Introduction to Protein–DNA Interactions

Structure, Thermodynamics, and Bioinformatics

ALSO FROM COLD SPRING HARBOR LABORATORY PRESS

Other Titles of Interest

Bioinformatics: Sequence and Genome Analysis, Second Edition

Genes & Signals

A Genetic Switch, Third Edition: *Phage Lambda Revisited*

Molecular Cloning: A Laboratory Manual, Fourth Edition

Introduction to Protein–DNA Interactions

Structure, Thermodynamics, and Bioinformatics

Gary D. Stormo, Ph.D.

COLD SPRING HARBOR LABORATORY PRESS
Cold Spring Harbor, New York • www.cshlpress.org

Introduction to Protein–DNA Interactions
Structure, Thermodynamics, and Bioinformatics

© 2013 by Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
All rights reserved
Printed in the United States of America

Publisher	John Inglis
Acquisition Editors	Ann Boyle and Kaaren Janssen
Director of Editorial Development	Jan Argentine
Developmental Editor	Judy Cuddihy
Project Manager	Maryliz Dickerson
Permissions Coordinator	Carol Brown
Production Manager	Denise Weiss
Production Editor	Rena Steuer
Compositor	Techset Ltd.
Cover Designer	Ed Atkeson

Front cover: Computer-generated structural diagram showing the overall geometry of the Lac repressor protein binding to *lac* operator DNA (image generated using Pymol software from data in the Protein Data Bank database, entry 2KEI).

Library of Congress Cataloging-in-Publication Data

Stormo, Gary.
Introduction to protein-DNA interactions : structure, thermodynamics, and bioinformatics / Gary D. Stormo.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-1-936113-49-1 (hard cover : alk. paper) – ISBN 978-1-936113-50-7 (pbk. : alk. paper)
I. Title.
[DNLM: 1. DNA-Binding Proteins–pharmacokinetics. 2. Binding Sites. 3. DNA–chemistry. 4. Protein Binding. 5. Transcription Factors. QU 58.5]

572.8′6459–dc23

2012035448

10 9 8 7 6 5 4 3 2 1

All World Wide Web addresses are accurate to the best of our knowledge at the time of printing.

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Cold Spring Harbor Laboratory Press, provided that the appropriate fee is paid directly to the Copyright Clearance Center (CCC). Write or call CCC at 222 Rosewood Drive, Danvers, MA 01923 (508-750-8400) for information about fees and regulations. Prior to photocopying items for educational classroom use, contact CCC at the above address. Additional information on CCC can be obtained at CCC Online at http://www.copyright.com.

For a complete catalog of Cold Spring Harbor Laboratory Press publications, visit our website at www.cshlpress.org.

To my parents, Milo and Claryce, who gave me the love of learning and the encouragement to follow wherever that led.

To my wife, Susan Dutcher, and my children, Ben and Adrienne, who have enriched my life immeasurably.

Contents

Preface, ix

1 Importance of Protein–DNA Interactions, 1

STRUCTURE

2 The Structure of DNA, 13

3 Protein Structure and DNA Recognition, 27

4 Sequence-Specific Interactions in Protein–DNA Complexes, 49

THERMODYNAMICS

5 Binding Affinity, Cooperativity, and Specificity, 67

6 Energetics and Kinetics of Binding, 89

BIOINFORMATICS

7 Bioinformatics of DNA-Binding Sites, 109

8 Bioinformatics of Transcription Factors and Recognition Models, 131

9 Transcriptional Genomics, 153

Index, 193

Preface

THE BIOLOGICAL IMPORTANCE of PROTEIN–DNA INTERACTIONS has been recognized since the early 1960s, starting with the discovery by Jacob and Monod of the *lac* operon and its regulation in *Escherichia coli*. In the intervening 50 years, studies of protein–DNA interactions have made significant contributions to most areas of molecular, cellular, and developmental biology. A wide range of approaches has been applied in those studies, but they can be broadly classified into the three types that are the focus of this book: structural, thermodynamic, and bioinformatic. The earliest studies used biochemical and biophysical methods to analyze the thermodynamic and kinetic aspects of protein–DNA interactions. The first binding site sequences were determined in the early 1970s, which led to hypotheses about recognition mechanisms and the information required for regulatory systems to function. Technological advances in the late 1970s and the early 1980s, including the ability to sequence and synthesize DNA and to clone, express, and purify large quantities of proteins, facilitated many new types of studies. The earliest bioinformatics approaches were developed in the late 1970s, as soon as there were enough sequences for statistical analyses to be worthwhile. Shortly after that, as it became much easier to synthesize and purify sufficient quantities of specific proteins and DNA sequences of interest, structural studies rapidly increased. Further technological advances in the last two decades have continued to accelerate the pace of discovery. Most important have been further efficiencies in DNA sequencing that have resulted not only in whole-genome sequences for many species but also whole-genome and mRNA sequences from individuals as well as a variety of other sequence-based data sets. Our understanding of protein–DNA interactions and their roles in a wide range of biological processes has grown enormously, but there is still much we do not know and the field continues to be ripe for further discovery.

The primary goal of this book is to provide an introduction to protein–DNA interactions that bridges the three classes of approaches. Experts in any of the fields are not

likely to learn anything new within their field; in fact, they will undoubtedly find examples of details being glossed over in favor of a simplified presentation. But experts in one area tend to have more cursory knowledge of the other fields and thus may learn from other sections of the book. Those who are new to the study of protein–DNA interactions or those outside the field with a casual interest in the topic may gain new insights throughout the book. If so, the book has succeeded even beyond the fact that I learned something in the process of writing every chapter.

The regulation of gene expression has fascinated me since my graduate school days. I have ventured into other topics, mostly related to how computer programs can help to uncover biological knowledge, but the majority of my efforts have been focused on understanding how networks of transcription factors regulate gene expression and control cell fates and phenotypes. I have been extremely fortunate to have been associated throughout my career with teachers and students, colleagues and collaborators, and most of all friends who have taught and encouraged me and made my whole adventure enjoyable. The list of those who made significant contributions to my research, many of whom I have never met but have benefited from immensely through reading their papers, is too long to include in this preface. But a few have had such a large influence that I must thank them here. Larry Gold, my graduate and postdoc advisor, kept research always fun and gave me the freedom and encouragement to follow an unconventional path. Tom Schneider, a fellow student in Larry's lab, and Andrej Ehrenfeucht, a mentor in all things computational, were there from the beginning and opened my eyes to new horizons that I would have missed without them. I have had many great collaborators over the years but special thanks go to John Heumann, Alan Lapedes, and Charles "Chip" Lawrence, each of whom has filled gaps in my knowledge and provided numerous insights into my own work that were initially invisible to me. I have also had many great students and postdocs who made progress possible and who taught me at least as much as I taught them.

This book would not have happened with the support and encouragement of the individuals at Cold Spring Harbor Laboratory Press, including Ann Boyle, Maryliz Dickerson, Kaaren Janssen, and Rena Steuer. Judy Cuddihy, in particular, made numerous improvements and helped at every step. I also thank those authors and publishers who allowed me to use their figures.

Index

A

Adenine (A), 13
A-DNA, 24–26
Affinity, binding. *See* Binding affinity
α helix secondary structure, 32–33
AMBER, 101
Amino acid properties
  categories, 29, 30f
  nonpolar hydrophobic amino acids, 29, 31
  polar acidic residues, 32
  polar basic residues, 31
  polar uncharged residues, 31
  special cases, 32
  structural function, 29
Artemisinic acid, 167
Association constant (K_A), 69, 98. *See also* Binding affinity
Avery, Oswald, 14

B

B1H (bacterial one-hybrid) methods, 84
Bacillus genus, 163, 164
Bacteriophage λ
  choice between lysis and lysogeny, 5, 7–8
  competition for binding between Cro and λ, 7
  operator binding sites, 7
  principles of sequence-specific TFs, 8
  regulatory region elements, 5–6
Basic-region leucine zipper (bZIP), 41, 95, 144–145
B-DNA, 24–26
Beadle, George, 18
bHLH (basic-region helix-loop-helix), 41, 95, 144–145
Binding affinity
  assay methods, 76
  binding probability equation, 69
  determination methods, 79–80
  factors influencing the rate of complex formation, 68, 70f
  K_D determination methods, 70f
    about, 69
    EMSAs, 71–72
    filter-binding assays, 71
    fluorescence anisotropy, 73
    SPR, 72–73
  measuring affinities of multiple sites simultaneously, 80–81
  nonspecific, 49
Binding cooperativity
  about, 73–74
  affinity assay methods, 76
  cooperativity constant, 74
  nuclease protection, 76–78
  physical basis of positive cooperativity, 75–76
  probability-of-each-state calculations, 74–75
Binding location analyses, 160–161
Binding-site motifs discovery
  expectation maximization, 126–127
  Gibbs sampling, 127–128
  greedy alignments, 125–126
  “motif discovery” problem, 123, 124f
  pattern searches, 123–125
  pros and cons of methods, 123
Binding specificity
  about, 78
  bioinformatics of DNA-binding sites and, 109–110
  estimating specificity needed for a regulatory system, 78–79
  limits to specificity determination, 149
  methods for determining
    bacterial one-hybrid, 84
    basis of, 80
    CSI, 83
    determining affinity and, 79–80

Binding specificity (*Continued*)
measuring affinities of multiple sites simultaneously, 80–81
MITOMI, 81–83
PBMs, 83, 148
SELEX, 83–84, 148
quantitative definition of specificity, 84–86
recognition model used to determine, 147–148
sequence-specific interactions (*See* Sequence-specific interactions)
specificity modeling by PWM
discriminatory models, 119
higher-order models, 121–122
probabilistic models, 113–119
regression models, 119–121
Bioinformatics of DNA-binding sites
position weight matrix (*See* Position weight matrix)
representation of the specificity of TFs, 109–110
Bioinformatics of TFs and recognition models
hidden Markov model
examples of TF profile HMMs, 143–146
probability of generating a particular sequence, 142–143
protein sequences alignment example, 138–140
pseudocounts additions, 141
sequence logos, 141–142
types of states, 140–141
identifying homologous TFs
assessing if two proteins are homologous, 134
BLAST database search method, 134, 138
BLOSUM62 substitution matrix, 132–134
methods used to predict the function of a protein, 131
mutations and, 132
optimal alignments with dynamic programming, 135–137
orthologs and paralogs, 132
recognition models
binding specificity determination method, 147–148
focus on developing a predictive model, 146–147
lack of a recognition code and, 146
limitations of the recognition code, 149–150
limits to specificity determination, 149
method to determine binding specificity, 147–148
phage-display method, 148–149
quantitative models, 150
BLAST database search method, 136, 138
BLOSUM62 substitution matrix, 132–134
Britten, Roy, 15
β strands and β sheets secondary structure, 33–34
bZIP (basic-region leucine zipper), 41, 95, 144–145

C

C2H2 zinc finger family, 138, 140–143, 144
Caenorhabditis elegans, 159
"Calling cards," 160–161
cAMP (cyclic AMP), 4, 5
cAMP receptor protein (CRP), 4, 94, 123, 156
Cancer Genome Anatomy Project (CGAP), 174
CAP (catabolite activator protein), 4
Carroll, Sean, 173
CGAP (Cancer Genome Anatomy Project), 174
CHARMM, 101
ChIP (chromatin immunoprecipitation), 123, 160–161
ChIP-seq experiments, 160, 162, 168–170
chip-seq experiments, 175
Cognate site identifier (CSI), 83
Coiled-coil helix dimers, 41, 42f
Cooperativity. *See* Binding cooperativity
COSY (correlation spectroscopy), 38
Crick, Francis H.C., 13
Cro protein, 5–8, 35–36
CRP (cAMP receptor protein), 4, 94, 123, 156
Cyclic AMP (cAMP), 4, 5
Cytosine (C), 13

D

DamID, 160–161
Delete states, 140–141
DHS (DNase I hypersensitive sites), 162, 170–172
Discriminatory models for specificity, 119
Dissociation constant (K_D), 70f. *See also* Binding affinity
about, 69
EMSAs, 71–72
filter-binding assays, 71
fluorescence anisotropy, 73
SPR, 72–73
DNA accessibility analyses, 162
DNase I, 162
DNase I hypersensitive sites (DHS), 162, 170–172
Double helix. *See* Structure of DNA
Drosophila
conservation of enhancers function, 181–182, 183f
embryonic development steps, 179–180
history as a model organism, 177
research advances, 180
dsDNA (double-stranded DNA), 15–17
Dynamic programming, 135–137

E

EcoCyc, 163
EcoRI, 53–54
Eggert, M, 125
EM (expectation maximization), 126–127
EMSAs (electrophoretic mobility-shift assays), 71–72
ENCODE project
about, 167–168
challenges in studying multicellular eukaryotes, 167

ChIP-seq experiments, 168–170
DNase I hypersensitive experiments, 170–172
project expansion, 168, 169f
endo16 gene, 177, 179f
Energetics and kinetics of binding. *See* Thermodynamics of TF binding
Enhancers, 9
Enthalpy (H), 90, 92, 94–95
Entropy (S), 90, 92–95, 117, 119
Escherichia coli
gene expression regulation and the *lac* operon, 3–5
gene regulatory networks study, 163–165
scaling up to human dimensions example, 17–18
even-skipped (*eve*), 180, 181, 182f, 183f
Expectation maximization (EM), 126–127
Expression analyses, 157–160

F

FAIRE (formaldehyde-assisted isolation of regulatory elements), 162
FFL (feed-forward loop), 156–157
Filter-binding assays, 71
Fluorescence anisotropy, 73
Fly-Ex, 180
FlyNet, 180
"Fly Room" laboratory, 177
Formaldehyde-assisted isolation of regulatory elements (FAIRE), 162
fushi tarazu (*ftz*), 180

G

Galas, DJ, 125
GATA family, 43–44, 143–144
Gel- or band-shift assays, 71–72
Gene expression regulation. *See also* Gene regulatory networks
bacteriophage λ
choice between lysis and lysogeny, 7–8
competition for binding between Cro and λ, 7
operator binding sites, 7
principles of sequence-specific TFs, 8
regulatory region elements, 5–6
lac operon of *E. coli* and, 3–5
mystery of, 2–3
principles of protein–DNA interactions and, 18–19
specificity of TFs and, 78
Generative probabilistic models, 115, 117
Gene regulatory networks (GRNs)
Binding-site information and, 157
characteristics of biological networks, 156
Drosophila embryonic patterning, 177, 179–182, 183f
feed-forward loop network motif, 156–157
genetic variation and, 172–175
modeling conventions, 154–155
model systems' characteristics, 175–176
sea urchin studies, 176–177, 178f, 179f
study of
bacteria based, 163–165
ENCODE project, 167–172
genetic variation, 172–175
limitations from studying only TFs and their targets, 162–163
synthetic biology, 165–166
yeast, 166–167
"wiring diagram" uses, 155–156
Genetic variation and GRNs
concept of the "human genome," 173–174
genome-wide association studies, 173–174
levels of DNA variation, 172–173
regulation differences focus, 173
sequence differences mechanisms focus, 173
Genome-wide association studies (GWAS), 174–175
Gibbs sampling, 127–128
Gibbs standard free energy of binding, 69, 89–90
Greedy alignments, 125–126
GRNs. *See* Gene regulatory networks
Guanine (G), 13
GWAS (Genome-wide association studies), 174–175

H

H (enthalpy), 90, 92, 94–95
HapMap project, 174
Helix-turn-helix protein family, 5, 35, 39–41, 145–146
Helix-turn-helix proteins, 51
Hidden Markov model (HMM)
examples of TF profile HMMs, 143–146
probability of generating a particular sequence, 142–143
protein sequences alignment example, 138–140
pseudocounts additions, 141
sequence logos, 141–142
types of states, 140–141
Higher-order models for specificity, 121–122
Homeodomain proteins, 41
Homodimers, 22
Homologous TFs
assessing if two proteins are homologous, 134
BLAST database search method, 134, 138
BLOSUM62 substitution matrix, 132–134
methods used to predict the function of a protein, 131
mutations and, 132
optimal alignments with dynamic programming, 135–137
orthologs and paralogs, 132

Human Microbiome Project, 164
Hydrophobic effect, 95

I

IC (information content) measurement, 118–119
Insert states, 140–141
International Genetically Engineered Machine (iGEM) Foundation, 165
Int protein, 7
Introns, 9
ITC (isothermal titration calorimetry), 92, 93

J

Jacob, François, 3

K

K_A (association constant), 69, 98. *See also* Binding affinity
K_D. *See* Dissociation constant (K_D)
Kendrew, John, 36
King, Mary-Claire, 173
Kullback-Leibler distance, 119

L

lac operon of *E. coli*
 compared to the λ repressor, 8, 9
 gene expression regulation and, 3–5, 61
Lac repressor
 binding specificity of, 99, 103, 156
 helix-turn-helix protein family, 39–40
 lactose regulatory system and, 3–5, 8, 47
 sequence-specific interactions, 61–63
Lactose, 3–5, 47
Lewis, Edward, 177
Likelihood ratios, 116–117
Log-odds PWM, 117
λ repressor protein, 5–8
Lysis/lysogeny decision of phage DNA, 5, 7–8

M

Major groove, 19–20, 51
Markov chain Monte Carlo (MCMC), 101
Match states, 140–141
MC (Monte Carlo) methods, 101
MD (molecular dynamics) simulations, 101
Melting DNA, 15–17
MicrobesOnline, 164
Minor groove, 20–21
MITOMI (mechanically induced trapping of molecular interactions), 81–83
Molecular dynamics (MD) simulations, 101
Monod, Jacques, 3
Monte Carlo (MC) methods, 101
Morgan, T.H., 177
Motif discovery problem, 123, 124f
mRNA (messenger RNA)
 measuring using microarrays, 157–158
 protein–DNA interactions and, 2, 3–5, 8–9
 role within a cell, 17
 sequencing, 159
Mullis, Dary, 16
Mutations and homologous TFs, 132

N

Ndt80, 59–60
NFAT (nuclear factor of activated T cells), 45
NF-**κ**B, 45
NMR (nuclear magnetic resonance), 37–39
NOESY (nuclear Overhauser effect spectroscopy), 38
Noncoding DNA, 9
Nonpolar hydrophobic amino acids, 29, 31
Nonspecific binding affinity, 49
Nuclear factor of activated T cells (NFAT), 45
Nuclear magnetic resonance (NMR), 37–39
Nuclear Overhauser effect spectroscopy (NOESY), 38
Nuclease protection, 76–78
Nucleosomes, 9
Nüsslein-Volhard, Christiane, 177

O

1D (one-dimensional) diffusion, 103
1000 Genomes Project, 174
One-dimensional (1D) diffusion, 103
Orthologs, 132

P

p53, 45
Paralogs, 132
Pauling, L., 172
PBMs (protein-binding microarrays), 83, 148
PCR (polymerase chain reaction), 16–17, 36
Perutz, Max, 36
PFM (position frequency matrix), 142
Phage display, 148–149
Phosphorylation, 46
Phylogenetic footprinting, 128, 129f
Polar acidic residues, 32
Polar basic residues, 31
Polar uncharged residues, 31
Polymerase chain reaction (PCR), 16–17, 36
Position frequency matrix (PFM), 142
Position weight matrix (PWM)
 advantages of, 111–112
 discovery of binding-site motifs
  expectation maximization, 126–127
  Gibbs sampling, 127–128
  greedy alignments, 125–126
  "motif discovery" problem, 123, 124f
  pattern searches, 123–125
  pros and cons of methods, 123
 phylogenetic footprinting, 128, 129f
 sequence and functional modeling using, 112–113

specificity modeling
discriminatory models, 119
higher-order models, 121–122
probabilistic models, 113–119
regression models, 119–121
uses, 110–111
Probabilistic models for specificity
generative model, 115, 117
information content measurement, 118–119
known binding sites basis, 113–115
likelihood ratios and information content, 116–117
Profile HMM. *See* Hidden Markov model
Promoters, 4
Protein-binding microarrays (PBMs), 83, 148
Protein cleavage, 46
Protein–DNA complexes. *See* Protein structure; Sequence-specific interactions
Protein–DNA interactions
accessibility of genomic DNA, 9
action-at-a-distance rule for eukaryotes, 9
approaches to the study of, 10–11
division of labor between proteins and DNA, 1, 2
functions performed by proteins on DNA, 1–2
messenger RNA and, 2
regulation of gene expression
bacteriophage λ, 5–8
lac operon of *E. coli* and, 3–5
mystery of, 2–3
TFs and eukaryotic gene regulation, 9–10
TFs in prokaryotes versus eukaryotes, 8–9
transcription factors and, 2
Protein structure
allosteric effectors, 47
amino acid properties
nonpolar hydrophobic amino acids, 29, 30f, 31
polar acidic residues, 32
polar basic residues, 31
polar uncharged residues, 30f, 31
side-chain categories, 29, 30f
special cases, 32
structural function, 29, 30f
β strands and β sheets secondary structure, 33–34
determination methods, 36–39
families
classifications, 35
coiled-coil helix dimers, 41, 42f
helix-turn-helix proteins, 35, 39–41
recognition with β strands, 44–45
recognition with loops, 45
zinc-coordinating proteins, 41–44
functional domains, 34–35
α helix secondary structure, 32–33
levels, 27, 28f
modifications, 46–47
multiprotein complexes, 46
protein–DNA complexes, 39–41
protein sequence determination, 27–28
PWM. *See* Position weight matrix

R

RAR (retinoic acid receptor), 42
Recognition helix, 35, 41
Recognition models
binding specificity determination method, 147–148
with β strands, 44–45
focus on developing a predictive model, 146–147
lack of a recognition code and, 146
limitations of the recognition code, 149–150
limits to specificity determination, 149
with loops, 45
phage display method, 148–149
quantitative models, 150
recognition code for zinc finger proteins, 57–58
Registry of standard biological parts, 165
Regression models for specificity, 119–121
Regtransbase, 164
RegulonDB, 163
Relative entropy, 117, 119
Rel-homology domain, 45
Retinoic acid receptor (RAR), 42
Retinoid X receptor (RXR), 42
Ribosomes, 2
Romanuka, J., 61
RTIDE, 125
runt, 45
RXR (retinoid X receptor), 42

S

S (entropy), 90, 92–95, 117, 119
Saccharomyces cerevisiae, 9, 166–167
Saccharomyces Genome database (SGD), 166
Sarai, A, 80
SBML (Systems Biology Markup Language), 156
Sea urchin studies, 176–177, 178f, 179f
Seeman, NC, 146
SELEX (systematic evolution of ligands by exponential enrichment), 57, 83–84, 147, 148
SELEX-seq, 148
Sequence-specific interactions
lessons on specificity of TFs, 64
profiles of specificity
EcoRI, 53–54
Lac repressor, 61–63
Ndt80, 59–60
zinc finger proteins, 54f, 55–59
specificity of protein–DNA interfaces, 50–52
specificity's meanings, 49–50

Sequence-specific interactions (*Continued*)
  structures of nonspecific binding, 63–64
σ factors (sequence-specific binding proteins), 164
SGD (*Saccharomyces* Genome database), 166
Simple consensus sequence, 109–110
Single-stranded DNA (ssDNA), 15–17
Smith, Michael, 17
Smith–Waterman algorithm, 134, 137
SNP variants, 175
spbase database, 176
Specificity, binding. *See* Binding specificity
SPR (surface plasmon resonance), 72–73
ssDNA (single-stranded DNA), 15–17
STAT factors, 45
Strongylocentrotus purpuratus (sea urchin), 176–177, 178f, 179f
Structure of DNA
  accessible surfaces of base pairs, 19–21
  alternative structures, 24–26
  base pairs, 13–14
  DNA melting, 15–17
  implications of, 18
  major groove, 19–20
  minor groove, 20–21
  modified bases, 22
  potential symmetry of DNA sequences, 21–22
  principles of protein–DNA interactions and gene regulation, 18–19
  scaling *E. coli* up to human dimensions, 17–18
  sequence-dependent variation, 22–23
Surface plasmon resonance (SPR), 72–73
Synthetic biology, 165–166
Systematic evolution of ligands by exponential enrichment (SELEX), 57, 83–84, 147, 148
Systems Biology Markup Language (SBML), 156

T

Takeda, Y, 80
TAL (transcription activator-like), 45
TATA-binding protein (TBP), 44, 95
Tatum, Edward, 18
TFs. *See* Transcription factors
Thermodynamics of TF binding
  computational modeling, 100–102
  contributions of entropy and enthalpy, 94–95
  enthalpy of an interaction, 92
  entropy change in an interaction, 92–94
  free energy equation, 89–92
  heat capacity changes, 95–96, 97f
  kinetics of binding-site location, 102–105
  molecular contributions to complex formation
    direct and indirect readout, 100
    electrostatic and nonelectrostatic contributions, 98–99
    nature of interactions, 96, 98
    specific and nonspecific contributions, 99–100
Thymine (T), 13
Transcription activator-like (TAL), 45
Transcriptional genomics
  binding location analyses, 160–161
  conclusions, 183–184
  developments in DNA studies, 153–154
  DNA accessibility analyses, 162
  expression analyses, 157–160
  gene regulatory networks (*See* Gene regulatory networks)
Transcription factors (TFs)
  allosteric effectors, 47
  families
    classifications, 35
    coiled-coil helix dimers, 41, 42f
    helix-turn-helix proteins, 5, 35, 39–41
    recognition with β strands, 44–45
    recognition with loops, 45
  function, 2
  functional domains, 34–35
  modifications, 46–47
  multiprotein complexes, 46
Tryptophan, 47

V

Van der Waals contacts, 19, 20

W

Waterman, MS, 125
Watson, James D., 13
WEEDER, 125
Wieschaus, Eric, 177
Wilson, Allan, 173
Winged HTH subfamily, 40–41
Wolfe, S.A., 57
Wüthrich, Kurt, 37

X

X-ray crystallography, 36–37

Z

Z-DNA, 25
Zif268, 55–56, 58–59
Zinc cluster, 42–43
Zinc finger domain, 34, 41–42
Zinc finger proteins, 54f, 55–59, 146–147
Zuckerkandl, E, 172