Gerhard Krauss

Biochemistry of Signal Transduction and Regulation

Third, Completely Revised Edition

WILEY-VCH Verlag GmbH & Co. KGaA

Gerhard Krauss

Biochemistry of Signal Transduction and Regulation

Related Titles from Wiley-VCH

H.-D. Höltje / W. Sippl / D. Rognan / G. Folkers

Molecular Modeling – Basic Principles and Applications (2nd Ed.)

2003, ISBN 3-527-30589-0

C.W. Sensen

Essentials of Genomics and Bioinformatics

2002. ISBN 3-527-30541-6

T.K. Lindhorst

Essentials of Carbohydrate Chemistry and Biochemistry (2nd Ed.)

2002, ISBN 3-527-30664-1

N. Sewald, H.-D. Jakubke

Peptides: Chemistry and Biology

2002. ISBN 3-527-30405-3

E. Keinan / I. Schechter / M. Sela (Eds.)

Life Sciences for the 21st Century

2003, ISBN 3-527-30588-2

Gerhard Krauss

Biochemistry of Signal Transduction and Regulation

Third, Completely Revised Edition

WILEY-VCH Verlag GmbH & Co. KGaA

Prof. Dr. Gerhard Krauss
Laboratorium für Biochemie
Universität Bayreuth
95440 Bayreuth
Germany
Gerhard.Krauss@uni-bayreuth.de

1st edition 1999
2nd edition 2001
3rd edition 2003
1st, corrected reprint 2005

Cover illustration by Hanno Krauss, Bayreuth

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 Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>

Printed in the Federal Republic of Germany

Printed on acid-free paper

Composition Mitterweger & Partner, Kommunikationsgesellschaft mbH, Plankstadt
Printing Druckhaus Darmstadt GmbH, Darmstadt
Bookbinding Litges & Dopf Buchbinderei GmbH, Heppenheim

ISBN-13: 978-3-527-30591-9
ISBN-10: 3-527-30591-2

For Silvia, Julia, Hannes, and Enno

Preface

This book has originated from lectures on regulation and signal transduction that are offered to students of biochemistry, biology and chemistry at the University of Bayreuth. The idea to write a book on signal transduction was born during the preparations of these lectures where I realized that it is extremely difficult to achieve an overview of the area of signal transduction and regulation and to follow the progress of this field. The first book appeared in 1997 and was written in German. It was soon substituted by two successive English editions that are now followed by the 3rd edition which includes data and references up to 2002.

Cellular signaling in higher organisms is a major topic in modern medical and pharmacological research and is of central importance in the biomolecular sciences. Accordingly, the book concentrates on signaling and regulation in animal systems and in man. Plant systems could not be considered, and results from lower eukaryotes and prokaryotes are only cited if they are of exemplary character. The enormous increase in data on signal transduction has led me to leave out the chapter on ion channels and nerve signaling found in the former editions. This topic has since evolved into a huge research area of its own that could not be considered adequately within this book.

Our knowledge of signal transduction processes has exploded in the past 10 to 15 years, and the basic principles of intra- and intercellular signaling are now quite well established. Signaling processes can be described nowadays more and more on a molecular level and structure-function relationships of many central signaling proteins have been worked out. Research on signal transduction is presently focused on the characterization of the distinct cellular functions of the huge number of different signaling proteins and their subspecies, on the supramolecular organization of signaling proteins and on the interplay between different signaling pathways. The enormous complexity of signaling systems revealed by these studies makes it increasingly difficult to write a book that provides a truly comprehensive overview on signal transduction and considers all of the major new achievements. In consequence, not all branches and fields of signal transduction could be treated here with the same thoroughness.

It is the aim of the present book to describe the structural and biochemical properties of signaling molecules and their regulation, the interaction of signaling proteins at

the various levels of signal transduction and to work out the basic principles of cellular communication. Numerous studies in very diverse systems have revealed that the basic principles of signaling and regulation are similar in all higher organisms. Therefore, the book concentrates on the best studied reactions and components of selected signaling pathways and does not attempt to describe distinct signaling pathways (e.g. the vision process) in its entirety. Furthermore, results from very different eucaryotic organisms and tissues have been included. Due to the huge number of publications on the topic, mostly review articles are cited. Only a few original articles have been selected on a more or less subjective basis.

I am grateful to all people who have encouraged me to continue with the book and who have supported me with many helpful comments and corrections. In first place I want to thank my colleague Mathias Sprinzl and my former coworkers Thomas Hey, Carl Christian Gallert and Oliver Hobert. I am also grateful to Hannes Krauss and Yiwei Huang for the figures and structure representations.

Bayreuth, June 2003 Gerhard Krauss

Contents

1
The Regulation of Gene Expression

1.1
Regulation of Gene Expression: How and Where? A Schematic Overview

The transfer of genetic information from the level of the nucleic acid sequence of a gene to the level of the amino acid sequence of a protein or to the nucleotide sequence of RNA is termed gene expression. The entire process of gene expression in eucaryotes includes the following steps:
– *transcription*: formation of a primary transcript, the pre-mRNA
– *conversion of the pre-mRNA into the mature mRNA*: includes processing, splicing, transport from the nucleus to the cytosol
– *translation*: synthesis of the protein on the ribosome.

The expression of genes follows a tissue- and cell-specific pattern, which determines the function and morphology of a cell. In addition, all development and differentiation events are characterized by a variable pattern of gene expression. The regulation of gene expression thus plays a central role in the development and function of an organism. Because of the multitude of individual processes which are involved in gene expression, there are many potential regulatory sites (Fig. 1.1).

Regulation of Transcription

At the level of transcription, it can be determined whether a gene is transcribed at a given point in time. The chromatin structure plays an important role in this decision. Chromatin structures exist that can effectively inhibit transcription and shut down a gene. This "silencing" of genes can be transient or permanent and is generally observed in development and differentiation processes. The regulated transcription of genes requires as an essential step a reorganization and modification of the chromatin, which is a prerequisite for the initiation of transcription and is influenced by epigenetic changes in the DNA in the form of methylation of cytidine residues. Following chromatin reorganization and modification, transcription initiation requires the selection of the target gene and formation of a transcription initiation complex at the starting point of transcription. A large number of proteins are involved in this step. The main components are the multisubunit RNA polymerase, general and specific transcription factors, and cofactors that help to coordinate the chromatin

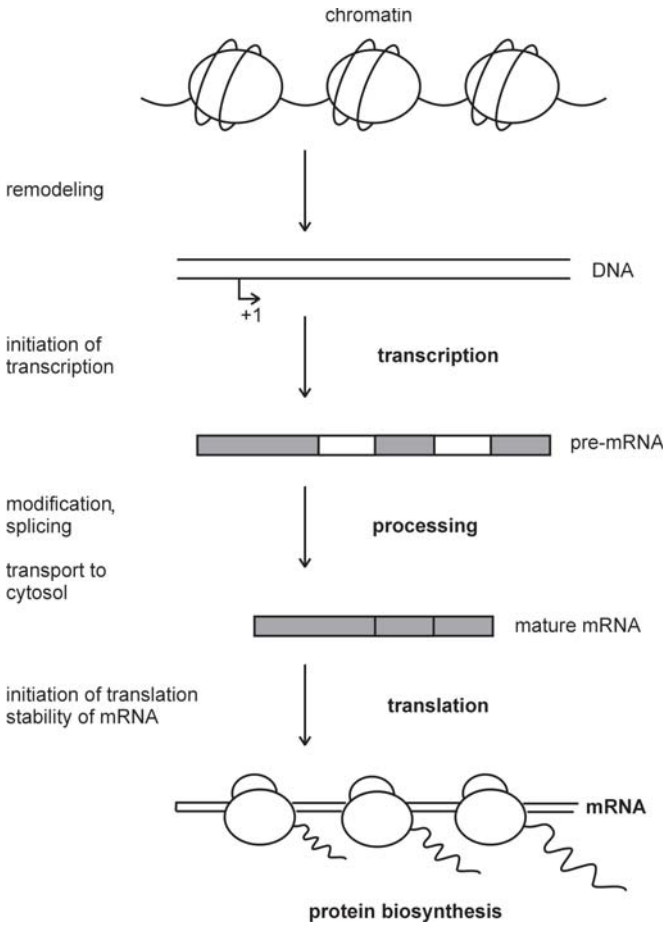

Fig. 1.1 Levels of regulation of eucaryotic gene expression.

structural changes and the process of RNA synthesis. The formation of a functional initiation complex is often the rate-limiting step in transcription and is subject to a variety of regulation mechanisms.

Conversion of the pre-mRNA into the Mature mRNA

Transcription of genes in mammals often initially produces a pre-mRNA, whose information content can be modulated by subsequent polyadenylation or splicing. Various final mRNAs coding for proteins with varying function and localization can be produced in this manner starting from a single primary transcript.

Regulation at the Translation Level

The use of a particular mature mRNA for protein biosynthesis is also highly regulated. The regulation can occur via the accessibility of the mRNA for the ribosome or via the

initiation of protein biosynthesis on the ribosome. In this manner, a given level of mature mRNA can specifically determine when and how much of a protein is synthesized on the ribosome.

Nature of the Regulatory Signals

Regulation always implies that signals are received, processed and translated into a resulting action. The nature of the signals which are employed in the course of the regulation of gene expression and are finally translated into a change in protein concentration can vary dramatically. Regulatory molecules can be small molecular metabolites, hormones, proteins or ions. The signals can be of external origin or can be produced within the cell. External signals originating from other tissues or cells of the organism are transferred across the cell membrane into the interior of the cell, where they are transduced by sequential reactions to the level of transcription or translation. Complex signal chains are often involved in the transduction.

1.2 Protein-Nucleic Acid Interactions as a Basis for Specific Gene Regulation

A recurring motif on the pathway of information transfer from gene to protein is the binding of proteins to DNA or RNA. At the DNA level, specific DNA-binding proteins aid in the identification of genes for regulation via transcriptional activation or inhibition. At the RNA level, specific RNAs are recognized in a sequence-specific manner to attain a controlled transfer of genetic information further on to the mature protein.

The basis of all specific regulation processes at the nucleic acid level is the recognition of nucleotide sequences by binding proteins. For the regulation of gene activity the specific binding of proteins to double-stranded DNA is of central importance. A specific DNA-binding protein usually recognizes a certain DNA sequence, termed the *recognition sequence* or *DNA-binding element*. Because of the enormous complexity of the genome, the specificity of this recognition plays a significant role. The binding protein must be capable of specifically picking out the recognition sequence in a background of a multitude of other sequences and binding to it. The binding protein must be able to discriminate against related sequences which differ from the actual recognition element at only one or more positions.

In the following, the basic features of specific recognition of DNA sequences by DNA-binding proteins will be presented.

1.2.1 Structural Motifs of DNA-binding Proteins

DNA-binding proteins contact their recognition sequences via defined structural elements, termed DNA-binding motifs. DNA-binding motifs are often found in structural elements of the protein which can fold independently from the rest of the protein and therefore represent separate DNA-binding domains.

The region of the binding protein which interacts with the recognition sequence often displays a characteristic small structural element which is stabilized through the help of other structural elements and is thereby brought into a defined position relative to the DNA. These structural elements contain short α-helical or β-sheet structures that in most cases contact the DNA sequence within the major groove: the dimensions of the major groove make it well suited to accept an α-helix. Accordingly, α-helices are often utilized as recognition elements. There are, however, examples of interactions with the minor groove of the double helix (TATA box-binding protein, see Section 1.2.3 and Fig. 1.9). We also know of DNA-binding proteins in which β-structures or flexible structures are involved in contact with the DNA.

The most common and well-characterized DNA-binding motifs can be characterized as described below.

Helix-turn-helix Motif

The helix-turn-helix motif (HTH motif) is – historically seen – the first DNA-binding motif whose structure could be solved in a complex with DNA. It is often found in bacterial repressors. Many eucaryotic DNA-binding proteins also utilize the helix-turn-

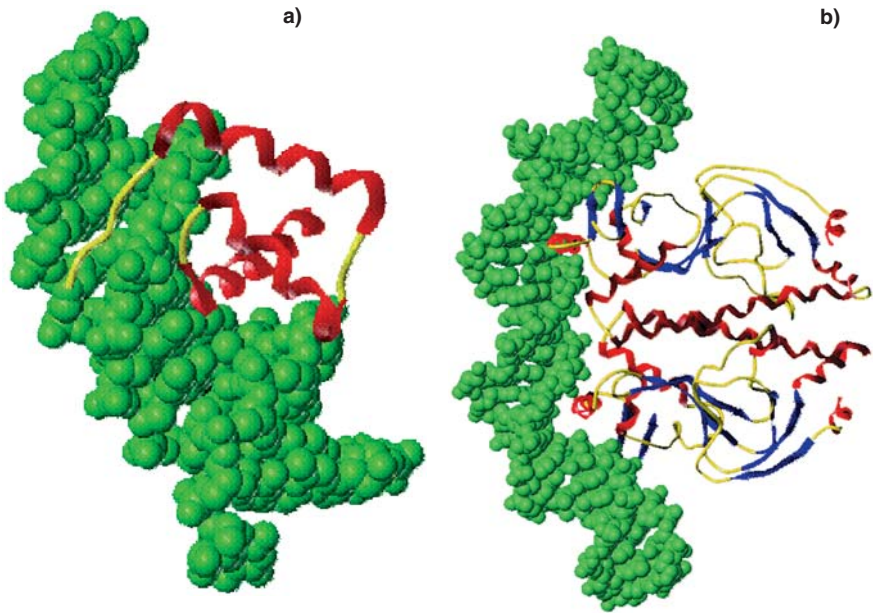

Fig. 1.2 The helix-turn-helix motif in complex with DNA

a) side view of the eucaryotic DNA binding motif of transcription factor MSX-1 in complex with DNA (Hovde et al., 2001). The homeodomain of MSX-1, which is a subgroup of helix-turn-helix motifs, shows a recognition helix (red) embedded in the large groove of the DNA (green). Two additional α-helices stabilize the arrangement of the recognition helix.

b) side view of the complex of a dimer of the catabolite activating protein (CAP) from *E. coli* in complex with DNA. This view displays the symmetrical embedding of two recognition helices, each from one CAP momomer, into the large groove of the DNA element. The DNA is bent nearly 90 deg in the complex. α-helices are in red, turns in yellow and β-strands in blue.

helix motif for specific binding on the DNA. Characteristic of the helix-turn-helix motif is the positioning of an α-helix in the major groove of DNA (Fig. 1.2). The recognition helix is connected by a turn to another helix, whereby the position of the recognition helix is fixed. The two helices occur at an angle of 120° to one other. The binding motif is usually stabilized by further helices of the same or another subunit. The detailed arrangement can differ significantly among the various helix-turn-helix motifs.

Binding Motifs with Zinc Ions

The zinc binding motifs contain Zn^{2+} complexed by four ligating Cys and/or His residues. Based on the stoichiometry of the complex, zinc fingers of the type Zinc-Cys_2His_2, Zinc-Cys_4 and $Zinc_2$-Cys_6 can be distinguished (Fig. 1.3).

The structures of two Zn-binding motifs are shown in Fig. 1.4. The zinc binding motifs play, above all, a structuring role by ensuring that a recognition helix is correctly oriented and stabilized. The zinc ion does not contact the DNA directly.

Fig. 1.5 shows the zinc binding motiv of Zif268, a regulatory DNA-binding protein of mice, in complex with DNA (Pavletich and Pabo, 1991). In Zif268, three zinc-fingers are arranged along the coil of the DNA. The DNA-binding element contains three repeats of the recognition sequence. This results in a modular construction of the protein, so that the periodicity of the DNA is reflected in the protein structure. Another example is found in the DNA-binding domain of the steroid hormone receptors which contains two non-equivalent zinc-Cys_4-motifs (see Section 4.3.2 and Fig. 4.6). In

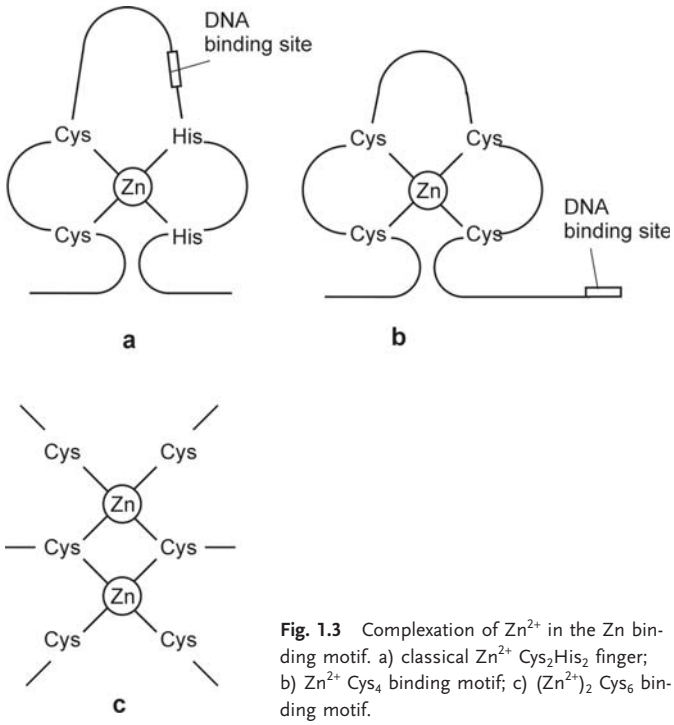

Fig. 1.3 Complexation of Zn^{2+} in the Zn binding motif. a) classical Zn^{2+} Cys_2His_2 finger; b) Zn^{2+} Cys_4 binding motif; c) $(Zn^{2+})_2$ Cys_6 binding motif.

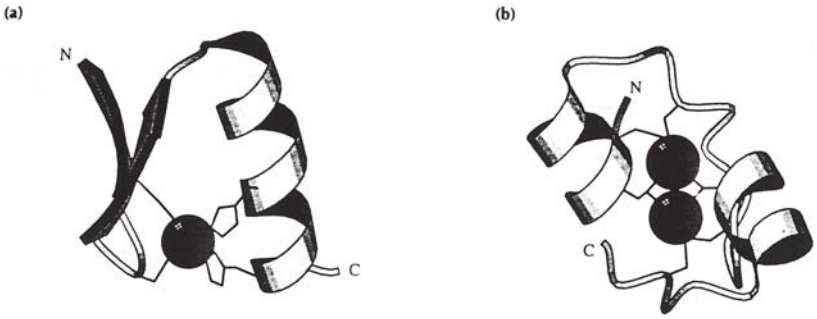

Fig. 1.4 Structures of Zn^{2+} binding motifs. a) TFIIIA-like Zn^{2+} Cys_2His_2 finger; b) the binuclear $(Zn^{2+})_2$ Cys_6 motif of the GAL4 transcription activator.

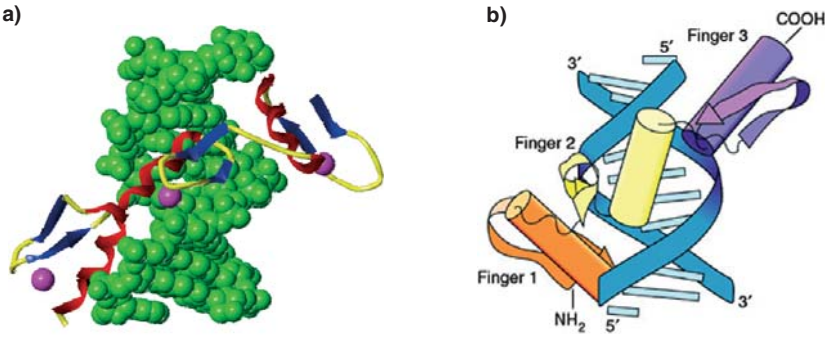

Fig. 1.5 The Zinc binding motif of Zif268 in complex with DNA
a) periodic arrangement of three Zn-fingers in the major groove of the DNA. Two of the Zn^{2+} ions are shown as violet spheres, the recognition helices are in red.
b) schematic drawing of the three Zn-fingers of Zif 268 in complex with DNA.

the transcriptional activator GAL4 of Yeast, two zinc ions are complexed by 6 Cys residues, whereby two of the Cys residues bind to both Zn^{2+} ligands (see Fig. 1.4).

Overall, the zinc-binding motifs display a great variety of structural diversity. The occurrence of a zinc-binding motif can often be predicted based solely on a characteristic series of Cys and His residues in a protein sequence. The complexation of a Zn^{2+} by His and Cys residues serves to bring the recognition element of the protein into a stable and unambiguous position relative to the DNA, thereby enabling specific contacts with the recognition sequence. It has to be pointed out that Zn-binding motifs are also found in many other proteins that do not act on DNA. One example is protein kinase C (see Section 7.4).

Basic Leucine Zipper

This group of binding motifs displays as characteristic structural element an extended bundle of two α-helices that are wound around each other in the form of a "coiled-coil". At their end is a basic region which mediates the DNA binding (review: Ellenberger,

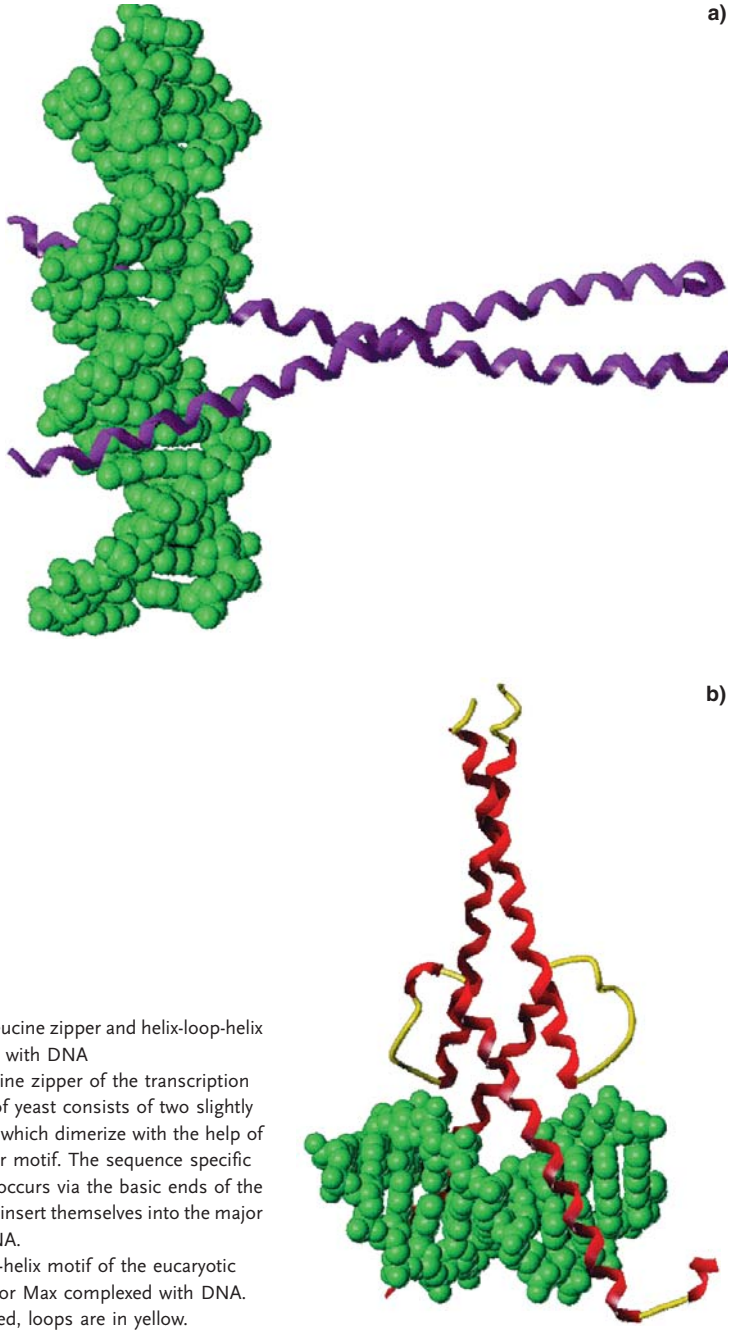

Fig. 1.6 Basic leucine zipper and helix-loop-helix motif in complex with DNA
a) The basic leucine zipper of the transcription activator GCN4 of yeast consists of two slightly curved α-helices, which dimerize with the help of the leucine zipper motif. The sequence specific binding of DNA occurs via the basic ends of the two helices. They insert themselves into the major groove of the DNA.
b) The helix-loop-helix motif of the eucaryotic transcription factor Max complexed with DNA. α-helices are in red, loops are in yellow.

1994). An example of the structure of a basic leucine zipper in complex with DNA is shown by the transcription factor GCN4 from yeast in Fig. 1.6A. *The leucine zipper takes its name from the regular occurrence of leucine residues (or other hydrophobic residues) in an α-helix*. A leucine or other hydrophobic amino acid is found at every seventh position of the helix. This sequential arrangement brings the hydrophobic residues all along one face of the helix, and the hydrophobic residues of two helices can interlock via hydrophobic interaction in a zipper-like manner. The leucine zipper is, above all, a tool to associate proteins in higher dimensions, whereby homodimers as well as heterodimers can be formed. The oligomerization of DNA-binding proteins is usually a prerequisite for strong binding to the cognate DNA element.

The leucine zipper itself does not participate in the recognition; it is only utilized for dimerization of the proteins. The N-terminal end of the basic leucine zipper motif is relatively unstructured in the absence of DNA. A helical structure is induced upon binding to DNA, allowing specific contacts to the recognition sequence. Dimer formation is a prerequisite for the exact positioning of the N-terminal basic end in the major groove of the DNA. Analogous to the dimeric structure of the protein, the DNA sequence displays 2-fold symmetry (see Section 1.2.4).

The Helix-Loop-Helix Motif

One example of the basic helix-loop-helix motif (HLH-motif) is found in the eucaryotic transcription factor Max (Fig. 1.6b and Section 14.3.3). The DNA binding occurs by a parallel bundle of 4 helices with two basic ends. As with the basic leucine zipper motif, the basic ends only attain a defined structure upon binding the DNA. The 4-helix bundle forms via dimerization of two subunits of the Max protein.

DNA Binding via β-Sheet Structures

β-sheet structures as DNA-binding motifs are found in pro- and eucaryotic DNA-binding proteins. Fig. 1.7 shows the structure of the eukaryotic transcription factor NFκB bound to its cognate DNA element. Noteworthy is the enshrouding of the DNA by the β-sheets of NFκB. The recognition of the DNA elements is achieved by interaction of the β-strands with the major groove of the DNA.

Flexible Structures in DNA-Binding proteins

A series of DNA-binding proteins utilize additional flexible structures aside from defined structural DNA-binding motifs in order to increase the stability and specificity of the complex. The λ repressor grabs around the DNA helix with the flexible N-terminal arm of the protein to contact the back of the helix. The basic region of the leucine zipper and HLH-binding proteins is a further example of the importance of protein flexibility in DNA binding. In the absence of DNA, the basic portion of this binding motif is poorly structured, and only following DNA binding is an α-helix formed in the basic region. The α-helix induced upon binding lies in the major groove of the DNA and establishes specific interactions with the recognition sequence.

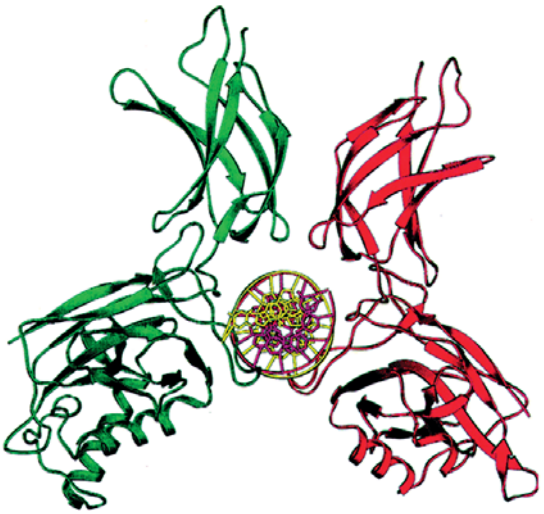

Fig. 1.7 The eucaryotic transcription factor NFκB in complex with DNA. Shown is the structure of the p50–p65 heterodimer of NFκB complexed with a specific NFκB DNA element. The view is along the DNA helical axis. Each of the subunits contains a bundle of β-sheets which envelops the DNA so that only the minor groove is exposed. The p50 subunit is shown in green and the p65 subunit in red The top strand of DNA is shown in pink, the bottom strand in yellow. From Chen et al., 1998; used with permission.

1.2.2
The Nature of the Specific Interactions in Protein-Nucleic Acid Complexes

The binding of a protein to nucleic acid is accomplished by weak, noncovalent interactions. The interactions are the same as those involved in the formation of the tertiary structure of a protein:
- hydrogen bonds (H-bonds)
- electrostatic interactions
- van der Waals interactions
- hydrophobic interactions.

H-bonds in Protein-Nucleic Acid Complexes

Of central importance for the formation of a specific protein-DNA complex are hydrogen bonds. The H-bonds are clearly identifiable in high-resolution structures. H-bonds occur where an H-bond donor and acceptor lie within 0.27–0.31 nm of each other. Energetically most favorable is the linear arrangement of the H-bond, with deviations from linearity leading to a reduction in energy. This characteristic is responsible for the stereospecific orientation of H-bond acceptors and donors. The H-bond thus contributes significantly to the spatial orientation between protein and nucleic acid.

There are many different H-bond donors as well as acceptors in proteins and nucleic acids which contribute to the specific recognition. Important H-bond donors and acceptors in proteins are Asn, Gln, Ser, Thr, Tyr, Glu, Asp, Arg, Lys, Cys and His. The peptide bonds of the backbone often participate also.

The heteroatoms and exocyclic functional groups of the bases within the nucleic acid can form H-bonds to residues of a binding protein, in addition to base pairing. Also, the oxygen of the ribose or deoxyribose and the phosphate moiety of DNA can be used as H-bond acceptors.

The available structural information on protein-DNA complexes reveals great variability and flexibility in the spectrum of H-bond interactions. Examples of the variety of H-bond interactions are shown in Fig. 1.8.

The following points are noteworthy:

- A base can be contacted by more than one amino acid residue. Furthermore, there are many examples of one amino acid residue, e.g. Arg, contacting two sequential bases. This type of interaction functions as a clip and maintains a spatially defined arrangement.
- The contact between protein and DNA can also be transmitted via bound water molecules.
- There are always numerous H-bond contacts formed between the recognition sequence and the binding protein. The pattern of H-bond donors and H-bond acceptors is determined by the sequence and conformation of the DNA as well as by the specific structure of the protein. Both together lay the foundation for a specific recognition of the DNA by the protein.
- An important factor in the structure of protein-DNA complexes can be the peptide backbone. The amide bond can function as an H-bond acceptor as well an H-bond donor. Because of the reduced flexibility of the backbone vs side chain (resonance stabilization of the peptide bond), H-bonds to the peptide backbone lead to a rigid and tight arrangement in the complex and contribute extensively to the exact fit between protein and nucleic acid.

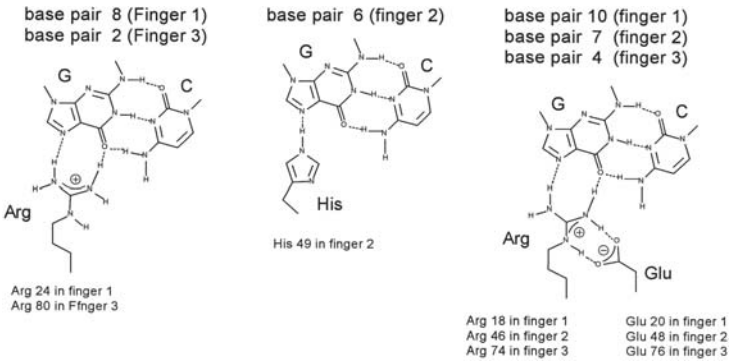

Fig. 1.8 H-bonds in the complex between the Zinc fingers of Zif268 with the cognate recognition helix. Zif268 contacts the DNA with three Zn-fingers (finger 1-3 in Fig. 1.5). Shown are some H-bond contacts formed between the fingers and the base pairs of the recognition sequence.

Ionic Interactions

Ionic interactions result from the electrostatic attraction or repulsion between charged groups. As opposed to H-bonds, ionic interactions are not directed and are effective over greater distances.

DNA presents itself to a binding protein as a negatively charged, anionic substrate. Accordingly, the protein displays a complementary positive potential resulting from an accumulation of basic amino acid residues. The electrostatic interaction between the two oppositely charged binding surfaces of DNA and protein make a significant energetic contribution to the formation of a stable complex.

The ionic interactions are, however, less suitable for distinguishing between various base pairs, since only the phosphates of the backbone from the DNA are involved in the interaction. Together with the specific H-bonds, the nonspecific ionic interactions contribute significantly to the formation of a stable complex. The positively charged surface of DNA-binding proteins is also the reason for the ability of many such proteins to bind DNA nonspecifically.

Van der Waals Contacts

The van der Waals contacts are a type of electrostatic interaction and arise from an interaction between permanent and/or induced dipoles in the bond pair. They are typically effective over a much shorter range than ionic interactions. The contribution of van der Waals contacts to the binding of a protein to a DNA sequence is difficult to estimate, since many small contributions must be considered. An example of a contact surface with many van der Waals interactions can be found in the complex of the TATA box-binding protein with the TATA box (see Fig. 1.9). In this case there are extensive van der Waals contacts between the sugar residues of the DNA backbone and the hydrophobic surface of the protein (Kim et al., 1993).

1.2.3 The Role of the DNA Conformation in Protein-DNA Interactions

The double helix of the DNA can only to a first approximation be considered a linear, rod-like structure with the typical coordinates of B-DNA. Actually, DNA possesses considerable flexibility and conformational variability. The flexibility and structural polymorphism of DNA are prerequisites for many of the regulatory processes on the DNA level (review: Alleman and Egli, 1997). Local deviation from the classical B-structure of DNA, as well as bending of the DNA, are observed in most protein-DNA complexes.

Local Conformational Changes of DNA

Crystal structures of DNA have shown that, apart from the structural motifs of the A-, B- and Z-forms of DNA, other, sequence-dependent structural variations exist which are observed when smaller sequence fragments are examined in detail. The structural variations can affect the width of the major groove, the extent of base stacking, and the tilt of the base pairs to each other. The local conformational changes are sequence

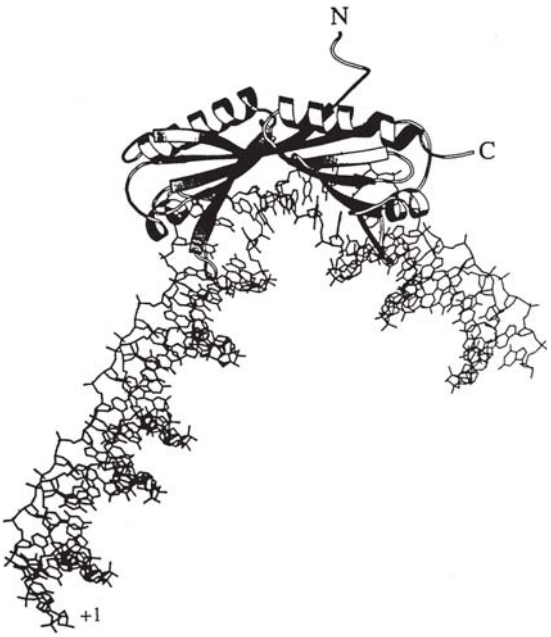

Fig. 1.9 Bending of DNA in the TATA box. The DNA is kinked in the complex of the TATA box binding protein (yeast) with the 8 base pair TATA box (Kim et al., 1993). The DNA is deformed in the region near the kink: the minor groove, which faces the protein, is clearly widened.

dependent and can be intrinsic properties and thus permanent occurrences; they can, however, also be induced by protein binding.

In most protein-DNA complexes, analysis of DNA structure in the region of contact with the binding protein reveals distinct divergence from the parameters of classical B-DNA structure. A specific sequence-determined conformation of the DNA is often a prerequisite for a specific recognition.

Bending of DNA

If one traces a longer stretch of a DNA molecule in solution, a clear divergence from linearity becomes evident. Thermally induced structural fluctuations allow a bending of DNA, which is why long DNA molecules are described as a random coil. This bending of the DNA occurs in molecules with a length of more than ca. 200 bp.

Bending of shorter fragments is observed in the presence of distinct sequence characteristics or upon binding of proteins. An intrinsic bending of short DNA fragments is induced when the DNA contains short dA-repeats (e.g. dA_5), and this bending can be enforced by protein binding.

Protein-induced Bending of DNA

There are numerous examples of protein-induced bending of DNA. The bending of a short segment of DNA (150-200 bp) leads to a loss of stacking interactions of the π-electron system of neighboring bases and is energetically unfavorable. Stacking interactions arise from interactions of the π-electron systems of bases atop one another and contribute extensively to the stability of the double helix. An active bending of a short