	223
9.2.1	RNA Annealing Activity	223
9.2.1.1	Annealing of Protein-bound Guide RNAs with Target RNAs	226
9.2.2	Nucleic Acid Melting Activity	227
9.3	RNA Binding and Restructuring	228
9.3.1	Proteins with RCA Interact with RNA only Weakly	228
9.3.2	Proteins with Specific RNA-binding Affinity	229
9.3.3	Protein Structure and RNA Chaperone Activity	230
	Acknowledgements	231
	References	231

## Chapter 10 Structure and Function of DNA Topoisomerases

*Ken C. Dong and James M. Berger*

10.1	Introduction	234
10.2	Type IA Topoisomerases	238

10.2.1	Overview	238
10.2.2	Structures and Mechanism	239
10.2.3	Type IA Topoisomerase Paralogs	242
10.2.3.1	Topoisomerase III	243
10.2.3.2	Reverse Gyrase	243
10.3	Type IB Topoisomerases	244
10.3.1	Overview	244
10.3.2	General Architecture	245
10.3.3	DNA Recognition and Cleavage	246
10.3.4	Mechanism	248
10.4	Topoisomerase V – The Defining Member of the Type IC Topoisomerases?	249
10.5	Type IIA Topoisomerases	249
10.5.1	Overview	249
10.5.2	Structural Organization	253
10.5.2.1	ATPase Domain	253
10.5.2.2	DNA Breakage/Reunion Domain and the DNA Binding/Cleavage Core	254
10.5.3	Duplex DNA Transport Mechanism	255
10.5.3.1	Type IIA Topoisomerase Paralogs: Role of the C-terminal Domain in Modulating Duplex Transport	256
10.5.4	Physiological Specialization of Type IIA Topoisomerases	258
10.6	Type IIB Topoisomerases	259
10.6.1	Overview	259
10.6.2	Structure	259
10.6.3	Mechanism	260
10.7	Conclusions	261
	Acknowledgements	261
	References	261

## Chapter 11 DNA Transposases

*Fred Dyda and Alison Burgess Hickman*

11.1	Introduction	270
11.1.1	Nomenclature, Classification, and Overview of Transposition Systems	271
11.1.1.1	Prokaryotic Elements	271
11.1.1.2	Eukaryotic Elements	273
11.2	Transposases and DNA	274
11.2.1	Ends of Transposons	274
11.2.2	Chemistry of DNA Transposition	276

11.2.3	DDE Transposases	276
11.2.4	Cut-and-paste, Copy-in, and Copy-out Transposition	277
11.2.4.1	Mechanism of DNA Cleavage and Strand Transfer	277
11.2.4.2	Second Strand Cleavage by Cut-and-paste Transposases	280
11.2.4.3	Replicative Transposition	281
11.2.5	Transposition Happens in Context	281
11.3	Tn5 Transposase: The Minimum Necessary	282
11.3.1	Overview of the Tn5 Transposition Pathway	283
11.3.2	Structure of a Tn5 Transposase Dimer Bound to DNA	283
11.4	Bacteriophage Mu	284
11.4.1	Organization of the Mu Genome Ends	285
11.4.2	Domain Structure of MuA	285
11.4.3	Putting Mu Ends and MuA together: The Mu Transpososome	287
11.5	Tc1/ <i>mariner</i> Transposases	288
11.5.1	Tc3 N-terminal Domains and DNA	289
11.5.2	<i>MosI</i> Transposase	290
11.5.3	What is the Active Assembly?	290
11.6	<i>hAT</i> Elements: A First Glimpse	290
11.6.1	Domain Organization of Hermes	292
11.7	Transposases that Form Covalent Phosphotyrosine or Phosphoserine Intermediates	293
11.7.1	Y and S Transposases	293
11.7.2	Y2 Transposases	294
11.7.3	Y1 Transposases	296
11.7.3.1	ISHp608 TnpA	296
	Acknowledgements	299
	References	299

## Chapter 12 Site-specific Recombinases

*Gregory D. Van Duyne*

12.1	Introduction	303
12.2	Tyrosine Recombinases	307
12.2.1	Cre and Flp Recombinases	309
12.2.2	$\lambda$ -Integrase and XerCD Recombinases	313
12.2.3	Integron Integrases	316
12.2.4	Other Tyrosine Recombinases	318
12.3	Serine Recombinases	318
12.3.1	Tn3 and $\gamma\delta$ -Resolvases	321
12.3.2	Hin and Gin Recombinases	323



12.3.3	Regulation by Accessory Sites	324
12.3.4	Large Serine Recombinases	326
12.4	Summary	327
	References	328

## Chapter 13 DNA Nucleases

*Nancy C. Horton*

13.1	Introduction	333
13.2	Summaries by Fold	334
13.2.1	Restriction Endonuclease-like Fold	334
13.2.2	RNaseH-like Fold	341
13.2.3	Homing Endonuclease-like Fold	343
13.2.4	His-Me Finger Endonucleases	345
13.2.5	SAM Domain-like/PIN Domain-like Fold	348
13.2.6	DNase I-like Fold	350
13.2.7	Phospholipase C/P1 Nuclease Fold	352
13.2.8	Phospholipase D/Nuclease Fold	354
13.2.9	TIM beta/alpha Barrel Fold	355
13.2.10	DHH Phosphoesterases Fold	357
13.2.11	GIY-YIG Endonuclease Fold	358
13.2.12	Metallo-dependent Phosphatases Fold	360
13.2.13	<i>Bacillus chorismate</i> Mutase-like Fold	361
13.3	Conclusion	361
	References	363

## Chapter 14 RNA-modifying Enzymes

*Adrian R. Ferré-D'amaré*

14.1	Introduction: Scope of RNA Modification	367
14.2	The tRNA Adenosine Deaminase Tada	369
14.2.1	RNA Recognition by Loop Eversion	369
14.2.2	Hydrolytic Deamination by a Zinc-activated Water	370
14.3	The tRNA Pseudouridine Synthase RluA	372
14.3.1	RNA Recognition through Protein-induced Base-pairing	374
14.3.2	In-line Displacement or Michael Addition?	374
14.4	The tRNA Archaeosine Transglycosylase ArcTGT	376
14.4.1	RNA Recognition by Tertiary Structure Rearrangement	376