

Investigation of Interactions between Homeodomain Proteins and DNA

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Darius Vainius
aus Sakiai, Litauen

Göttingen 2004

D7

Referent: Prof. Dr. E. Neher

Koreferent: Prof. Dr. W. Lauterborn

Tag der mündlichen Prüfung: 18.05.2004

*To Ruta
and to my parents.*

Abbreviations

ADC	analog-to-digital converter
a.u.	arbitrary units
Antp	Antennapedia
Bcd	Bicoid
bp	base pair
CD	circular dichroism
DNA	deoxyribonucleic acid
DTT	dithiothreitol
DMSO	dimethylsulfonic acid
<i>E. coli</i>	<i>Escherichia coli</i>
En	Engrailed
eve	even-skipped
Eq.	equation
FCS	fluorescence correlation spectroscopy
Fig.	figure
FRET	fluorescence resonance energy transfer
ftz	fushi tarazu
GFC	gel filtration chromatography
HEPES	N-(2-Hydroxyethyl)-piperazin-N'-2-Ethansulfonic acid
HLH	helix-loop-helix
HTH	helix-turn-helix
IAEDANS	5-(((2-iodoacetyl)amino)ethyl)amino) naphthalene-1-sulfonic acid
kDa	kilodalton
NMR	nuclear magnetic resonance
mBBr	monobromobimane
PAGE	polyacrylamide gel electrophoresis
pH	negative decimal logarithm of H ₃ O ⁺ ion concentration
PMT	photomultiplier tube
prd	paired
RISC	reduced instruction set computer
SDS	sodiumdodecylsulfate
TBP	TATA-binding protein
Ubx	Ultrabithorax
v/v	volume/volume

Abbreviations of amino acids:

A	(Ala)	Alanine
C	(Cys)	Cysteine
D	(Asp)	Aspartic acid
E	(Glu)	Glutamic acid
F	(Phe)	Phenylalanine
G	(Gly)	Glycine
H	(His)	Histidine
I	(Ile)	Isoleucine
K	(Lys)	Lysine
L	(Leu)	Leucine

M	(Met)	Methionine
N	(Asn)	Asparagine
P	(Pro)	Proline
Q	(Gln)	Glutamine
R	(Arg)	Arginine
S	(Ser)	Serine
T	(Thr)	Threonine
V	(Val)	Valine
W	(Trp)	Tryptophan
Y	(Tyr)	Tyrosine

Contents

1. INTRODUCTION	1
2. REVIEW OF THE LITERATURE.....	3
2.1 Specific protein-DNA interactions	3
2.1.1 Forces between proteins and nucleic acids	3
2.1.2 Structural tools of binding – DNA-binding motifs.....	5
2.1.3 Common principles	8
2.1.4 Changes in structure and dynamics of protein and DNA induced by interaction	9
2.1.5 Thermodynamics and kinetics of Protein-DNA interaction.....	15
2.1.6 Influence of environmental factors	22
2.2 Aspects of homeodomain-DNA interaction.....	26
3. AIMS OF THE PRESENT STUDY	30
4. MATERIALS AND METHODS	31
4.1 Materials	31
4.2 Protein preparation.....	31
4.3 DNA preparation.....	33
4.4 Steady state absorption spectroscopy.....	34
4.5 Circular dichroism	34
4.6 Steady state fluorescence spectroscopy	35
4.7 Fluorescence titrations	36
4.8 Measurements of fluorescence resonance energy transfer (FRET).....	37
4.9 Stopped-flow kinetics	40
4.10 Fluorescence correlation spectroscopy (FCS)	42
4.11 Molecular modeling.....	46
5. RESULTS	47
5.1 Spectroscopic characterization of homeodomain-DNA complexes at equilibrium.....	47
5.1.1 Changes in intrinsic fluorescence of homeodomains upon binding to DNA	47
5.1.2 Changes of circular dichroism spectra of homeodomains upon binding to DNA ..	49
5.1.3 Use of fluorescent labels to study the homeodomain-DNA interaction.....	51
5.2 Analysis of homeodomain-DNA binding at equilibrium conditions.....	54

5.3 Quantitative analysis of FRET between the fluorescein and Cy3 labels linked to the DNA	58
5.4 Analysis of the translational diffusion properties of the double-labeled oligonucleotide bFC and the Bcd-oligonucleotide complex using fluorescence correlation spectroscopy (FCS).....	61
5.5 Investigation of the homeodomain-DNA interaction kinetics	62
6. DISCUSSION	69
7. SUMMARY	78
APPENDIX	80
A1. Function of Ultrabithorax and Engrailed homeodomain proteins	80
A2. Calculation of the kinetic fluorescence intensity profiles	80
A3. Numerical methods of kinetic data analysis	81
REFERENCES	84

1. Introduction

The information required by every living organism for obtaining and maintaining its structures and functions is encoded in its DNA. At various levels of information handling such as storage (DNA packing into chromatin), maintenance (e.g. repair of damaged DNA), copying (replication) and transfer (i.e. the expression of genes involving transcription and translation) protein-nucleic acid interactions play a central role. Particularly important for metabolism, replication and development of each organism is the expression of its genes as proteins at correct locations, in proper amounts and with correct timing, relative to cellular and developmental cycles. Gene expression is primarily controlled at the level of transcription by the proteins that bind to specific DNA regulatory regions (operators, enhancers etc.) acting as activators or repressors [Jacob and Monod 1961]. These proteins, called transcription factors, constitute one of the largest and most diverse classes of DNA-binding proteins. Therefore, the elucidation of molecular details of specific interactions between transcription factors and DNA is critical for understanding the mechanisms involved in the control of the expression of genetic information.

Structural studies of transcription factors have identified a number of different DNA-binding domains (see chapter 2.1.2), the molecular platforms upon which the protein components of the complementary recognition are positioned in space. Some of these domains form independently folding protein substructures, the DNA-binding properties of which often resemble those of the full protein [Struhl et al. 1989; Wilson et al. 1996]. If other regions of the protein do not influence these properties, the object of investigation can be reduced to the DNA-binding domain and studied separately.

The present study focuses on the analysis of the DNA-binding properties of one of such domain present in homeotic transcription factors, the homeodomain (chapter 2.2). Homeodomain-containing proteins play fundamental roles at early developmental stages of eukaryotes [Gehring et al. 1994]. These proteins control the pattern formation and determine the identity of body segments by governing the choice between alternative developmental pathways depending on spatial position in the embryo [Gehring et al. 1994]. Such positional information in the fruit fly (*Drosophila melanogaster*) embryo is provided by the exponential antero-posterior gradient of the maternal morphogen protein Bicoid (Bcd), a homeodomain-containing transcription factor. This protein activates a number of target genes at certain threshold levels of concentration and thus initiates position-specific developmental programs,

which are controlled by the products of these target genes (mostly other homeodomain-containing transcription factors). The ability of this protein to bind to specific DNA regions and activate the appropriate genes at concentrations varying by approximately two orders of magnitude [Ephrussi and St Johnston 2004] is one of the major albeit unresolved control mechanisms of early embryonic development. This fact stimulated the choice of the Bcd homeodomain as the central topic for this thesis.

The analysis of specific protein-DNA interactions addresses some important questions: How does the protein locate and discriminate between the specific binding target among millions of competing nonspecific sites? Which kinetic pathways does it use in searching for the specific site? What types of interaction and which structural elements of the macromolecules are involved in specific recognition? What determines the stability of specific protein-DNA complexes? Achieving answers to these questions necessitate structural, kinetic and thermodynamic studies, which have to be combined in order to obtain a complete image of the function of the DNA-binding protein. A large body of work has been performed over the last two decades in studying the genetic, structural and biochemical aspects of homeodomain-DNA interactions (for review see [Gehring et al. 1994; Billeter 1996; Wolberger 1996]). However, very little is known so far about the kinetics of homeodomain-DNA binding. This is related to the fact that the vast majority of *in vitro* studies of homeodomain-DNA interactions [Affolter et al. 1990; Ekker et al. 1991; Ades and Sauer 1994; Burz et al. 1998] rely on equilibrium binding measurements based mostly on electrophoretic mobility shift assays [Sambrook et al. 1989]. Although these can yield qualitative and quantitative (e.g. equilibrium binding constant) information about the binding affinity, the analysis of homeodomain-DNA binding at equilibrium does not provide information about intermediate steps of interaction. Furthermore, the binding equilibrium, measured using the electrophoretic techniques is biased by the electrostatic field acting on the protein-DNA complex. These problems were overcome in this study by the use of transient kinetic techniques such as stopped-flow in parallel with spectroscopic measurements. Some new insights into the kinetic mechanisms of homeodomain-DNA interactions were gained with these techniques.

2. Review of the literature

2.1 Specific protein-DNA interactions

2.1.1 Forces between proteins and nucleic acids

The intermolecular forces that determine how proteins interact with DNA can be classified into four types: electrostatic, dipolar, dispersion force and the hydrophobic effect of solvation.

Electrostatic forces are involved in interactions between groups of opposite charge that are called “salt bridges”. They typically provide about -40 kJ/mol stabilization energy and occur between the ionized phosphates of the nucleic acids and either the ϵ -ammonium group of lysine, the guanidinium group of arginine, or the protonated group of imidazole of histidine in the protein. Compared to other forces between proteins, and nucleic acids, salt bridges are relatively long range and rather insensitive to the relative orientation of the charges. They therefore are considered to contribute primarily to nonspecific protein-DNA binding.

The force of dipole-dipole interaction is the physical basis of hydrogen bonds that form between a partially negatively charged oxygen or nitrogen (bond acceptor) and a partially positively charged hydrogen atom (bond donor). Nucleic acids present numerous functional groups that can serve as hydrogen bond donors or acceptors. These include oxygens and amide groups of the bases as well as the phosphodiester oxygens of the DNA backbone. At the “side” of the protein, the appropriate groups for hydrogen bonding can be provided by amides and carbonyls of the peptide backbone as well as by most amino acid side chains. Hydrogen-bonding interactions can also be mediated by water molecules whose positions and orientations are fixed by simultaneous hydrogen-bonding to both the protein and the DNA. The strength of the hydrogen bond declines with the inverse third power of the donor-acceptor distance and also decreases greatly if the bond is bent (i.e., if there is an angle between the dipoles). These features account for the sensitivity of the hydrogen bonding interactions to the conformation and flexibility of the DNA and, furthermore, for its contribution to sequence-specific DNA recognition. Hydrogen bonding is one of the most abundant and important interactions in protein-DNA complexes, providing both sequence-specific and non-sequence-specific contacts between protein and DNA.

Specific protein-DNA complex formation requires a close fit between the surfaces of the protein and DNA molecules. A close apposition of neutral atoms causes a correlation between their electron distributions, resulting in an attractive dispersion force. Since this type of interaction is essentially an induced-dipole–induced-dipole interaction, the magnitude of the associated potential energy scales with the inverse sixth power of the distance between the interacting surfaces. Thus, large neutral surfaces, such as thymine methyl groups on DNA are sites of potential van der Waal’s interactions of this type with the uncharged side-chains on proteins. Such interactions with deoxyribose rings can contribute to a nonselective DNA binding. In some cases, disruption of the stacking of bases in the DNA helix allows contacts between the aromatic rings of the bases and hydrophobic residues, causing a large dispersion force. In general, dispersion forces between proteins and DNA are relatively small and influence mainly DNA binding specificity much more than binding strength (affinity).

The hydrophobic effect of solvation, which is often considered to originate from so-called “hydrophobic interaction”, is rather a “virtual force”. It does not represent any particular type of interaction, but is rather a thermodynamic consequence of water behavior at the surface of the macromolecule in aqueous solution. Any molecule in water creates a sharply curved interface and thus arranges a layer of water molecules around itself, thus decreasing the entropy and changing the enthalpy of the system. Interactions between protein and DNA involve large changes in the organization of water at the surfaces of these macromolecules. These changes are caused by water-macromolecule and water-water interactions. Water associated with polar surfaces is hydrogen-bonded to donor or acceptor groups on the surface. Removal of such water requires the breaking of these hydrogen bonds and their replacement with hydrogen bonds to other water molecules. This desolvation of polar surfaces generally increases the enthalpy. In contrast, the transfer of water molecules associated with nonpolar surfaces to the bulk phase decreases the enthalpy. The entropy changes caused by the displacement of bound water constitute an even larger contribution to the thermodynamics of many protein-DNA complexes [Kerppola 2002]. Water molecules near nonpolar surfaces are more constrained in motion and orientation than water molecules in bulk solution. Thus, removal of nonpolar surfaces from solution through complex formation releases water molecules into bulk solution, resulting in higher entropy. This effect promotes the association of hydrophobic groups and is therefore called *the hydrophobic effect*.

2.1.2 Structural tools of binding – DNA-binding motifs

Although it is not possible to cover the whole spectrum of DNA-binding proteins in such a brief introduction, considering DNA-binding motifs as families simplifies the understanding of the structural features involved in protein-DNA binding and thus helps the classification of various types of protein-DNA recognition.

Many DNA-binding proteins recognize specific sites through small, discrete domains – independently folding substructures that protrude from the surface of the DNA-binding protein and penetrate the DNA grooves. In some cases these domains can be interchanged between proteins. On the basis of structural studies and comparisons of primary sequences, several families of DNA-binding proteins have been identified, whose members share the same DNA-binding motifs. Large, well-established families include helix-turn-helix (HTH) proteins, the homeodomains (HDs), zinc finger (coordinating) proteins, the steroid receptors, leucine zipper proteins and the helix-loop-helix proteins. There are also two smaller families that use β -sheets for DNA binding, but they are beyond the scope of this brief introduction (for a review see [Luscombe et al. 2000]).

The helix-turn-helix motif is the most common and most thoroughly studied motif found in DNA-binding proteins in both prokaryotes and eukaryotes. This motif contains about 20 amino acids that form two α -helices interspaced by a short four-residue turn, which keeps the helices at a relatively fixed angle. The second of the two α -helices, referred to as the recognition helix, inserts into the major groove and forms both base and sugar-phosphate backbone contacts (Figure 2.1 A). The first helix, while not embedded in the major groove, in some cases makes additional DNA contacts. Typical representatives of this family are prokaryotic transcription factors, like: Lac, Trp, 434 and lambda repressors, as well as the lambda Cro and CAP proteins. It is worth noting here that the HTH motif, unlike many other motifs, is not a separate stable unit. The HTH motif cannot fold or function by itself but always occurs as part of a larger DNA-binding domain.

A good example of such a domain is the homeodomain, which is sometimes considered a separate family of DNA-binding motifs. The homeodomain typically has 60 amino acids and folds into three α -helices, of which the second and third helix resemble the HTH motif (Figure 2.1 B). Unlike the isolated HTH unit, the homeodomain forms a stable folded structure and can bind DNA by itself [Sauer et al. 1988; Qian et al. 1989; Affolter et

al. 1990]. The third helix, which is more extended, docks into the major groove of DNA, making most of the specific contacts. The rest of the critical contacts are achieved by the extended N-terminal arm, which embraces the DNA and fits into the minor groove. Further aspects of homeodomain-DNA interaction will be covered in chapter 2.2.

Protein domains with one or more coordinated zinc ions at their core form a superfamily of eukaryotic DNA-binding proteins. In all cases, zinc plays only a structural role in maintaining the protein fold and does not interact with the DNA. There are significant differences in the overall fold and DNA binding of zinc-coordinating DNA-binding domains within this big superfamily. I will focus only on two most significant and abundant subfamilies: zinc fingers and steroid receptors. Members of the zinc finger family typically contain tandem repeats of a compact, ~30 amino acids containing DNA-binding motif, connected by short linker regions.

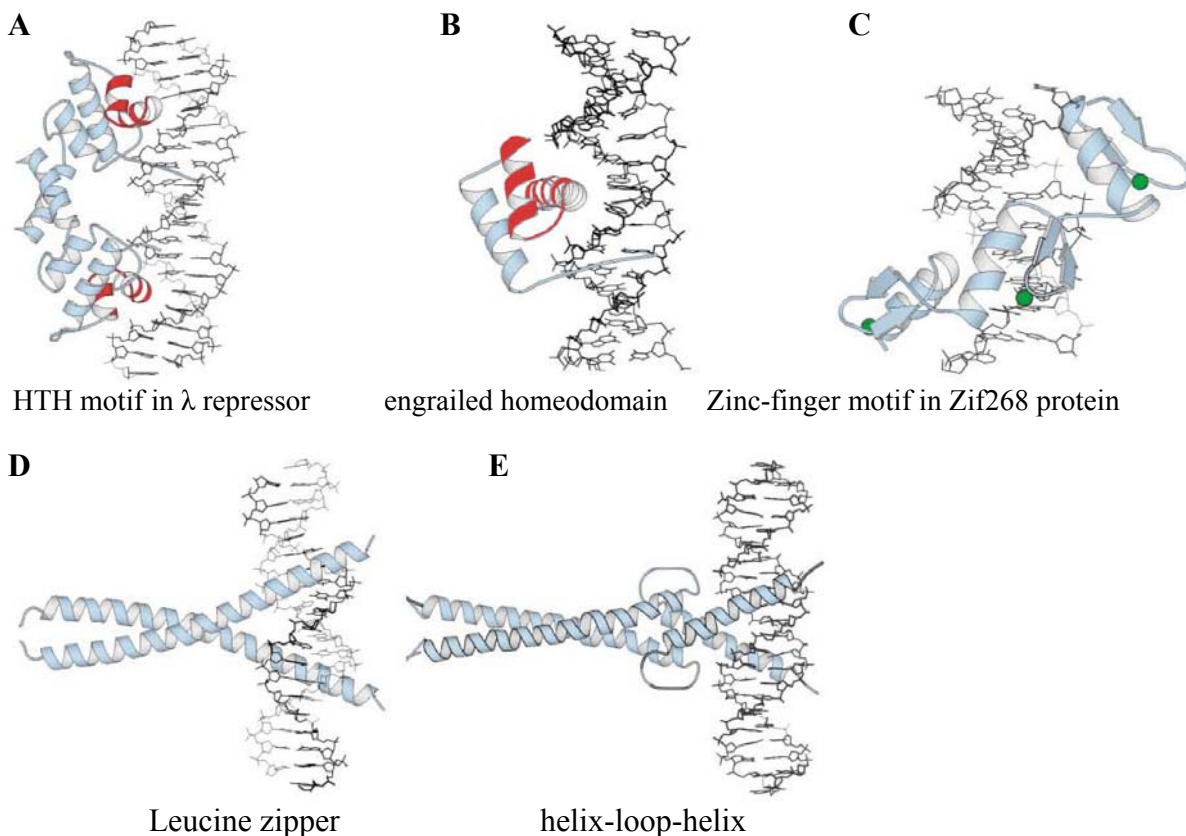


Figure 2.1. DNA-binding motifs. HTH motifs are highlighted. Small dark spheres in the Zinc-finger motif are Zn atoms.

These motifs have the sequence pattern Cys-X₂-4-Cys-X₁₂-His-X₃-5-His, which folds to a domain containing a relatively short α -helix, two antiparallel strands of the beta sheet, and a

core Zn^{2+} ion coordinated by the two cysteine and two histidine residues [Pavletich and Pabo 1991] (Figure 2.1 C). The binding pattern of the zinc finger is quite similar to that of the HTH proteins: the α -helix is inserted into the major groove of DNA and makes specific contacts, recognizing a site consisting of 3 to 4 base pairs. For a specific DNA recognition at least two successive fingers tracking the DNA major groove are necessary, whereas one zinc finger can bind the DNA only nonspecifically.

The DNA-binding domain of steroid receptor proteins has approximately 70 residues. It contains 8 conserved cysteines, which serve as coordinating points for two Zn^{2+} ions. The structural data of the well-analyzed glucocorticoid and estrogen receptors [Hard et al. 1990; Schwabe et al. 1990] show that each of these peptides folds into a single globular domain with a pair of α -helices. The two extended helices are roughly perpendicular and are held together by hydrophobic contacts. A zinc ion binds near the start of each helix and holds a peptide loop against the N-terminal end of the helix. These proteins, as with the other domains mentioned above, bind DNA by inserting the α -helix into the major groove and making the specific contacts with the edges of the base pairs. In contrast to the modularity of the zinc fingers, steroid receptors bind DNA either as homodimers or heterodimers.

Yet another mode of DNA binding is represented by the leucine zipper and helix-loop-helix (HLH) proteins, which have a common mechanism of DNA binding, but differ in the dimerisation modes they apply. Leucine zipper proteins consist of long, uninterrupted α -helices of about 60 residues. The amino acid sequences of these proteins exhibit a unique feature: in a region of about 30–40 residues, every seventh residue is a leucine. Since the helical repeat of the α -helix is 3.6 residues per turn, the leucines are located at approximately two-turn intervals, and on the same side of the helix, forming the dimerisation region of the zipper-like pattern. Leucine zipper proteins form parallel oriented coiled-coil homodimers and heterodimers associating via this dimerisation region, which is located in a C-terminal half of the α -helix [O'Shea et al. 1991] (Figure 2.1 D). The N-terminal portions of the dimerized helices splay and insert into the major groove on either side of the DNA [Ellenberger et al. 1992] (Figure 2.1 D). An interesting feature of these proteins is that the helical structure of the entire DNA sequence reading head is coupled to DNA binding, as these residues are unstructured in the absence of DNA [Weiss et al. 1990]. The helix-loop-helix proteins share a very similar mode of DNA binding with the leucine zipper proteins (Figure 2.1 E). The salient difference lies in the dimerisation region, which in the HLH proteins is composed of two helices separated by a loop.

2.1.3 Common principles

Diversity of known DNA-binding motifs and contacts suggests that there are no simple rules or patterns for describing site-specific recognition. Nevertheless, it is possible to make some generalizations comparing known structures of protein-DNA complexes [Pabo and Sauer 1992; Luscombe et al. 2000].

Specific recognition always involves a set of contacts with the bases and with the DNA backbone [Pabo and Sauer 1992]. Most of critical contacts are achieved by the protein side chains via hydrogen bonds, which are very important for site-specific recognition (although other types of interaction occur, too). Contacts with DNA backbone usually also involve salt bridges with the phosphodiester oxygens. These contacts may serve as “reference marks” that stabilize the protein against the bases and thereby enhance the specificity of the side chain-base interactions.

There is no simple recognition “code” consisting of a one-to-one correspondence between protein residues and the bases of DNA they contact. Many amino acid side-chains can interact with more than one type of base and any given type of base can be contacted by different side-chains. Often, more than one side-chain contacts a given base, and in other instances a single side-chain may contact more than one base pair simultaneously. Nevertheless, a lot of work has been devoted to trying to identify and understand the mechanisms and origins of “typical” specific side chain-base pair interactions. Since the “classical” specific residue-base pairings (arginine with guanine and glutamine or asparagine with adenine) were predicted by Seeman, Roenberg and Rich [Seeman et al. 1976] and confirmed later by others [Pabo and Sauer 1992; Mandel-Gutfreund et al. 1995], many structures of protein-DNA complexes were published ([Luscombe et al. 2000; Pabo and Nekludova 2000], see references therein) and an extensive and comprehensive analysis was performed [Kono and Sarai 1999; Pabo and Nekludova 2000; Luscombe et al. 2001], leading to some common principles. The “classical” set of specific side chain-base interaction pairs mentioned above was extended by the lysine-guanine, which seems to be as common as arginine-guanine contact in specific protein-DNA complexes [Luscombe et al. 2001]. Furthermore, a set of rules governing the interactions seen for members of some families, in particular the zinc-finger proteins [Suzuki and Yagi 1994; Choo and Klug 1997] may exist. Because of their lack of directional requirements, Van der Waals interactions are thought to play a lesser role in specificity. Nevertheless, the high proportion of van der Waals contacts

found at most protein-DNA interfaces [Luscombe et al. 2001] imposes steric constraints on the types of side chains and bases that can be accommodated at particular positions, thereby also playing a role in sequence selectivity. Since the DNA-binding motifs not only have a conserved folding, but in many cases also incorporate a conserved docking mechanism and conserved set of contacts [Pabo et al. 1990], it is thought that folding and docking of entire protein can help to control the purpose that any particular side chain may have for specific DNA recognition [Pabo and Nekludova 2000].

Because of its very convenient geometric and electrostatic environment, the major groove of B-form DNA provides twice as many of specific contacts with bases as the minor groove does [Seeman et al. 1976]. This may explain why most of DNA-binding proteins dock into the major groove of DNA (although a number of proteins exist that specifically contact the DNA in the minor groove only). The geometric shape of the major groove of B-form DNA defines the form of the protein secondary structure elements that dock into it. The most “popular” secondary structure element, that DNA-binding motifs employ for docking, is the α -helix. Although β -sheets or regions of extended polypeptide chain can also make contacts, α -helices are used much more frequently. A reason is the structure of α -helix, being complementary to the major groove of B-form DNA, provides the potentially highest density of possible contacts with the bases and backbones of DNA.

Very often multiple DNA-binding domains are required for site-specific recognition. The same motif may be used more than once when the DNA-binding protein forms a homodimer or heterodimer (e.g. steroid receptors, leucine zipper, HLH), or when a single polypeptide contains tandem recognition motifs (e.g. zinc fingers).

Recognition is a detailed structural process. There are many other factors that can play a critical role in recognition like hydration or sequence-dependent aspects of the DNA structure. Some of them will be briefly discussed in the following chapters.

2.1.4 Changes in structure and dynamics of protein and DNA induced by interaction

The formation of a sequence-specific protein-DNA complex typically involves changes in the structures of both binding partners (protein and DNA). These structural changes serve a wide variety of purposes, including specific DNA recognition, DNA packing, assembly of multiprotein complexes (e.g. transcription initiation), etc.

The average structure of DNA in solution is well described by the regular B-form double helix [Bram and Beeman 1971] proposed by Watson and Crick [Watson and Crick 1953]. However, there are perceptible deviations from this average structure in both free DNA and complexes with protein. In fact, the detailed topology of the DNA molecule itself may vary intramolecularly, depending on local base sequence and composition. The variations of structure present in free DNA not do only include local differences in stacking, tilt and twist of the bases relative to the helix axis, but often also result in noticeable changes of such global parameters of the intrinsic DNA curvature as the helical repeat and deflection of the double-helical axis [Travers 1995]. The extent of DNA bending estimated from the electrophoretic mobility in gels is highly sensitive both to temperature and to the ionic environment. In particular, the apparent intrinsic curvature is greatest at low temperatures and in the presence of a divalent cation [Diekmann 1987; Shliakhtenko et al. 1990].

Three decades ago von Hippel and co-workers [von Hippel and McGhee 1972; Berg and von Hippel 1988] proposed that these sequence-specific conformational differences of DNA [von Hippel and McGhee 1972; Berg and von Hippel 1988] “could provide a coarse identifying criterion” and therefore increase the specificity of protein-DNA interaction. This stimulated the question: to which extent do DNA-binding proteins sense and exploit the structural information coming from variations of structural parameters along with physical (e.g. stiffness) and topological properties (e.g. superhelicity)¹ of DNA sequences. Over the years of studies in this field a continuously increasing number of examples of “indirect protein-DNA recognition” emerged, supporting the original idea. The most prominent of these is the ability of the tumour suppressor protein p53 to discriminate between the different topologies of the DNA, including the relaxed B-form and the supercoiled non-B-form, and therefore modulate its own sequence-specific activity [Kim et al. 1997; Kim et al. 1999]. It has also been reported that this protein can specifically bind four-way and three-way junctions, facilitating their cleavage *in vivo* [Lee et al. 1997]. Apparently, this structure-specificity feature of p53 is a control component of promoter selectivity during transcriptional activation [Göhler et al. 2002]. Further examples of the structure-specific DNA recognition are high mobility group (HMG) domains of both sequence-specific and nonsequence-specific subfamilies that have been shown to recognize distorted DNA structures, such as four-way junctions [Kuhn et al. 1994], cisplatin lesions [Pil and Lippard 1992], and bulged DNA [Payet et al. 1999]. The biological role of these recognition

¹ Superhelicity – a measure of the topological state of covalently closed circular DNA double helices.

capabilities of HMG domains is not yet completely understood. In contrast, the ability of T7 RNA polymerase to recognize the mismatch bubble structures in double-stranded DNA [Aiyar et al. 1994] has a very clear purpose, that is, the initiation of transcription. There is a group of enzymes (resolvases and few endonucleases) involved in recombination, replication and reparation of DNA, whose primary function is structure-specific DNA recognition coupled with catalytic nuclease activity. Some of these enzymes are also able to recognize specific DNA sequences. In general, depending on the sequence-specificity of the DNA binding protein, structure-specific interactions may function as a means of a coarse or fine-tuning during complex formation and therefore result in different levels of functional activity control.

Protein-induced changes of the DNA structure are a much more frequent event in protein-DNA interaction than an “indirect readout” of pre-bent DNA structure. The most common changes are bending and twisting of the DNA helix. Crystallographic and biophysical studies of duplex DNA suggest that bending and twisting are sequence-dependent and that bending is highly anisotropic [Hagerman 1988; Hagerman 1990]. The sequence-dependence of these mechanical DNA features arises, as we will see later, from the distinctive stacking properties of the ten unique base steps² [Delcourt and Blake 1991]. On the basis of bending geometry data from known crystal structures of protein-DNA complexes, three major mechanisms of protein-induced DNA bending have been identified [Kerppola 2002]. First, proteins such as CAP and MATA1/a2 heterodimers contain a curved interaction interface that wraps the DNA partially or completely around the protein. These interaction interfaces are formed by different combinations of hydrogen bonding and electrostatic, hydrophobic, and van der Waal’s interactions. Second, proteins such as the TATA binding protein (TBP) and LEF-1 insert large hydrophobic amino acid side-chains between the base pairs in the minor groove. This causes partial unstacking of the base pairs and kinking of the DNA helix towards the major groove at the position of the inserted side-chain. Third, proteins such as Fos-Jun heterodimers and SRF bend DNA through electrostatic interactions. Positively charged amino acid residues bend DNA towards the residues, whereas negatively charged residues bend DNA away from the residues [Leonard et al. 1997]. Members of the same protein family can induce distinct DNA bends and even bend DNA in opposite directions [Kerppola and Curran 1991].

² Base step is a combination of two neighboring bases (nucleotides) in a DNA molecule.

Untwisting of DNA is a necessary concomitant to the fundamental processes of transcription initiation, replication initiation, and site-specific recombination, and is frequently directly associated with DNA bending. The primary role of the untwisting of DNA is to provide unhindered access to the minor groove and to facilitate the separation of the strands prior to DNA replication and transcription [Calladine and Drew 1992; Huang and Kowalski 1993]. A characteristic feature of the DNA sequences directing these processes is the presence of the dinucleotide TpA, either in isolation or, more frequently, as the tandem repeat TATA. The pyrimidine-purine step TpA is intrinsically less stable than all other dinucleotides (purine-pyrimidine, purine-purine or pyrimidine-pyrimidine), a property which is a direct consequence of the relatively low stacking overlap between the component base-pairs, and is therefore ideally suited for the nucleation of untwisting and bending [Drew et al. 1985; Kim et al. 1993]. Although transient untwisting at such sequences can occur in a naked DNA, particularly under negative superhelical strain, in biological systems this process must necessarily be mediated and regulated by DNA-binding proteins. A classical example of the protein-induced DNA unwinding is the complex formed by TBP and the TATA box [Chasman et al. 1993; Kim et al. 1993]. The binding mechanism of TBP to the TATA box was shortly introduced above. It is worth noting that seven base steps of the kinked DNA region facing towards the protein are undertwisted in this complex with a net deficit of approximately 110°.

It has been recognized for some time that not only DNA structure is altered in protein-DNA complexes. The structures of many DNA-binding proteins also undergo significant changes during sequence-specific interaction with DNA. These structural changes range from reorientation of side-chains to folding of the DNA binding domain or quaternary rearrangement of subunits ([Spolar and Record 1994] and references therein). Coupling of protein folding to specific DNA binding may be mediated by base contacts required for stabilizing the folded conformation. Alternatively, protein folding may be induced by interaction with a specific DNA structure. The DNA-binding domains of many eukaryotic transcription factors have disordered regions in the uncomplexed protein, which assume well defined folded structures upon binding to the specific DNA binding site [Frankel and Kim 1991].

The most common large-scale structural change induced by a sequence-specific DNA binding is the stabilization of an α -helical conformation. This type of structural transition has, for instance, been observed for leucine zipper dimerization domain of GCN4 [Talanian et al. 1990; Weiss et al. 1990], *Antp* homeodomain [Otting et al. 1990] and Fos Jun transcription

factor [Patel et al. 1990]. Folding of N-terminal arm of λ repressor [Clarke et al. 1991] and refolding of N-terminal subdomains of the *lac* repressor [Lewis et al. 1996] are other typical examples of structural changes of transcription factors. DNA binding can also result in the unfolding of protein domains and disruption of intramolecular interactions within a protein [Newman et al. 1995; Petersen et al. 1995]. These structural transitions can be coupled to structural changes in other parts of the proteins. Such intramolecular interactions can regulate DNA-binding activity and other aspects of protein function.

DNA binding also regulates the oligomerization state of many proteins. Most DNA-binding proteins bind palindromic recognition elements as dimers or higher-order oligomers [Deibert et al. 1999; Siksnys et al. 1999]. These dimers can be preformed in solution or the proteins may dimerize on DNA [Schwabe et al. 1993]. Coupling of DNA binding to dimerization can effect cooperative DNA binding and a nonlinear response to variations in protein concentration. In the case of proteins able to form heterodimers, the sequence of the binding site can influence the choice of dimerization partner. Furthermore, the recognition sequence can influence the orientation of binding by asymmetric heterodimers [Leonard and Kerppola 1998]. Thus, DNA binding frequently influences the quaternary structure of protein complexes even in those cases where their tertiary structures remain unaffected by DNA binding. DNA binding can also stabilize interactions between proteins by increasing their local concentrations. Changes in protein conformation upon DNA binding can influence the biological function of the protein. For example the DNA-binding site can function as an allosteric effector of the protein and alter its function, thereby preventing nonproductive functions of the free protein (for a review see: [Perutz 1990]. In addition, the conformation of the protein can vary at different recognition sequences, allowing the possibility of differential regulation of protein function at different binding sites [Diebold et al. 1998].

The conformations of proteins and DNA in solution are not restricted to rigid structures in their free and complexed states. Both protein and DNA undergo permanent fluctuations of their structures and are in dynamic equilibrium with their surroundings. Consequently, the experimentally observed structures of these molecules under given conditions are often time-averaged snapshots.

The extent of conformational dynamics of a DNA molecule depends on its axial and torsional flexibility, which are further defined by the range of conformations readily available to individual base steps. For any given step this range will be dependent on the available energy, as determined by both the immediate environment and the local superhelical strain [Sarai et al. 1989]. In general, however, the range will be more extensive for some base steps,

for example the pyrimidine-purine steps (e.g. TpA) which can adopt alternative conformations [Hunter 1993].

In the case of proteins, however, no such simplified description of factors defining the conformational dynamics can be given. This is due to the relatively high disorder of protein structures and the number of amino acids having different structural properties, which results in many degrees of freedom not only for the movement of separate residues, but also for global conformational transitions of the entire protein (for a review see: [McCammon and Harvey 1987]). When studying protein-DNA interactions, we seek to determine the changes in conformational dynamics of these molecules upon formation of the complex. These often involve the restriction of protein side-chain motions at the DNA interacting surface or constraint of torsional motions of DNA. Such effects make an energetically unfavorable contribution to the binding and are therefore reflected in the thermodynamic description of the interaction. As already mentioned above, formation of a protein-DNA complex is typically characterized by a large negative ΔC_p . Thermodynamic studies of several protein-DNA complexes have shown that this change of the heat capacity cannot be attributed solely to the hydrophobic effect [Sturtevant 1977; Ladbury et al. 1994; Spolar and Record 1994]. For the *trp*-repressor-operator interaction, some authors [Ladbury et al. 1994] suggest that the “unattributed” decrement of the heat capacity of the complex reflects a stereospecific restriction of the hydrated polar elements in the specific interface, i.e. “tightening of soft internal modes” in the vicinity of this interface.

NMR relaxation techniques have been used to study the changes in dynamics of a single arginine side chain in the Sso7d protein. This residue forms a part of the protein-DNA interface [Berglund et al. 1995]. A comparison of NMR order parameters measured for arginine side chains at a protein surface, at a protein-DNA interface and in a protein core indicate that the concerted effect of flexibility restriction of many side chains at an interface may be comparable to other determinants of binding thermodynamics. However, some of the inherent flexibility of the interacting molecules needs to be retained in the complex to increase the specificity and affinity. NMR studies of *lac* repressor headpiece-operator complexes have shown that specific high-affinity protein-operator complexes retain the inherent flexibility of the free operator, whereas the phosphate esters are conformationally restricted in the lower affinity complexes [Karslake et al. 1992].

2.1.5 Thermodynamics and kinetics of Protein-DNA interaction

The investigation of protein-DNA interactions is typically related to a number of questions:

- 1) How stable and how specific is the protein-DNA complex?
- 2) What are the structural details or mechanisms of binding/recognition?
- 3) What are the kinetic pathways of the binding reaction:
 - a. Is the binding diffusion- or reaction-controlled?
 - b. Is the binding just a simple bimolecular association event or does it contain some intermediate steps? If yes, what are these steps?
 - c. Are there any cooperativity effects (e.g. interaction between the binding proteins in specific and non-specific binding modes)?

In order to answer most of these questions, thermodynamic and mechanistic information is required. These data and, in some cases, molecular information can be deduced from the dependence of equilibrium and kinetic quantities (observed equilibrium “constants” and rate “constants”) on temperature and salt concentration. The thermodynamic parameters obtained from this type of analysis can be directly related to the degree of stability and specificity of the protein-DNA complex and provide an insight into the structural details of interaction.

On a thermodynamic level, questions regarding the origins of stability and the specificity of site-specific protein-DNA interactions are interrelated. The stabilities of site-specific (PS) and nonspecific (PD) complexes are defined by the standard free energy changes (ΔG°) for the respective processes of complex formation:



where P denotes a protein, S – a specific DNA site, and D – a nonspecific DNA site.

In thermodynamic terms, specificity is defined as the standard free energy change $\Delta G_{\text{PD} \rightarrow \text{PS}}^\circ$ for the process of transferring a protein from a nonspecific site to a specific site:

$$\Delta G_{\text{PD} \rightarrow \text{PS}}^\circ = \Delta G_{\text{PS}}^\circ - \Delta G_{\text{PD}}^\circ = (\bar{G}_{\text{PS}}^\circ - \bar{G}_{\text{PD}}^\circ) - (\bar{G}_{\text{S}}^\circ - \bar{G}_{\text{D}}^\circ) \quad (2.2)$$

\bar{G}_i° is here the dilute solution standard state chemical potential (partial molar Gibbs free energy) of molecular species $i = \text{P}, \text{S}, \text{PS}$. This quantity is not experimentally accessible.

If the protein-DNA complex is in equilibrium with its free components (protein and DNA), the free energy change of the binding process can be evaluated from the observed equilibrium association constant K_{obs} of the binding reaction:

$$\Delta G_{\text{obs}}^{\circ} = -RT \ln K_{\text{obs}} \quad (2.3)$$

Free energy change provides the thermodynamic measure of stability of the protein-DNA complex relative to the reactants under reversible binding conditions. The observed equilibrium association constant defines the extent of complex formation at the reaction equilibrium:

$$K_{\text{obs,PS}} \equiv \frac{[\text{PS}]}{[\text{P}][\text{S}]} \quad (2.4)$$

This definition of K_{obs} requires knowledge of the equilibrium concentrations of both the complex [PS] and the free reactants [P], [S]. In practice, these quantities are difficult to determine independently. Therefore, in order to express K_{obs} as a function of [PS], the assumption is required that no coupled or competitive equilibria involving protein, DNA and/or complex exist in significant amounts.

$$K_{\text{obs,PS}} \equiv \frac{[\text{PS}]}{(\text{P}_{\text{T}} - [\text{PS}])(\text{S}_{\text{T}} - [\text{PS}])} \quad (2.5)$$

The subscript “T” denotes the total concentration of the appropriate species. However, if such equilibria (e.g., nonspecific binding, protein aggregation) do occur and/or if significant concentrations of intermediate complexes are present at equilibrium, then $[\text{P}] \neq \text{P}_{\text{T}} - [\text{PS}]$ and/or $[\text{S}] \neq \text{S}_{\text{T}} - [\text{PS}]$, and the definitions of K_{obs} in Eqs. (2.4) and (2.5) are not equivalent. In these cases K_{obs} as defined by Eq. (2.5) may depend on the total concentration of protein and/or DNA [Berg et al. 1981].

Equilibrium binding of proteins to their DNA recognition sites can be determined by a variety of methods (nitrocellulose filter binding, gel retardation assay, footprinting, fluorescence spectroscopy techniques), all of which depend on either physical separation of complexes from free molecules or the measurement of a signal (e.g. changes in fluorescence intensity or anisotropy) that differentiates the free molecule(s) from the complex. Regardless of the method used, experimentally determined equilibrium binding constants reflect only the distribution between the free molecules and the complexes detected, and should thus be

considered “observed” constants, K_{obs} . For example, a given protein may form its final complex through a series of intermediates:



or may form several distinct complexes that may or may not be directly interconvertible. The existence of such situations cannot be discerned from equilibrium binding measurements alone, but may be crucial to the interpretation of the experimentally determined K_{obs} . In addition, this constant depends on the solution conditions. The use of macromolecular concentrations in the definition of K_{obs} that do not take into account differences in the association extent of ions, other small solutes, and solvent with the complex and with the reactants, as well as the neglect of activity coefficients describing nonideality arising from solute-solute interactions, results in a dependence of K_{obs} (and standard thermodynamic quantities derived from it) on solution variables (such as pH and ion concentrations) as well as on temperature and pressure.³ Since the stability and specificity of protein-DNA complexes are thermodynamically defined as free energy changes, which are highly dependent on solution conditions, they must be considered as relative rather than absolute quantities.

Thermodynamic studies of protein-DNA interactions provide a description of the forces that drive macromolecular complex formation and thus are complementary to structural studies. Understanding the energetics of DNA binding requires the measurement of the associated changes in enthalpy, entropy, and heat capacity. These components of the Gibbs free energy change of macromolecular association represent the fundamental forces, which drive the processes and can be obtained from calorimetric experiments by direct determination of $\Delta H_{\text{obs}}^{\circ}$ and $\Delta C_{p,\text{obs}}^{\circ}$ (which is defined as a slope of $\Delta H_{\text{obs}}^{\circ}$ change versus temperature at constant pressure) or from van't Hoff analysis of the temperature dependence of K_{obs} .

A general purpose of kinetic studies of protein-DNA interactions is to determine the mechanism, i.e., the sequence of elementary kinetic steps, employed by the protein for binding to a specific DNA site. However, kinetic experiments can provide much more information, including kinetic and thermodynamic characteristics of kinetically significant intermediates, the activation energy barriers between intermediates and cooperativity effects.

³ The effects of the concentration and type of salt on K_{obs} of protein-DNA interactions will be discussed in detail in chapter 2.1.6.

The association between proteins and DNA typically exhibits second order kinetics:

$$\frac{d[\text{PS}]}{dt} = k_{\text{assoc}} (\text{P}_T - [\text{PS}])(\text{S}_T - [\text{PS}]) \quad (2.6)$$

with an experimentally determined second order rate constant k_{assoc} ($M^{-1}s^{-1}$). A number of proteins (including *lac* repressor, RNA polymerase $E\sigma^{70}$, TATA binding protein [von Hippel and McGhee 1972], [Lohman 1986], [Petri et al. 1995]) have been observed to follow this rate law in the association process with DNA.

The association mechanism of a protein with a specific site in a DNA molecule surrounded by nonspecific sequences typically involves some or all of the following classes of steps:

1. Changes in the state of protein aggregation, conformational changes in the protein or DNA, and/or nonspecific binding of the protein to DNA, prior to the elementary bimolecular association step
2. An elementary bimolecular step in which an initial complex at the specific DNA site is formed at the diffusion-collision rate or at the rate that is slower than the estimated maximum diffusion-collision rate because of orientation effects
- 2'. An elementary bimolecular step in which an initial complex is formed at a distant nonspecific site at the (orientation-corrected) diffusion-collision rate, followed by a mechanistically distinct diffusion process in the domain of the DNA molecule to locate the specific site
3. Local and/or global conformational changes that occur in the initial complex at the specific site subsequent to the elementary bimolecular step, and result in formation of the functional specific complex

In general, information about the binding mechanism and its composite steps is obtained by using the following approach to the analysis of second order kinetic data: the rate constant k_{assoc} (Eq. (2.6)) is measured as a function of temperature and solution conditions (especially salt concentration, pH, solvent viscosity, as well as DNA length, if the process is expected to contain a facilitating step as in class 2' above). Depending on the magnitude of k_{assoc} and its dependence on temperature and salt concentration the results might indicate the presence of one of these situations:

- $k_{\text{assoc}} \approx k_{\text{dc}}$ (diffusion-collision) and weak dependence on temperature T and salt concentration $[\text{MX}]$
→ diffusion-limited reaction

- $k_{\text{assoc}} < k_{\text{dc}}$ and weak dependence on T and [MX]
→ diffusion-collision mechanism with severe orientation restrictions
- $k_{\text{assoc}} < k_{\text{dc}}$ and strong dependence on T and [MX]
→ intermediates before and/or after diffusion-collision step
- $k_{\text{assoc}} > k_{\text{dc}}$
→ facilitating mechanisms involving nonspecific DNA sites on the same DNA molecule.

The kinetics of dissociation of site-specific protein-DNA complexes follows a first order rate law:

$$-\frac{d[\text{PS}]}{dt} = k_{\text{dissoc}}[\text{PS}] \quad (2.7)$$

where k_{dissoc} (s^{-1}) is the observed (generally composite) first order rate constant. The mechanism must involve passage through the same steps (in the reverse direction) as in the association mechanism, according to the principle of microscopic reversibility. For example, if association is facilitated by the presence of contiguous nonspecific sites, these sites and facilitating mechanisms will also play a role in dissociation, and the equilibrium constant and thermodynamic quantities derived therefrom will be independent of this path-dependent effect.

Since cations may be considered as fundamental and omnipresent competitors with proteins for the vicinity of the DNA polyanion, all elementary protein-DNA dissociation rate constants are actually of pseudo first order, because cations are reactants in the elementary step of dissociation of the protein from DNA, and the cation concentration is in vast excess [Lohman et al. 1978; Lohman 1986]. Typically, measurements of association and dissociation kinetics are performed and compared at identical concentrations of electrolyte ions, so there is no need to eliminate this effect.

The rates of all association processes in solution are ultimately limited by the time it takes for the reactants to meet by diffusion. If there are no other chemical or conformational processes preceding the association that act to slow down the overall reaction [association] rate, such a reaction is called diffusion-controlled. Another extreme, a “reaction-controlled” association, describes the situation, in which the reaction steps necessary for a successful bimolecular association are much slower than the rate of intermolecular collisions. The diffusion-controlled upper limit of the association rate of any particular protein-DNA system

is important as a reference point for the analysis of the measured association rates and for the development of the appropriate association reaction model.

The maximum rate of an association reaction depends on the encounter probabilities of the components, which for a bimolecular association of two uniformly reactive spherical molecules (A and B) corresponds to the Smoluchowski limit [von Smoluchowski 1917], [Noyes 1961],

$$k_{dc}^o = \frac{4\pi N_A}{10^3} (D_A + D_B)(r_A + r_B) \quad (2.8)$$

where the factor 4π is the spherical solid angle (indicating that all directions of approach of the spheres lead to reaction), D_A and D_B are the translation diffusion constants (in cm^2/s) and r_A and r_B (in cm) are the hydrodynamic radii of molecules A and B, respectively. N_A is Avogadro's number, and the factor 10^3 normalizes the units of k_{dc} to $\text{M}^{-1} \text{s}^{-1}$. Equation (2.8) is valid for uncharged spheres and assumes no orientational constraints. The diffusion constants of such spherical molecules depend on their sizes, on temperature (T), and on solvent viscosity (η), as defined by the Stokes-Einstein relation.

$$D_A = k_B T / 6\pi\eta r_A; \quad D_B = k_B T / 6\pi\eta r_B \quad (2.9)$$

When A and B are approximately equal in size, one finds $k_{dc} = 8k_B T / 3\eta$, which corresponds to $k_{dc} \simeq 10^9\text{--}10^{10} \text{ M}^{-1}\cdot\text{s}^{-1}$ under normal solution conditions. This is sometimes taken as the maximum diffusion-limited association rate constant for molecular interactions.

Proteins and nucleic acids are of course not adequately modeled as uncharged, uniformly reactive spheres. DNA is a highly charged locally cylindrical polyanion. Proteins may or may not be spherical, and, more importantly, are polyampholytes⁴ with an overall charge that is not uniformly distributed and is function of pH. The entire molecular surface of either protein or DNA is not uniformly "reactive". Long-range Coulomb interactions may increase or decrease the probability of collision. Introduction of these effects leads to an improved estimate of the diffusion-collision rate constant for noncovalent interactions of the macromolecules (P and S) [Berg and von Hippel 1985; Lohman 1986]:

$$k_{dc} = \frac{4\pi\kappa f N_A}{10^3} (D_P + D_S) R_{PS} \quad (2.10)$$

⁴ A polyampholyte is defined as any molecule that contains many positively and negatively charged groups.

where κ is the probability that the collision has the correct mutual orientation to lead to interaction, $R_{PS} = r_A + r_B$ is an effective reaction radius and f is a dimensionless factor that accounts for nonspherical geometry and long-range Coulomb interactions ($f < 1$ for repulsion; $f > 1$ for attraction). Von Hippel and Berg [von Hippel and Berg 1989] applied this equation to estimate the diffusion-limited rate constant of bimolecular association between *lac* repressor and specific DNA operator site on a large piece of DNA. Based on several assumptions (that approximately one fifth of the repressor surface represents active site and that approximately one-fourth of the cylindrical surface of the operator DNA sequence actually interacts with repressor, which gives $\kappa = 0.05$; neglecting geometrical and electrostatic corrections, which are expected to be relatively small) they obtained the value of k_{dc} of the order of $10^8 \text{ M}^{-1}\text{sec}^{-1}$. However, this value is as much as 500 times smaller than the association rates of *lac* repressor with λ phage DNA measured by Riggs *et al.* [Riggs *et al.* 1970] and confirmed by others [Barkley 1981; Winter *et al.* 1981]. This discrepancy as well as an unusually strong dependence of k_a on ion concentration was explained by Richter and Eigen [Richter and Eigen 1974] who developed a theoretical model based on the idea that the reduction of dimensionality in the diffusion may enhance reaction rates [Adam and Delbrück 1968]. In the actual physical situations, this principle represents a nonspecific binding of the protein (repressor) to any place on the DNA macromolecule and subsequent one-dimensional sliding along the DNA chain until a specific site (operator) is found. This sliding model was extended later by Berg *et al.* [Berg *et al.* 1981], taking into account other effects like intersegment transfer, inter- and intradomain dissociation and reassociation. Thus, in general, the values of observed association rate constants exceeding the estimation of these constants for diffusion controlled interactions may point to the presence of facilitated docking processes or to enhancements by shaped electrostatic fields.

Diffusion limited association reactions exhibit weak dependence of association rate constant on temperature and salt concentration. Through substitution of the diffusion constants in (2.8) by its expressions from the Stokes-Einstein relation one immediately obtains $k_{dc}^0 \propto T/\eta$. In water, an Arrhenius plot of this temperature dependence predicts a very small but positive activation energy of $\approx 4 \text{ kcal/mol}$ for a diffusion-controlled process. This effect can be employed to examine whether the association process is diffusion-controlled or contains some preceding or following additional steps that usually make the activation energy barrier significantly higher. The weak dependence of k_a on monovalent salt concentration [MX] was explained by Lohman *et al.* [Lohman *et al.* 1978] as a consequence of screening of

the interaction between protein and DNA by electrolyte. For double-helical DNA, screening of phosphates is equivalent to the association of 0.12 univalent cation per phosphate, which result in:

$$d \ln k_{dc} / d \ln [MX] = -0.12z \quad (2.11)$$

where z is the number of ionic interactions in the equilibrium protein-DNA complex.

2.1.6 Influence of environmental factors

Water

The chemical properties of water molecules and their thermodynamic behaviour play a critical role in the formation of protein-DNA complexes in aqueous solutions. As mentioned in chapter 2.1.1, formation of the complexes between protein and DNA involves the displacement of surface water molecules located in the interface region of the macromolecules. The role water molecules play in protein-DNA binding has essentially two aspects. First, water molecules caged in the interface between protein and DNA in the formed complex are often involved in the formation of water mediated specific contacts. Second, rearrangement of water molecules effects the thermodynamical costs of entropy and enthalpy related to the formation and stability of the complex. Protein-DNA complexes are quite diverse when it comes to their use of water. Striking contrasts can be observed even within the same HTH family of transcription factors. In the structure of the 434 repressor-operator complex [Anderson et al. 1987; Aggarwal et al. 1988] the protein is in direct contact with DNA, which seems to be important for sequence-specific recognition. Although few water-mediated contacts can be found here, they seem to be of secondary importance relative to the direct contacts which are critical for specific recognition. A completely different situation with regard to the protein-DNA interface is observed in the structure of the trp repressor and DNA complex [Otwinowski et al. 1988]. There are very few side chain-base contacts, and they do not seem to be important for sequence-specific recognition [Bass et al. 1988]. Furthermore, in the protein-DNA interface there are three ordered water molecules that hydrogen-bond with protein side chains and bases important for the specificity. Experimental studies have demonstrated that water molecules in trp repressor complex do indeed play an important role in DNA-recognition mediating critical contacts between bases and protein side chains which do not otherwise interact [Bass et al. 1988; Joachimiak et al. 1994]. Similarities

in the position of bound water in the complex and free operator further suggest that such water molecules provide a non-covalent extension of the DNA surface utilized in recognition [Shakked et al. 1994].

There are many more structures of protein DNA complexes that show features of water conceptually similar to those observed in the trp repressor-operator complex (e.g. [Newman et al. 1995]). Studies of the structures of several homeodomain-DNA complexes revealed another interesting implementation of water in protein-DNA interactions. NMR and molecular dynamic studies of the *Antennapedia* homeodomain-DNA complex showed the presence of multiple water molecules in the protein-DNA interface having nanosecond residence times [Billeter et al. 1993; Billeter et al. 1996]. Furthermore, two critical residues (Glu50 and Asn51) in the recognition helix of this homeodomain structure appear to be involved in a fluctuating network of water-mediated hydrogen bonds and are in a dynamic equilibrium between two or more contact sites on the DNA [Wilson et al. 1996]. A similar behaviour of Glu50 has also been shown in the structure of the even skipped (*eve*) homeodomain-DNA complex [Hirsch and Aggarwal 1995]. Such a highly hydrated and relatively disordered interface represents a new role for water – a “molecular lubricator” reducing the entropic costs of protein docking and one-dimensional diffusion on the DNA until the specific sequence is found, and at the same time providing the dense network of interactions required for the recognition.

The large number of high-resolution crystal structures of protein-DNA complexes that emerged in recent years made it possible to perform an extensive structural analysis of water function in protein-DNA recognition. It has been shown that only a relatively small part (~6%) of the crystallographic water molecules are involved in the recognition process, whereas the primary function of the other water molecules observed in crystal structures is thought to buffer the electrostatic repulsions between phosphate groups of DNA and the electronegative atoms on the protein [Reddy et al. 2001].

Formation of protein-DNA complexes is also coupled with a large negative change in heat capacity (ΔC_p) [Sturtevant 1977]. The primary determinant of this effect is attributed to the release of water molecules from nonpolar surfaces of protein and DNA [Ha et al. 1989], although the restriction of vibrational degrees of freedom inside the complex (partially due to local folding events) seems to also contribute strongly [Spolar and Record 1994].

ions

Macromolecules in aqueous solutions are associated with a large number of solvent molecules including ions of different kind, charge and valence. The majority of protein-nucleic acid interactions are affected to a very large extent by all of these (three) parameters [Record et al. 1991; Lohman and Mascotti 1992]. Due to the high density of negatively charged phosphate groups, DNA is associated with cations (M^+) that partially neutralize the net charge of the nucleic acid [Manning 1978]. In this context DNA can be considered as a linear polyelectrolyte, having a constant fraction ψ of one counterion thermodynamically bound per nucleic acid phosphate [Record et al. 1976]. This fraction is dependent only on the structural charge density along the nucleic acid and the counterion valence, and is independent of the bulk salt concentration as long it exceeds the concentration of phosphate charges [Manning 1969; Manning 1978; Record et al. 1978]. It has been shown that for helical B-DNA the fraction of neutralized structural charges is approximately 88% ($\psi = 0.88$) [Record et al. 1976; Anderson and Record 1982]. Binding of a protein to DNA in monovalent salt solution is a cation-exchange process in which the DNA structural charge density is reduced and cations (M^+) are released from DNA to the bulk solution. Defining the protein with positively charged groups as an oligocation with net charge $+z$, its association with a linear nucleic acid in the presence of a monovalent salt (MX) can be expressed by the following cation-exchange reaction:



where $DNAP^-$ is a phosphate charge of the of DNA. The predicted dependence of the observed equilibrium constant for this binding reaction, K_{obs} , on the concentration of monovalent cations ($[M^+]$) at constant temperature (T) and pressure (P) is [Record et al. 1976]:

$$\left(\frac{\partial \log K_{obs}}{\partial \log [M^+]} \right)_{T,P} = -z\psi \quad (2.13)$$

Because the coefficient, $z\psi$, is constant for a given ligand and nucleic acid, $\log K_{obs}$ is expected to decrease as a linear function of the increase in $\log [M^+]$. From the slope of this line (further referred to as SK_{obs}), the net number of cations released upon complex formation ($z\psi$) and the valence (z) of the oligocation may be determined.

Although the counterion condensation at the DNA reduces the absolute enthalpy change (ΔH) of protein–nucleic acid interactions, the release of these counterions into the bulk solution upon protein-DNA complex formation provides a favourable entropic contribution ($\Delta S > 0$) to the free energy of binding [Record et al. 1976]. This free energy term has been referred to as the polyelectrolyte effect and is given for a ligand-nucleic acid equilibrium occurring in a monovalent salt solution by the following equation [Record et al. 1976]:

$$\Delta G_{\text{PE}}^0 = z\psi RT \ln[M^+] \quad (2.14)$$

This equation indicates that the contribution of the polyelectrolyte effect to the free energy of binding is eliminated at 1 mol/L of M^+ , which is the so-called pseudo-standard state for this reaction. This behaviour has been verified for a number of simple oligocations (e.g. oligolysines and polyamines) binding to both duplex [Record et al. 1976; Lohman et al. 1980] and single stranded nucleic acids [Mascotti and Lohman 1990]. It is thought that in these cases as well as in the case of nonspecific DNA binding of some proteins [deHaseth et al. 1977] the polyelectrolyte effect provides the major driving force of oligocation-linear nucleic acid complex formation. [Lohman and Mascotti 1992].

There are substantial differences in the influence of the polyelectrolyte effect to the specific and nonspecific protein-DNA interactions. Since nonspecific binding of proteins to DNA is expected to primarily involve electrostatic interactions between the phosphates of DNA and any charged residues on the DNA-binding surface of the protein, the model of the protein as a cationic ligand (2.12) can be applied. As already mentioned above, formation of a protein-DNA complex, is stabilized by the polyelectrolyte effect and the observed equilibrium constant decreases with an increase in salt concentration (Equation (2.13)). A very good example of this behaviour is the nonspecific DNA binding of the *E. coli* lac repressor, which in the presence of a monovalent cation shows linear dependence of $\log K_{\text{obs}}$ on $\log[M^+]$ with $\partial \log K_{\text{obs}} / \partial \log[M^+] = -11 \pm 2$ and the intercept of a linear extrapolation of $\log K_{\text{obs}}$ to a 1 mol/L of M^+ representing the non-electrostatic component of the binding free energy, approaches zero [Lohman et al. 1980].

In the case of sequence-specific DNA binding, there is a considerable contribution of the free energy change arising from specific contacts in addition to stabilization caused by the polyelectrolyte effect. Therefore, the equilibrium constant for the site-specific DNA binding (K_{sp}) is significantly higher than observed at identical conditions for nonspecific DNA binding (K_{ns}). The experimental results of the most well-studied protein, *E. coli* lac repressor,

related to the sequence-specific binding to its operator show that the observed K_{sp} value varies less dramatically than the K_{ns} for unspecific binding (respectively $\partial \log K_{sp} / \partial \log [M^+] = -7 \pm 2$ versus $\partial \log K_{ns} / \partial \log [M^+] = -11 \pm 2$). This difference, as well as the fact that the value of $\log K_{sp}$ for the specific binding mode of lac repressor at 1 mol/L M^+ (where the polyelectrolyte effect is blocked out) is large and positive (6.1 ± 0.1 in KCl) reflects the favourable free energy contribution arising from specific contacts and conformational changes induced in the specific complex.

As with the phosphate groups of the DNA, the charged residues of the protein can also recruit ions from solution. Thus, protein-DNA interaction is not only affected by the cations bound to DNA but also by anions associated with the positively charged amino acids in the DNA binding region of proteins. Studies of anion effects on K_{obs} and of the interaction of lac repressor with the lac operon on λ plac5 DNA [Barkley et al. 1981] and with isolated synthetic (strong-binding) lac operator [Ha 1990] have shown that there are remarkable differences in K_{obs} when different types of anions (Acetate⁻, Glutamate⁻, F⁻, Cl⁻ etc.) are used as the Na⁺ salt in binding reactions. For example, there was an approximately 30-fold increase in K_{obs} (without a significant change in $\partial \log K_{obs} / \partial \log [MX]$) of both specific and nonspecific binding of lac repressor to the wild-type operon upon replacement of Cl⁻ by Ac⁻ at constant salt concentration [Barkley et al. 1981]. Although these anion-specific effects are quite substantial, there is as yet no unambiguous explanation accounting for their mechanisms.

2.2 Aspects of homeodomain-DNA interaction

Unusual changes in the shape and pattern of animals, e.g. the exchange of one body part for another (e.g. an antenna into a leg for insects), attracted in the beginning of the last century the attention of biologists studying the classical genetics of fruit fly *Drosophila*. Later with the advent of the gene cloning techniques, these mutations, called homeotic, could be located to their genes, which were named after the mutation. The homeotic genes are involved in the genetic control of higher organisms, in particular in the specification of the body plan, pattern formation, the determination of the cell fate, and several other basic developmental processes (for a review see [Gehring 1987; Scott et al. 1989]). These genes share a common sequence of 180 bp, the homeobox, which was first discovered in *Drosophila* [McGinnis et al. 1984]

and later in vertebrates [Carrasco et al. 1984]. Subsequently, hundreds of homeoboxes were discovered, spread over a large variety of species ranging from yeast to humans (for a guide to homeobox genes see [Duboule 1994]).

The homeobox encodes a self-folding, stable protein domain of about 60 amino acids, called the homeodomain, which is composed of three α -helical regions folded into a tight globular structure: helix I is preceded by a flexible N-terminal arm and separated by a loop from helix II, which forms with helix III a HTH motif (see Figure 2.1 B).

Even though a large variety of homeodomain proteins have arisen in the course of evolution, their amino acid sequences have been conserved to a high degree [Gehring et al. 1994]. There are seven positions in the homeodomain sequence that are occupied by the same amino acid in more than 95% of the 346 known homeodomains (Figure 2.2). These include

	N-terminus		helix 1		helix 2		helix 3		C-terminus
	1	10	20	30	40	50	60		
Bcd:	PRRTRTTFTS	SQIAELEQHF	LQGRYLTAPR	LADLSAKLAL	GTAQVKIWFK	NRRRRHKIQS			
Ubx:	RRRGRQTYTR	YQMLELEKEF	HTNHYLTRRR	RIEMAHALCL	TERQIKIWFQ	NRRMKLKKEI			
En:	EKRPRTAFASS	EQLARLKREF	NENRYLTERR	RQQLSSELGL	NEAQIKIWFQ	NKRAKIKKST			

Figure 2.2 Sequence comparison of homeodomains. Amino acids are represented in a single-letter code (see the abbreviation list). The highly conserved amino acids are highlighted in gray. For details on function see appendix A1.

Leu16, Phe20, Trp48 and Phe49, which are hydrophobic core amino acids preserving the tertiary structure of the homeodomain, and Arg5, Asn51 and Arg53, which are involved directly in DNA binding. Another characteristic feature of the homeodomains is a high content of basic amino acids (in particular in the helix III), which implies that the long-range electrostatic interactions play an important role in homeodomain-DNA binding.

About a dozen of structures of homeodomain-DNA complexes available today provide a detailed structural image of interaction between homeodomain and DNA (for a review see [Wolberger 1996]). The homeodomains typically contact the DNA with residues at the N-terminus, around the beginning of helix II and within helix III. The recognition helix (III), which is considerably longer than that of the HTH motif in prokaryotic repressors, spans the entire major groove of DNA molecule (see Figure 2.3). The residues from three helical turns, namely at positions 47, 50, 51 and 54 form specific contacts with DNA bases. It was shown [Billeter 1996] that despite some variations in amino acid composition, the available structures of homeodomain-DNA interfaces (the amino acids 47-54 and the appropriate

fragment of the DNA molecule) could be superimposed with an average of root mean square difference of 0.81 Å, which clearly indicates that amino acid side chains at equivalent positions in the recognition helix are positioned next to equivalent DNA bases in all complexes. Additional specific contacts are established in the minor groove of the DNA by two basic amino acids (typically arginines) at position 3 and 5 of flexible N-terminal arm. The very high conservation of the Arg5 indicates that these contacts may play a critical role in specific homeodomain-DNA binding. A 130-fold decrease in the DNA-binding affinity reported for the fushi tarazu (*ftz*) homeodomain mutant lacking 6 N-terminal [Percival-Smith et al. 1990] witnesses the importance of the N-terminal arm-DNA contacts.

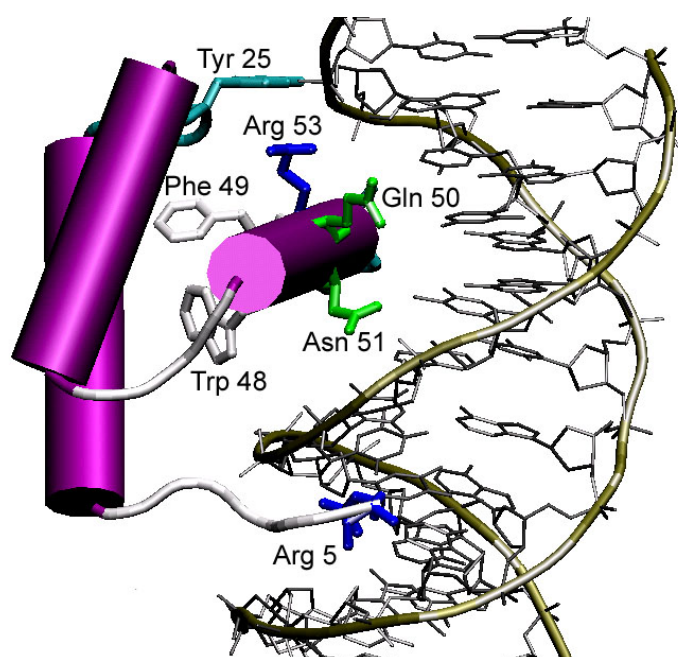


Figure 2.3. Structure of Ubx homeodomain-DNA complex [Passner 1999]

The homeodomain-DNA interface is remarkable for long-range organization of an interfacial water cavity [Wilson et al. 1995]. This cavity, a channel in the DNA major groove adjacent to the recognition helix, is proposed on the basis of molecular dynamics simulations to be essential for homeodomain-DNA recognition [Billeter et al. 1996]. Originally predicted by nuclear magnetic resonance studies [Billeter et al. 1993], interfacial water molecules are well visualized by X-ray diffraction and participate in a network of protein-water-DNA hydrogen bonds [Hirsch and Aggarwal 1995; Li et al. 1995; Wilson et al. 1995]. Reorganization of this network is suggested [Billeter 1996] to account for differences in the specificity among functional subclasses of homeodomains.

Most of the DNA sequences that interact favorably with homeodomains contain a tetranucleotide ATTA (TAAT on the complementary strand). Structural data of the available homeodomain-DNA complexes suggest that the conservation of this motif is closely related with the extreme conservation of the four amino acids Arg3, Arg5, Ile47 and Asn51 interacting with this DNA fragment as well with the previously mentioned conserved docking arrangement of divergent homeodomains against DNA. The bases immediately next to the core motif are contacted by considerably less conserved amino acids of the recognition helix (e.g. residues at position 50 and 54) and therefore are homeodomain-specific. For example the homeodomains containing a Lys residue at position 50 (e.g. Bicoid and Orthodenticle that belong to so-called K-50 family of homeodomains) exhibit a strong preference of GG dinucleotide succeeding the ATTA core [Driever et al. 1989]. Corresponding nucleotides in the Antp and Ubx homeodomain-DNA complexes (CC) [Billeter et al. 1993; Passner et al. 1999] are contacted by the Gln side chain at position 50 and the Met54 (Figure 2.3). Furthermore, it was shown that the specificity of the ftz homeodomain can be switched to that of Bcd by substituting the Gln50 into Lys in the recognition helix [Schier and Gehring 1993].

Another interesting feature of the homeodomains is their ability to bind to so-called low-affinity binding sites. These sites differ from the specific recognition sequences having high affinity and often do not contain the ATTA core motif. It was proposed that the binding to the low-affinity sites might play an important role in the function of several homeobox proteins *in vivo* [Driever et al. 1989; Pick et al. 1990].

The high similarity between the recognition sites and at the same time relatively low selectivity for the binding sequences of a large number of homeodomains is in sharp contrast to the distinctly different biological effects exerted by these proteins *in vivo*. This apparent inconsistency stimulates a lot of questions regarding the mechanisms of homeodomain DNA interaction, most of which are still unanswered.

3. Aims of the present study

Since the first homeodomain protein has been discovered [McGinnis et al. 1984] various aspects of homeodomain-DNA interactions have been investigated. A number of structures of homeodomain-DNA complexes were solved by NMR or X-ray crystallography over a last decade providing a detailed insight into the structural mechanisms of homeodomain-DNA recognition. Some thermodynamic parameters of homeodomain-DNA interaction have been determined using microcalorimetry and theoretical methods [Carra and Privalov 1997; Fogolari et al. 1997]. However, very little has been known about the kinetics of homeodomain-DNA binding, which could reveal the answer to one of the major unanswered questions about the homeodomains: how do these proteins, having relatively promiscuous DNA binding properties, achieve high spatial and temporal specificity required for their function? The present study does not presume to offer a direct answer to this question. Though the main aim was to elucidate the kinetic mechanisms reflecting the interaction of Bcd and Ubx homeodomains with DNA, mostly by means of the stopped-flow kinetic technique. A further aim was to assess the structural alterations occurring in DNA and/or protein upon specific complex formation using various spectroscopic methods.

4. Materials and methods

4.1 Materials

Standard chemical reagents were of analytical purity grade and were obtained from the companies: Baker (Deventer, Netherlands), MERCK (Darmstadt, Germany) or Sigma-Aldrich (Deisenhofen, Germany). The fluorescent compounds (Alexa 488, Alexa 546, IAEDANS and monobromobimane) used for the protein labeling were from Molecular Probes (Eugene, OR, USA). Media for cell cultures was obtained from Difco (Detroit, MI, USA) and Gibco (Paisley, Scotland). Materials for chromatography were obtained from Amersham Pharmacia Biotech (Uppsala, Sweden), Bio-Rad (Richmond, CA, USA) and Quiagen (USA). *E. coli* strain BL21(DE3)(pLysE) used for protein expression was obtained from Novagen (USA). Materials used for the polyacrylamide gel electrophoresis were from Roth (Karlsruhe, Germany) and Invitrogen (Carlsbad, CA, USA).

The composition of buffers used in the experiments is provided in the description of the experimental conditions in appropriate sections of methods.

4.2 Protein preparation

Protein expression and purification

Plasmids pRSbcd89 and pRSN-HisbcdHD (provided by R. Rivera-Pomar) were used to generate Bcd homeodomain. pRSbcd89 [Rivera-Pomar et al. 1995] encodes a protein spanning amino acids 89 to 154 of Bcd protein fused to 6 histidines at the C-terminus. The resulting homeodomain protein will be referred further as Bcd_{CHis}. pRSN-HisbcdHD [Niessing et al. 2000] encodes the same sequence of Bcd homeodomain as pRSbcd89, but contains 6 histidines at the N-terminus of the protein and will be further referred to as Bcd_{NHis}. Bcd mutants K50A and R54A were derived from pRSN-HisbcdHD [Niessing et al. 2000]. pRSbcd89c plasmid (R. Rivera-Pomar, unpublished) was derived from pRSbcd89 by adding a cystein residue at the C-terminus of the histidine tag. The encoded protein will be referred further to as Bcd_{CHis}-Cys. The proteins were expressed in *E. coli* BL21 (DE3) and purified as described [Rivera-Pomar et al. 1996]. Purity (>94%) was determined by SDS polyacrylamide gel electrophoresis (PAGE). Polyacrylamide gels were stained with SYPRO

Orange (Molecular Probes, Eugene, OR, USA). A frequent 24 kDa⁵ contaminant protein was removed by chromatofocusing using Mono P columns (Pharmacia, Uppsala, Sweden) equilibrated with 25 mM triethanolamine-HCl (pH 7.0). The Ubx homeodomain was kindly provided by P.A. Beachy (Howard Hughes Medical Institute Research Laboratories, The Johns Hopkins University School of Medicine). The Engrailed (En) homeodomain was kindly provided by W. Hoyer (MPI for Biophysical Chemistry). Concentrations of the proteins were determined by measuring absorption at 280 nm and using extinction coefficients calculated as described in [Gill and von Hippel 1989].

Fluorescent labeling of Bcd_{CHis}-Cys protein

Before the labeling reaction the disulphide bridges between the cysteines of the Bcd_{CHis}-Cys protein molecules were reduced by adding a ~100-fold excess of the reducing agent dithiothreitol (DTT) to the protein solution and incubating at 21 °C for 4.5 hours. After the reaction was completed DTT was removed by dialysis of the sample at 5 °C against a labeling reaction buffer 1 (25 mM Na-phosphate, 500 mM NaCl, pH 7.0), used for labeling with Alexa 488, Alexa 546 and IAEDANS or labeling reaction buffer 2 (15 mM Tris-HCl, 500 mM NaCl, pH 7.0), used for labeling with mBBr. The dialysis was performed using a dialysis tube (Roth, Karlsruhe, Germany) with a molecular mass cut-off of 2000 Da for 9 hours by replacing the dialysis buffer every 3 hours. The labeling reaction was initiated by adding a 20-fold excess of the fluorescent label (Alexa 488, Alexa 546, IAEDANS, prepared in DMSO, or mBBr, prepared in acetonitrile). The samples were incubated in the dark at 21 °C for 8 hours by shaking periodically. The labeled protein was separated from the excess of the free fluorescent label by gel filtration chromatography (GFC) using a column (bed volume: 1 mL) of Bio-Gel P4 (Bio-Rad, Hercules, CA, USA) and equilibrated with a 25 mM Na-phosphate buffer pH 7.0, 500 mM NaCl. After purification, the labeled protein was concentrated on a Microcon-3 (Amicon, Beverly, MA, USA) centrifugation filter (molecular mass cut-off: 3000 Da).

⁵ Da (Dalton) – a non-SI-unit of mass that equals the weight of a hydrogen atom (1.657×10^{-24} g).

4.3 DNA preparation

All oligonucleotides used in this study (Table 4.1) were obtained from Metabion GmbH (Martinsried, Germany). The fluorescent labeling of the oligonucleotides b1_5F, b1_3F, b1_5C, b1_3C, b1_FC, u1_FC and ns1_FC was performed by the above mentioned company during the oligonucleotide synthesis by covalent coupling of the fluorescein- or Cy3-phosphoramidite at the appropriate position of the DNA molecule as indicated in the table 4.1.

Table 4.1 Structures of oligonucleotides.

Oligonucleotide	Structure
b1	AATCTAATCCCTATA
b2	TATAGGGATTAGATT
b1_5F	5' -Fluorescein-AATCTAATCCCTATA-3'
b1_5C	5' -Cy3-AATCTAATCCCTATA-3'
b1_3F	5' -AATCTAATCCCTATA-Fluorescein-3'
b1_3C	5' -AATCTAATCCCTATA-Cy3-3'
b1_FC	5' -Fluorescein-AATCTAATCCCTATA-Cy3-3'
u1	AATTTAATGGCTATA
u2	TATAGCCATTAAATT
u1_FC	5' -Fluorescein-AATTTAATGGCTATA-Cy3-3'
ns1	AATCCACAGCCTATA
ns2	TATAGGCTGTGGATT
ns1_FC	5' -Fluorescein-AATCCACAGCCTATA-Cy3-3'

The concentration of the oligonucleotides was determined from their absorption spectra at 260 nm using absorption extinction coefficients calculated for single-stranded sequences at 25 °C according to [Puglisi and Tinoco 1989]. In the case of the fluorescently labeled oligonucleotides the contributions of the fluorescent labels (fluorescein and Cy3) to the absorption at 260 nm were calculated from the ratios $\epsilon_{260}^{Fl} / \epsilon_{490}^{Fl} = 0.3$ and $\epsilon_{260}^{Cy3} / \epsilon_{550}^{Cy3} = 0.15$ (determined from the absorption spectra of fluorescein and Cy3 labels in solution without DNA) and subtracted from the net absorption at 260 nm of the labeled oligonucleotides.

The double-stranded oligonucleotides (b, u, b5F, b3F, b5C, b3C, bFC, uFC, ns, nsFC) were prepared in 10 mM Na-cacodylate, 100 mM NaCl, pH 7.0 by mixing equimolar amounts of the appropriate complementary single-stranded oligonucleotides (b1 and b2, u1 and u2, b1_5F and b2, b1_3F and b2, b1_5C and b2, b1_3C and b2, b1_FC and b2, u1_FC

and u2, ns1 and ns2, ns1_FC and ns2) and heating at 70 °C for 5 min. Then the samples were slowly (in a course of ~12 hours) cooled to room temperature in order to promote a complete hybridization of the separate complementary strands. The resultant double-stranded oligonucleotides were separated from any residual non-hybridized single-stranded precursors by running an electrophoresis in 18% polyacrylamide gel at native conditions in 89 mM Tris-borate pH 8.4, 2 mM MgCl₂ at room temperature. The double-stranded oligonucleotides were recovered from the polyacrylamide gel by electroelution with a Biotrap BT 1000 system (Schleicher & Schuell, Dassel, Germany) in the same buffer as used for the gel electrophoresis, and concentrated in a Na-cacodylate buffer by using a Microcon-3 (Amicon, Beverly, MA, USA) centrifugation filter (molecular mass cut-off: 3000 Da).

4.4 Steady state absorption spectroscopy

Absorption spectra were recorded with a Cary 100 (Varian, Mulgrave, Australia) spectrophotometer at 25 °C with 1 nm resolution and 0.1 s integration time. Spectra were baseline-corrected using the absorption of the solvent as a reference. Absorption spectra of proteins and of unlabeled oligonucleotides were collected in the range of 220 – 350 nm. Absorption spectra of fluorescently labeled oligonucleotides were measured in the range of 220 to 600 nm. Samples of all oligonucleotides were measured in 10 mM Na-cacodylate, 100 mM NaCl, pH 7.0. Absorption of the proteins was measured in 20 mM Na-phosphate, 500 mM NaCl, pH 7.0. Time resolved measurements were performed at 5 °C at fixed wavelengths (220, 260, 280 and 350 nm) with 1 s integration time. All spectroscopic measurements were performed in quartz cuvettes (Hellma, Müllheim, Germany).

4.5 Circular dichroism

Circular dichroism (CD) spectra were acquired with a JASCO J-720 CD spectropolarimeter (Jasco Corporation, Tokyo, Japan) equipped with the PTC348 peltier temperature control accessory. Spectra were acquired in 25 mM Na-Phosphate buffer, pH 7.0, 100 mM NaCl at 5 °C. Bcd-specific (b), Ubx-specific (u) and nonspecific (nsb) oligonucleotides were used for the measurements. Homeodomain-DNA complexes were prepared by mixing equal amounts

(in moles) of protein and oligonucleotide. The final concentration of both the protein and the oligonucleotide in the sample solution was 5 μM .

4.6 Steady state fluorescence spectroscopy

Fluorescence spectroscopy measurements were performed at 5 $^{\circ}\text{C}$ with 1 nm resolution on a PTI Quanta Master 6 SE photon counting spectrofluorometer (Photon Technology International, Lawrenceville, NJ, USA), equipped with a double excitation and single emission monochromator. The emission spectra were corrected for the instrument response, lamp fluctuations and solvent background contributions. Polarization artifacts were avoided by using “magic angle” conditions (i.e. exciting sample with vertically polarized light and detecting emission at the angle of 54.7° relative to the orientation of the excitation polarization). The intrinsic fluorescence of homeodomains was measured in the range of 290 to 450 nm by exciting the aromatic amino acids at 280 nm. Tryptophan fluorescence of homeodomains was recorded over 310-450 nm range by exciting at 295 nm. Emission spectra of the single and double-labeled oligonucleotides were collected in the range of 505 to 650 nm with excitation at 493 nm. Excitation spectra of the double-labeled oligonucleotides were recorded in the 400 – 590 nm range with emission at 600 nm. All measurements were performed using constant slit apertures (4 nm of excitation monochromators and 8 nm of emission monochromator) and gain settings. Fluorescence anisotropies (r) were determined by measuring fluorescence intensities with vertically polarized excitation and either vertically (I_{VV}) or horizontally (I_{VH}) oriented emission polarizers

$$r = (I_{VV} - GI_{VH}) / (I_{VV} + 2GI_{VH}) \quad (4.1)$$

where the G factor corrects for gain inequality in instrumental response to polarized light. The G factor was determined using horizontally polarized excitation and either vertically (I_{HV}) or horizontally (I_{HH}) oriented emission polarizers

$$G = I_{HV} / I_{HH} \quad (4.2)$$

Samples were prepared in 25 mM Na-HEPES, pH 7.0, 100 mM NaCl, 0.001% v/v surfactant P20 (polyoxyethylenesorbitan, a non-ionic detergent).

4.7 Fluorescence titrations

Background

Titration experiments imply measurements of protein-DNA binding extent (also called “binding density”) at fixed concentration of one of the binding partners e.g. DNA, when the concentration of the other partner e.g. homeodomain, is varied. Binding density is defined as the ratio of bound protein concentration (P_C) to the total concentration of DNA (D_T)

$$\nu = \frac{P_C}{D_T} \quad (4.3)$$

The concentrations of protein-DNA complexes are usually determined under equilibrium conditions by spectroscopic approaches, e.g. by monitoring the changes in steady state fluorescence intensity induced upon formation of the complex. In order to use this signal change for the calculation of the titration curve, it is commonly assumed that a linear relationship exists between the signal change and the fractional saturation of the fluorescently labeled DNA molecule by protein. In other words, if it can be assumed that initial fluorescence intensity F_o of the sample in absence of the protein is linearly proportional to the total concentration of the DNA $F_o = f_F D_T$ and the observed total fluorescence intensity F_{obs} in presence of the protein can be expressed as $F_{obs} = f_F D_F + f_C D_C$, where D_F and D_C are the concentrations of free and bound DNA respectively, then according to the law of mass $D_F = D_T - D_C$ and assuming that $P_C = D_C$,

$$\frac{F_{obs} - F_o}{F_o} = \frac{f_C - f_F}{f_F} \nu \quad (4.4)$$

where $(f_C - f_F)/f_F$ is a constant.

The titration curve is constructed by plotting $(F_{obs} - F_o)/F_o$ versus the total concentration of the protein (P_T). If a simple model, such as reversible binding of protein to a single site on a DNA molecule, can be applied, the value of equilibrium association constant, that reflects the macroscopic affinity of the interaction, can be directly obtained from the analysis of this curve using a known relation between the concentration of complex and total concentration of protein at equilibrium:

$$\nu = \frac{D_T + P_T + K_d - \sqrt{(D_T + P_T + K_d)^2 - 4 \cdot D_T \cdot P_T}}{2 \cdot D_T} \quad (4.5)$$

where K_d is the equilibrium dissociation constant. However, in some cases, e.g. protein binds in several modes, each possessing a different spectroscopic signal change, the linear relationship between the signal change and the fractional saturation does not hold. Therefore, model independent equilibrium binding isotherms [Lohman and Bujalowski 1991] have to be used for the analysis. These isotherms are obtained from series of titration curves at different concentrations of DNA as described in [Lohman and Bujalowski 1991].

Experimental conditions

In this study titration experiments were performed at fixed concentration of DNA (20 nM of oligonucleotides bFC, uFC and nsFC) with step-wise increases of the concentration of homeodomain (Bcd, Ubx and En) from 5 nM to 180 nM. Changes in the steady state fluorescence intensities of fluorescein and Cy3 on double-labeled oligonucleotides (Table 4.1) were employed to monitor the formation of homeodomain-DNA complexes. Fluorescence emission spectra, covering the emission range of donor and acceptor, were measured as described in chapter 4.6. The recorded fluorescence spectra were decomposed into the donor and acceptor components as described in [Clegg 1992]. Fluorescence intensity values of donor at 525 nm and of acceptor at 562 nm were used to calculate the titration curves. Furthermore, fluorescence excitation spectra of donor and acceptor pair were acquired at each titration step, detecting the emission of acceptor at 600 nm. Samples were prepared in a buffer, containing 25 mM Na-HEPES, pH 7.0, 100 mM NaCl, 10^{-3} % surfactant P20.

4.8 Measurements of fluorescence resonance energy transfer (FRET)

Background

Fluorescence resonance energy transfer is a photophysical process by which energy is transferred from a fluorophore (the energy-donor D) in an excited state to another chromophore (the energy acceptor A) by means of intermolecular long-range dipole-dipole coupling [Förster 1946]. The energy transfer is non-radiative: acceptor does not reabsorb the photons emitted by donor.

There are four essential requirements for an effective energy transfer over distances from 10 Å to 100 Å:

- 1) The emission spectrum of the donor and the absorption spectrum of the acceptor must overlap adequately.
- 2) The quantum yield of donor in the absence of acceptor, Φ_D should be sufficiently high e.g. $\Phi_D \geq 0.1$.
- 3) The absorption coefficient of acceptor ϵ_A should be sufficient, e.g. $\epsilon_A \geq 1 \text{ mM}^{-1} \cdot \text{cm}^{-1}$.
- 4) The emission dipole of the donor and the absorption dipole of the acceptor must be oriented favorably to each other or at least one must have a certain degree of rotational freedom to allow effective dipole-dipole interactions between the two chromophores (see κ^2 below).

The rate of Förster dipole-dipole energy transfer is given by the following equation [Förster 1946]:

$$k_T = \frac{1}{\tau_D} \left(\frac{R_0}{R} \right)^6 \quad (4.6)$$

where the τ_D is the fluorescence lifetime of donor (in the absence of acceptor), R is the distance between the donor and acceptor, R_0 is the distance parameter (also called “characteristic Förster distance”), which depends on the spectroscopic parameters of the dyes and their mutual orientation:

$$R_0^6 = 8.79 \times 10^{-28} \times (n^{-4} \Phi_D \kappa^2 J(\lambda)) \quad (4.7)$$

where the n is refractive index of the medium and Φ_D is the fluorescence quantum yield of the donor. κ^2 is the orientation factor of dipole-dipole interaction and is determined by the unit vectors \vec{r}_D and \vec{r}_A representing the orientation of transition dipoles of the donor and acceptor respectively and the vector \vec{r}_{DA} connecting their origins

$$\kappa = \vec{r}_D \cdot \vec{r}_A - 3 \left(\vec{r}_{DA} \cdot \vec{r}_D \right) \left(\vec{r}_{DA} \cdot \vec{r}_A \right) \quad (4.8)$$

$J(\lambda)$ is the normalized spectral overlap integral, given by

$$J(\lambda) = \frac{\int_0^{\infty} F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda}{\int_0^{\infty} F_D(\lambda) d\lambda} \quad (4.9)$$

where $F_D(\lambda)$ is the fluorescence emission at wavelength λ , and $\varepsilon_A(\lambda)$ is the molar absorption coefficient of acceptor at the wavelength λ . The characteristic Förster distance $R_0 = 56 \text{ \AA}$ for the fluorescein-Cy3 pair was obtained by using the value of the spectral overlap integral, calculated numerically as $3.58 \times 10^{32} \text{ nm}^6 \text{ mol}^{-1}$ [Jares-Erijman and Jovin 1996], fluorescence quantum yield of donor $\Phi_D = 0.45$ [Eis and Millar 1993], $n = 1.33$ and $\kappa^2 = 2/3$ (assuming rapid orientational randomization of donor and acceptor transition dipoles; for a justification of this assumption see chapter 5.3).

Once the R_0 is determined, the donor-acceptor distance R can be calculated from the experimental value of energy transfer efficiency E via expression:

$$E = \frac{R_0^6}{R_0^6 + R^6} \quad (4.10)$$

Various approaches exist for calculation of FRET efficiency, based on changes in fluorescence intensity, anisotropy or lifetime of the fluorophores [Clegg 1992; Wu and Brand 1994; Jares-Erijman and Jovin 2003]. In this work I used a method based on the normalization of sensitized acceptor fluorescence by a directly excited fluorescence signal of acceptor, which has proven to be useful for donor-acceptor pairs covalently attached to DNA/RNA structures [Clegg 1992]:

$$\frac{F_{\text{em}}^A(\lambda_1, \lambda')}{F_{\text{exc}}^A(\lambda_2, \lambda'')} = \left(Ed^+ \frac{\varepsilon^D(\lambda')}{\varepsilon^A(\lambda'')} + \frac{\varepsilon^A(\lambda')}{\varepsilon^A(\lambda'')} \right) \frac{\phi^A(\lambda_1)}{\phi^A(\lambda_2)} \quad (4.11)$$

where $F_{\text{em}}^A(\lambda_1, \lambda')$ is the sensitized acceptor fluorescence intensity at wavelength λ_1 , excited at wavelength λ' , $F_{\text{exc}}^A(\lambda_2, \lambda'')$ is the acceptor fluorescence intensity at wavelength λ_2 , obtained from its excitation spectrum at excitation wavelength λ'' , where only acceptor is excited, d^+ is the fractional donor labeling, ε is the molar absorption coefficient of donor or acceptor (superscript) at the indicated wavelength, ϕ^A is an emission spectrum shape function of the acceptor, where the integral over the λ is proportional to the quantum yield.

Experimental conditions

The fluorescence emission spectra of fluorescein-Cy3 pair were measured by exciting the donor (fluorescein) at wavelength $\lambda' = 493$ nm and the excitation spectra of acceptor (Cy3) were recorded by observing the emission at wavelength $\lambda_2 = 600$ nm. The obtained emission and excitation spectra were decomposed into donor and acceptor components as described in [Clegg 1992] and the acceptor fluorescence spectra were extracted for further analysis. The intensity value of sensitized acceptor fluorescence at the emission wavelength $\lambda_1 = \lambda_2 = \lambda = 600$ nm was used for the calculation of E , leading to a simplified expression of the normalized enhanced acceptor fluorescence:

$$\frac{F_{em}^A(\lambda, \lambda')}{F_{exc}^A(\lambda, \lambda'')} = Ed^+ \frac{\varepsilon^D(\lambda')}{\varepsilon^A(\lambda'')} + \frac{\varepsilon^A(\lambda')}{\varepsilon^A(\lambda'')} \quad (4.12)$$

The intensity of the directly excited acceptor fluorescence was taken from the excitation spectrum of acceptor at the excitation wavelength $\lambda'' = 552$ nm. The molar absorption coefficients of donor and acceptor at the selected wavelengths λ' and λ'' used for the calculation of E were $73 \text{ mM}^{-1}\text{cm}^{-1}$ and $130 \text{ mM}^{-1}\text{cm}^{-1}$, respectively. $\varepsilon^A(\lambda')/\varepsilon^A(\lambda'')$ was directly obtained from the excitation spectrum of the acceptor.

4.9 Stopped-flow kinetics

Measurements of homeodomain-DNA association kinetics were performed on a commercial stopped-flow apparatus SX.18MV (Applied Photophysics Ltd., UK) equipped with 20 μl optical detection cell. The source of light was a 150 W Xenon arc lamp (Hamamatsu, Japan) coupled to a 05-109 SpectraKinetic monochromator, which was used for the excitation wavelength selection. The light from monochromator output was directed by an optical fiber into the 20 μl optical cell (excitation pathlength is 1 cm, emission pathlength – 2 mm). In the emission path a long-pass filter KV550 (SCHOTT, Germany) was used to separate the excitation light from the fluorescence of Cy3 label. During the experiments of simultaneous measurement of donor and acceptor fluorescence kinetics, fluorescence of donor (fluorescein) was detected on a second optical detection window in the direction perpendicular to the excitation light beam using a band-pass filter BP520-540 (Delta Light & Optics, Lingby,

DK). Emission light passed through the emission filters was detected by a 12-stage end window emission photomultiplier R928 (Hamamatsu, Japan). The photomultiplier (PMT) output signal was detected by the photometric control unit and further redirected to ADC card. Digitized signal from the ADC card was recorded by an Acorn RISC PC computer. In order to avoid possible changes in fluorescence intensity caused by the fluctuations of the excitation light intensity, the measured signal was normalized by the excitation light intensity simultaneously detected⁶ by the side window absorption PMT as shown in Figure 4.1. The amount of the background fluorescence intensity coming from the excitation light leaked through the emission filter was subtracted and the kinetic fluorescence intensity profiles were calculated [see Appendix A2.]

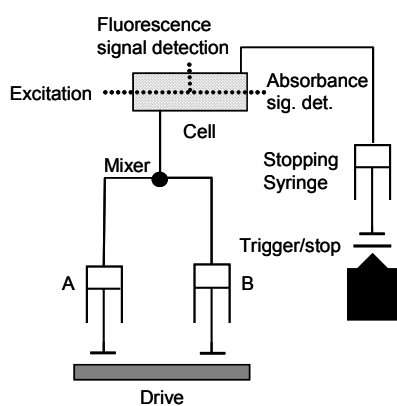


Figure 4.1. Principal scheme of the stopped-flow system

The fluorescence of the donor-acceptor pair on the oligonucleotides was excited at 490 nm. Experiments with oligonucleotides containing only Cy3 labels were performed by exciting the fluorescence at 517 nm (emission spectra are shown in Figures 5.4 and 5.6 A in chapter 5). The binding reaction was initiated by mixing two equal volumes (50 μl each) of oligonucleotide (sample A) and protein (sample B) solutions respectively. Acquisition was started when the displacement of the piston in the stop-syringe activated the trigger and terminated the flow of the mixed solution (Figure 4.1). The measurements were carried out in a buffer containing 25 mM, 100 mM NaCl, 0.001% v/v surfactant P20, pH 7.0 at 5 $^{\circ}\text{C}$. The concentration of the oligonucleotide solution (before mixing) was 40 nM. Concentration of homeodomain (Bcd or Ubx) solution (before mixing) varied from 20 nM to 160 nM. Sample solutions were degassed and filtered with a sterile 0.22 μm filter (Schleicher & Schuell,

⁶ Assuming that the absorption of the sample solution at the excitation wavelength is negligibly small and does not change during the time of measurement.

Dassel, Germany). For each injection, 1,000 data points were recorded in a time window of either 0.2, 0.5, 1, 10, or 50 s depending on reaction conditions in order to obtain optimal resolution of the ascending or descending phase. The time scale of the measured kinetic traces was corrected for the dead-time of the stopped flow system. The determination of the dead-time (1.4 ms) was performed by measuring the reaction rates of 2,6-dichlorophenolindophenol reduction by L-ascorbic acid according to a published method [Tonomura et al. 1978].

The obtained kinetic fluorescence intensity profiles were analyzed with the kinetic data analysis program Pro-KII (Applied Photophysics Ltd., UK). A global analysis was performed on the series of the data sets acquired at different concentrations of homeodomain. The principle of this procedure is described in detail in Appendix A3.

4.10 Fluorescence correlation spectroscopy (FCS)

Background

The principle of FCS is based on the confocal detection of concentration fluctuations of sample molecules in a very small detection volume ($\approx 1\text{fL}$) monitored by fluorescence excited with a laser beam focused through a microscope objective (Figure 4.2 A). The emission is detected through the same optics; the excitation and emission wavelengths are separated by dichroic mirrors and filters.

The primary data obtained in FCS measurement is the time-dependent fluorescence intensity $F(t)$, which is proportional to the number of particles in the detection volume at time t . The autocorrelation function of $F(t)$ contains the information related to the diffusion of the fluorophores. The normalized autocorrelation function $G(\tau)$ is computed as

$$G(\tau) = \frac{\langle F(t)F(t+\tau) \rangle}{\langle F(t) \rangle^2} \quad (4.13)$$

In order to determine quantities such as diffusion coefficients, concentrations or reaction rate constants, the measured $G(\tau)$ has to be fitted by a theoretical correlation function which is based on the model that contains these quantities as free parameters. For example, $G(\tau)$ for a single fluorescent species with diffusion coefficient D and molar concentration C , assuming

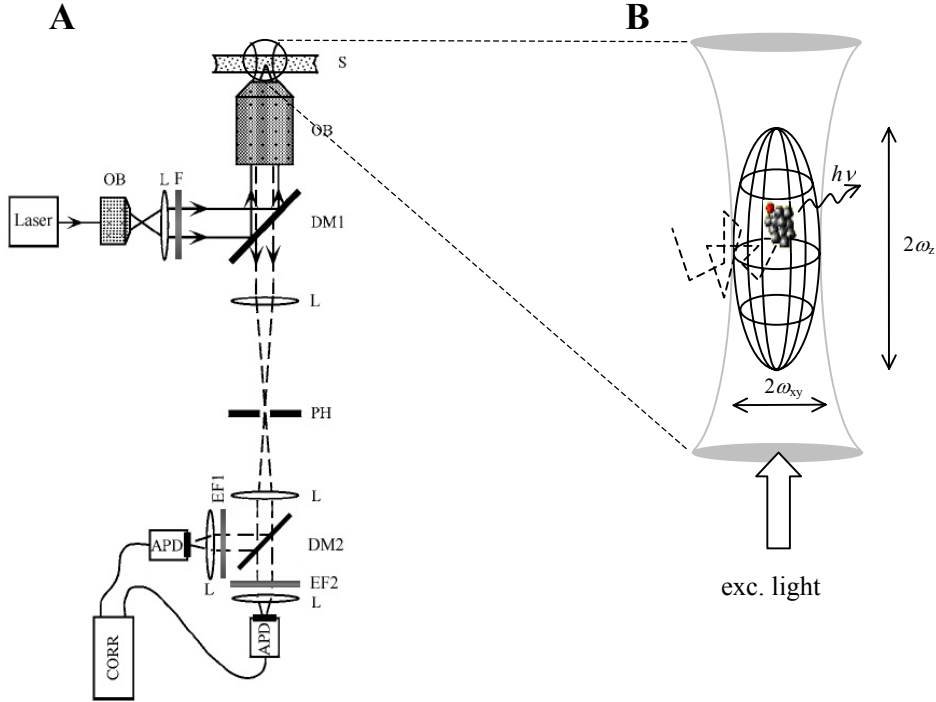


Figure 4.2. Panel **A**: Schematic principle of the FCS using two-channel detection. Notations: OB – objective, L – lenses, F- excitation filter, DM1 and DM2 – dichroic mirrors, S – sample, PH – pinhole, EF1 and EF2 – emission filters, CORR – correlator. Panel **B**: Geometry of the detection volume.

Gaussian profiles for the detection volume in the lateral and axial directions, is equal to [Rigler et al. 1993]

$$G(\tau) = \frac{1}{CV_{\text{eff}}} \cdot \frac{1}{\left(1 + \frac{4D\tau}{w_{xy}^2}\right) \sqrt{1 + \frac{4D\tau}{w_z^2}}} + 1 \quad (4.14)$$

Here, V_{eff} is the effective observation volume, which depends on the geometry of the focus for excitation and emission, ω_{xy} and ω_z are the half-widths of the focus in the x - y plane (the observation plane of the lens) and the z -direction, respectively (Figure 4.2 B). V_{eff} , ω_{xy} and ω_z can be measured independently by calibration with a solution of a fluorophore of known concentration and diffusion coefficient. Introduction of the average particle number $N = CV_{\text{eff}}$, the average diffusion time $\tau_D = \omega_{xy}^2/4D$ and of the structural factor $S = \omega_z/\omega_{xy}$ enables one to simplify the expression of $G(\tau)$

$$G(\tau) = \frac{1}{N} \cdot \frac{1}{\left(1 + \frac{\tau}{\tau_D}\right) \sqrt{1 + \frac{\tau}{S^2\tau_D}}} + 1 \quad (4.15)$$

Taking into account the fraction T of the fluorophore molecules in the triplet state, leads to the following expression of $G(\tau)$ [Widengren et al. 1995]

$$G(\tau) = \frac{1}{N(1-T)} \cdot \frac{1}{\left(1 + \frac{\tau}{\tau_D}\right) \sqrt{1 + \frac{\tau}{S^2 \tau_D}}} \left(1 - T + T \cdot e^{-\frac{\tau}{\tau_T}}\right) + 1 \quad (4.16)$$

where τ_T is a lifetime of the triplet state.

A logical consequence in extending the established approach of autocorrelation is the concept of cross-correlation, the analysis of underlying mechanisms relating the measured quantities (e.g. F_a and F_b) with each other:

$$G^{\times}(\tau) = \frac{\langle F_a(\tau) F_b(t + \tau) \rangle}{\langle F_a(t) \rangle \langle F_b(t) \rangle} \quad (4.17)$$

A cross-correlation of two fluorescence signals arising from two different fluorescent labels emitting at two distinct wavelengths (dual-color) has been shown to be a sensitive method to analyse the molecular association processes that otherwise cannot be detected by means of fluorescence autocorrelation [Schwille 1997]. The suppression of the background fluorescence signal arising from the free fluorophores in solution has been one of the major reasons for using the fluorescence cross-correlation analysis to study the diffusion properties of the double-labeled oligonucleotides. In fact, the dual-color cross-correlation function is equal to the autocorrelation function Eq. (4.15) if the detection efficiencies are equal and the detection volumes exactly overlap for the two emission wavelengths (which is generally true as a first approximation for modern FCS systems).

Experimental conditions

FCS measurements were performed on the FCS microscope based on Zeiss Axiovert 35 model (Carl Zeiss, Jena, Germany) equipped with a C-Apochromat 40x 1.2 numerical aperture water immersion objective (Carl Zeiss, Jena, Germany) to which a multi-channel confocal side-port with an adjustable pinhole was added.

The fluorescence of fluorescein and Cy3 labels on the oligonucleotide bFC was excited using the 488 nm line of the Argon-ion laser 2313-150MLYV (Uniphase, Eching, Germany). Laser power was attenuated to 50 μ W using neutral density filters (New Focus,

Santa Clara, CA, USA) and coupled into a single mode optical fiber with a Fiber Coupler (Point Source, Winchester, UK). In the excitation path a band-pass filter BP488/5 (Delta Light & Optics, Lingby, DK) was used to block the light coming from other sources. For the excitation and detection of fluorescence, a dichroic mirror 505DRLP (Omega Optical, Brattleboro, VT, USA) was used. After passing through a pinhole (the diameter of which was controlled from the computer via an DAC card) the fluorescence of donor and acceptor was divided with a dichroic mirror 555DRLP (Omega Optical, Brattleboro, VT, USA) into two beams for separate detection of donor and acceptor fluorescence, using band-pass filters respectively 535DF45 and 580DF30 (Omega Optical, Brattleboro, VT, USA) and avalanche photodiodes (APDs) SPCM-AQ-132 (EG&G Optoelectronics, Canada). Signals measured on the APDs were directed into a ALV-5000 Multiple Tau Digital Correlator Board (ALV-Laservertriebsgesellschaft mbH, Langen, Germany), where the cross-correlation between the donor and acceptor signals was calculated. Control of the focus positioning, pinhole diameter and data acquisition was implemented with software Labview v 4.0 (National Instruments, Austin, TX, USA) via a custom-built FCS control unit based on a microcontroller SAB80C537 (Siemens, Germany).

Measurements were performed at room temperature in solutions containing 25 mM Na-HEPES, pH 7.0, 100 mM NaCl, 0.001% surfactant P20. The concentration of the oligonucleotide bFC was 20 nM. The sample was analysed in a compartment of a Lab-Tek chambered borosilicate coverglass (Nalge Nunc Int., Rochester, NY, USA). Sample volume was 500 μ L. The laser beam was focused \sim 50 μ m above the bottom plate of the sample vial. The duration of measurements was usually 15 s. Cross-correlation curves obtained from 20 individual measurements of the sample containing free oligonucleotide bFC were fitted using the $G(\tau)$ expression (4.16) and results were averaged.

For the time course measurements of Bcd homeodomain interaction with the oligonucleotide bFC a 3-fold molar excess of the Bcd homeodomain was added to the sample containing 20 nM of the oligonucleotide bFC. A total of 50 measurements were performed with a period of 30 s. Individual cross-correlation curves were fitted using a theoretical correlation function defined in (4.16).

In order to determine the parameters V_{eff} , ω_{xy} and ω_z (see Figure 4.2 B and Eq. (4.14)), the autocorrelation curve of 10 nM fluorescein in water ($D_t = 5.9 \times 10^{-6}$ cm²/s) was recorded at the same configuration of the FCS system as used for the measurements of the

fluorescently labeled oligonucleotide bFC. The analysis of the obtained autocorrelation curve yielded $V_{\text{eff}} = 1.5 \pm 0.2 \text{ fL}$, $\omega_{xy} = 0.38 \pm 0.02 \text{ } \mu\text{m}$, $\omega_z = 1.9 \pm 0.15 \text{ } \mu\text{m}$.

4.11 Molecular modeling

A model of the Bcd_{NHis} tertiary structure was constructed based on the primary sequence homology with Engrailed homeodomain using the tertiary structures: 1HDD (Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution [Kissinger et al. 1990]), 2HDD (Engrailed homeodomain Q50K variant DNA complex at 1.9 Å resolution [Tucker-Kellogg et al. 1997]), 3HDD (Engrailed homeodomain-DNA complex at 2.2 Å resolution [Fraenkel et al. 1998]). Calculations were performed using the ProModII program [Peitsch 1996] on the SWISS-MODEL Server (<http://swissmodel.expasy.org/>) [Schwede et al. 2003]. Structures of the amino acids in the linker containing the His-tag at N-terminus and of three amino acids at C-terminus of the protein sequence could not be simulated using this program. These amino acids were included later by using the molecular modeling program SYBYL. The modeled structure of the entire Bcd_{NHis} protein was further energy-minimized using the AMBER molecular dynamics package [Ponder and Case 2003]. A tertiary structure of the double-stranded 15 bp oligonucleotide “b” was generated using the molecular modeling program SYBYL.

5. Results

5.1 Spectroscopic characterization of homeodomain-DNA complexes at equilibrium

Before starting measurements of homeodomain-DNA binding kinetics, a series of steady state experiments were performed under equilibrium conditions.

5.1.1 Changes in intrinsic fluorescence of homeodomains upon binding to DNA

A classical and minimally invasive way to study protein-DNA interactions with fluorescence spectroscopy is to employ fluorescence changes (if there are any available) of intrinsic fluorophores of the protein. Typically, the fluorescence of tryptophan (Trp) is used for the detection because of its relatively high extinction coefficient and quantum yield, compared with other two aromatic aminoacids: tyrosine (Tyr) and phenylalanine (Phe). However, in some cases (when tryptophans are absent or highly quenched) the fluorescence of tyrosine and/or phenylalanine can provide useful information about protein-DNA interaction.

The homeodomains selected for this study (Bcd, Ubx and En) contain a single tryptophan amino acid (Trp48). Bcd and En have one tyrosine and Ubx has three tyrosines. Phenylalanines are not considered here, since their contributions to the total fluorescence are generally negligible.

In this work the fluorescence spectra of Bcd, Ubx and En homeodomains were measured in the presence and in the absence of specific (“b”, “u”) and non-specific (“ns”) non-labeled double-stranded 15-mer oligonucleotides (see Table 4.1 in chapter 4.3). The concentration of the protein and the oligonucleotide in sample solution was 5 μ M each. Tryptophan was selectively excited at 295 nm and both Trp and Tyr at 280 nm. The fluorescence spectra of all three homeodomains obtained when exciting at 280 nm (panels A and C in the Figures 5.1 and 5.2) show a decrease of fluorescence intensity by ~28% in the presence of specific DNA and by ~40% in the case of non-specific DNA. This fluorescence signal is largely determined by the Tyr25, which is quenched upon its interaction with phosphate groups of DNA backbone, as can be seen from the crystal structures of Ubx-DNA and En-DNA complexes (see Figure 2.3 in chapter 2.2). However tryptophan fluorescence

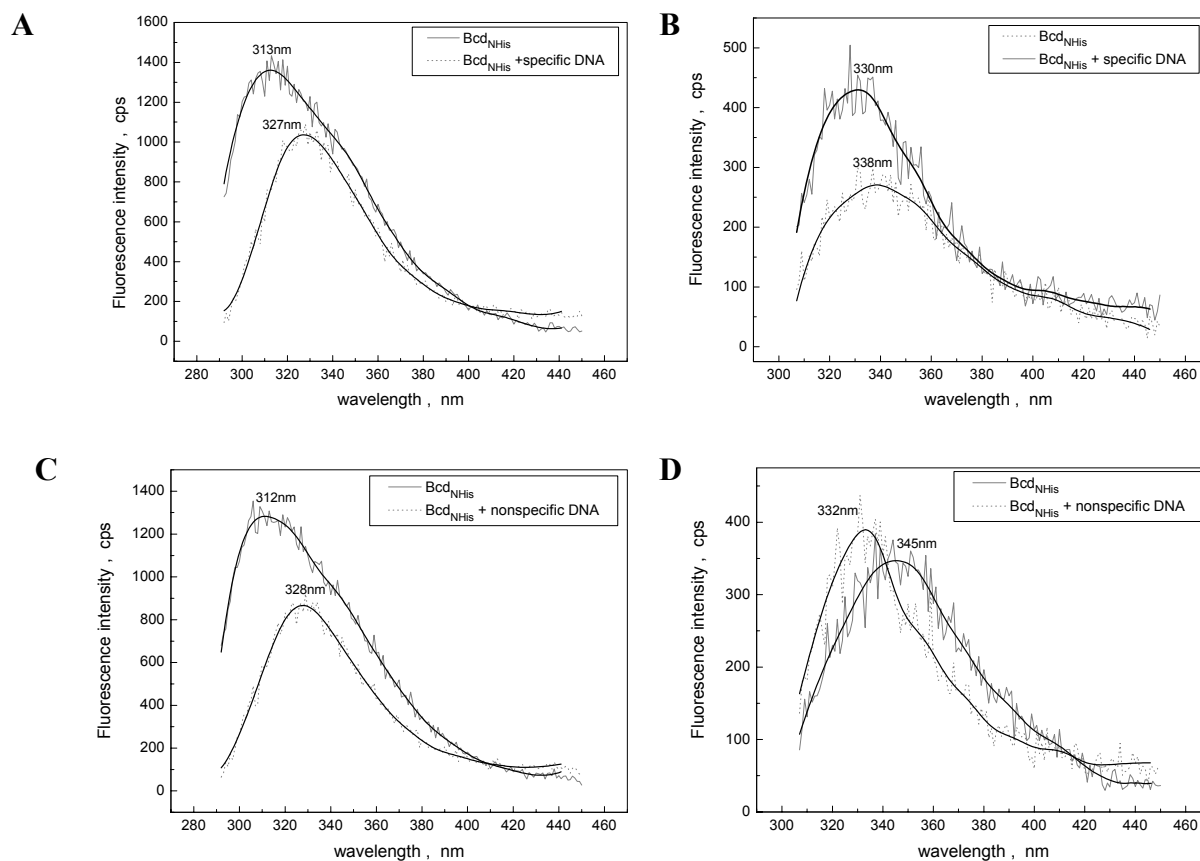


Figure 5.1. Fluorescence spectra of 5 μM Bcd_{NHIS} in 25 mM Na-HEPES, 100 mM NaCl, 0.001% surfactant P20, pH 7.0 buffer at 5.0 °C in presence and absence of specific (panels **A** and **B**) and nonspecific (panels **C** and **D**) double-stranded 15-mer oligonucleotides. The samples were excited at 280 nm (panels **A** and **C**) and at 295 nm (panels **B** and **D**). The concentration of the protein and the oligonucleotide in sample solution was 5 μM each.

exhibits an opposite effect – the intensity increases upon homeodomain binding to DNA, indicating that the environment of Trp48 has changed such that the quenching of its fluorescence by neighboring residues, (e.g. as Phe8), become less effective. In addition, there is a blue-shift (by ~ 8 nm for the specific complex and ~ 13 nm for the non-specific complex) of the Trp fluorescence spectrum of the bound Bcd in comparison with free Bcd, indicating that the environment of Trp48 in this protein becomes more hydrophobic after DNA is bound. Since the tyrosine fluorescence spectrum has a maximum at 304 nm and tryptophan at 330 nm (in a hydrophobic environment), it is obvious that quenching of tyrosine fluorescence and an increase of tryptophan fluorescence intensity will result in the red shift of the fluorescence spectrum when excited at 280 nm, as it is well expressed in the case of Bcd and to a lesser degree in Ubx-DNA binding.

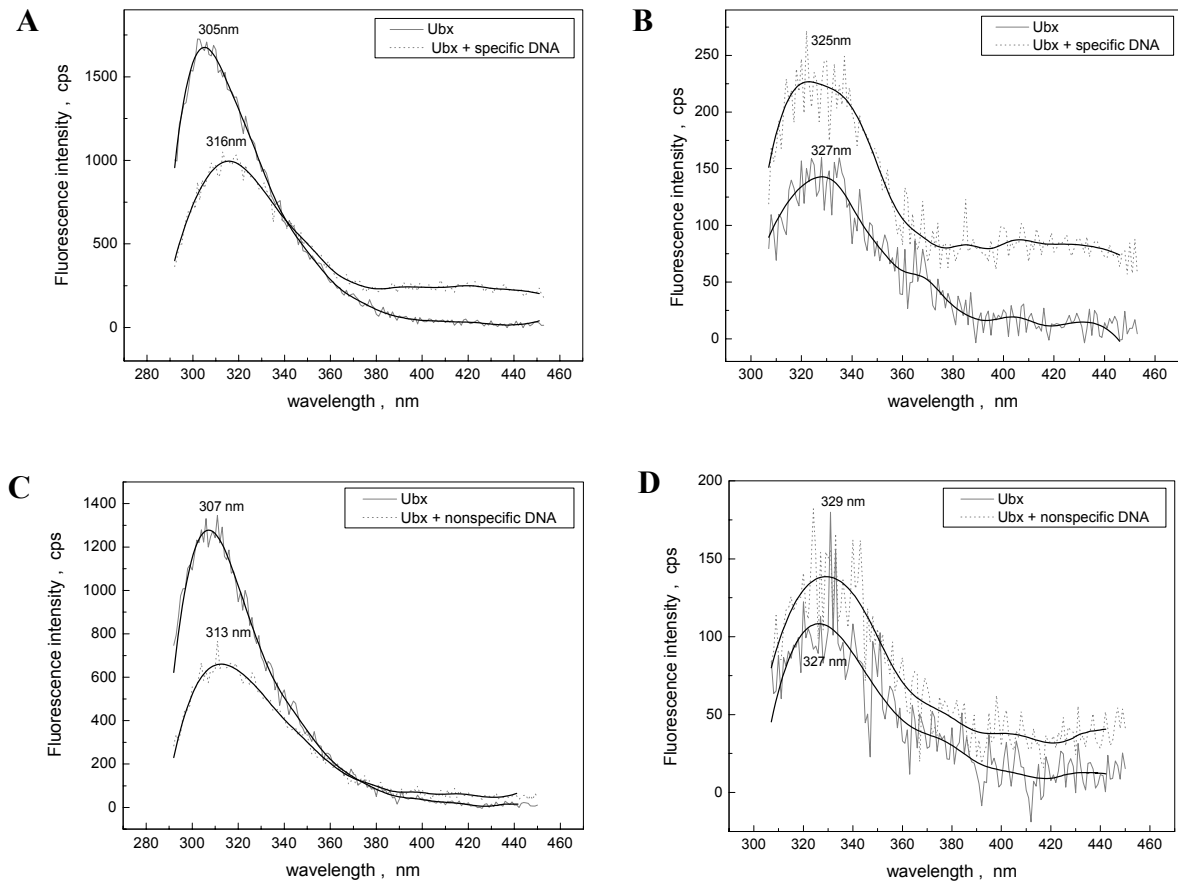


Figure 5.2. Fluorescence spectra of 5 μ M Ubx in 25 mM Na-HEPES, 100 mM NaCl, 0.001% surfactant P20 pH 7.0 buffer at 5.0 $^{\circ}$ C in presence and absence of specific (panels **A** and **B**) and nonspecific (panels **C** and **D**) doublestranded 15-mer oligonucleotide. The samples were excited at 280 nm (panels **A** and **C**) and at 295 nm (panels **B** and **D**). The concentration of the protein and the oligonucleotide in sample solution was 5 μ M each.

5.1.2 Changes of circular dichroism spectra of homeodomains upon binding to DNA

Measurements of circular dichroism spectra (CD) of the homeodomain proteins in its free state and in complex with specific and nonspecific DNA were performed in order to detect possible changes in secondary structure of homeodomains arising from their binding to DNA. The CD spectra of homeodomains Bcd and Ubx (Figure 5.3, panels A and B) show two minima at 208 and 222 nm, characteristic of α -helical secondary structure, in accordance with similar data obtained for other homeodomains [Ades and Sauer 1994; Damante et al. 1994; Carra and Privalov 1997]. A comparison of the CD spectra of free Bcd and Ubx with the CD spectra of these homeodomains in a complex with respective specific DNA sequences shows an increase (by up to 24%) in ellipticity at 222 nm, indicating formation of additional α -helical secondary structures (possibly in the C-terminal and/or N-terminal regions) of the

proteins. Such a behavior is typical for many other DNA binding proteins [Gray 1996; Lagunavicius et al. 1997; Kamadurai et al. 2003]. Engrailed, in contrast to Bcd and Ubx, exhibits only a very small increase in ellipticity at 222 nm when bound to the specific oligonucleotide (Figure 5.3, C), indicating a weak influence of protein-DNA interactions to the secondary structure of the homeodomain. In a complex with nonspecific DNA no significant change in the ellipticity of En was observed. Binding of Bcd to the non-specific DNA seems to cause even a larger increase in the α -helicity of the protein relative to the case of specific DNA (Figure 5.3, panels A and B). This finding suggests that a relative orientation of the protein and DNA molecule in a complex (which might differ in complexes with the specific and nonspecific DNA) is more critical for the induction of the secondary structure changes of the protein than the specific contacts between protein and DNA.

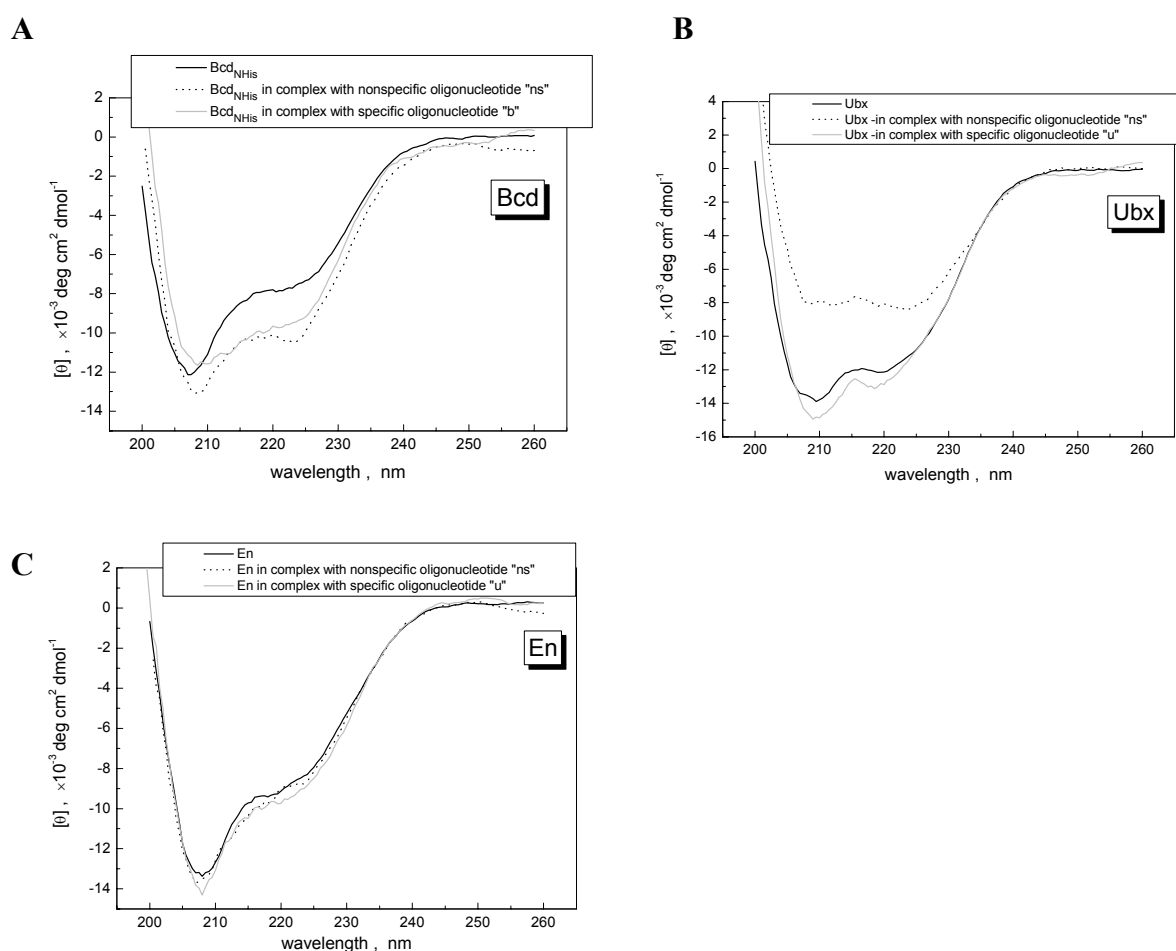


Figure 5.3. CD spectra of Bcd (panel **A**), Ubx (panel **B**) and En (panel **C**) in presence and absence of specific (“b” and “u”) and nonspecific (“ns”) oligonucleotides. Spectra were acquired in 25 mM Phosphate buffer pH 7.0, containing 100 mM NaCl at 5 °C. The concentration of the protein and the oligonucleotide in sample solution was 5 μ M each. The CD spectrum of the oligonucleotide was subtracted from the respective CD spectrum of the protein-DNA complex, assuming that it does not change upon binding.

A comparison of the CD spectrum of free Ubx and its complex with nonspecific DNA reveals a striking loss of α -helical content in a secondary structure of the protein after binding to the DNA. The effect could be accounted for processes such as complex-complex interaction, aggregation etc.

5.1.3 Use of fluorescent labels to study the homeodomain-DNA interaction

The fluorescent labels are also widely used to investigate protein-DNA interactions. For this analysis both Bcd homeodomain and DNA were fluorescently labeled and tested for their applicability for measurements of the homeodomain-DNA binding equilibrium and kinetics.

Fluorescent labeling of DNA

As first, fluorescently labeled double-stranded 15-mer oligonucleotides (b5F and b5C), containing a fluorescein or Cy3 molecule at 5'-end of one of the strands (see Table 4.1 in chapter 4.3), were used for spectroscopic measurements. These oligonucleotides contained a specific binding sequence of Bcd (TAATCC). Buffer conditions were optimized in order

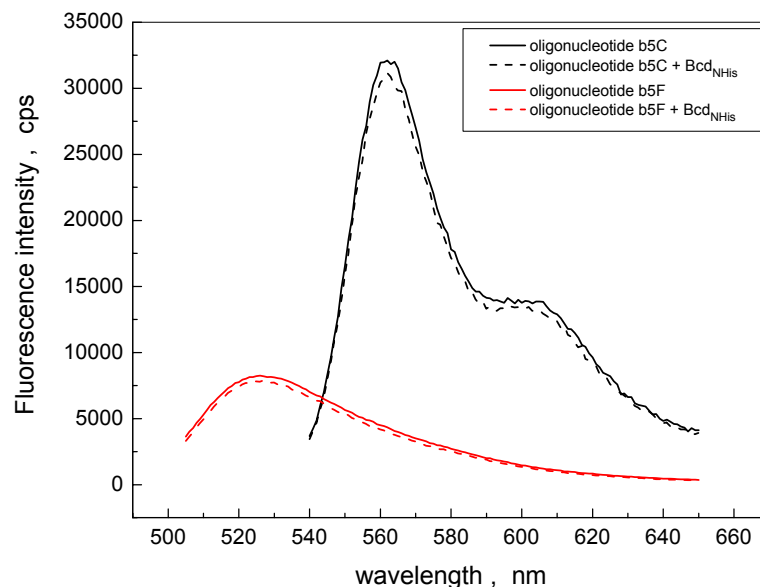


Figure 5.4. Changes in fluorescence spectra of fluorescein and Cy3 of oligonucleotides b5F and b5C respectively, induced by the binding of Bcd. Concentrations of each oligonucleotide and the protein in the sample solution was 20 nM. Spectra were acquired in 25 mM Na-HEPES, 100 mM NaCl, 0.001% surfactant P20 pH 7.0 buffer at 5.0 °C. Fluorescein fluorescence was excited at 490 nm and that of Cy3 at 530 nm.

to improve the stability of fluorescence signal from these oligonucleotides in solution. The best signal stability was achieved in a solution containing 25 mM Na-HEPES, 100 mM NaCl, 0.001% v/v surfactant P20 at pH 7.0. These conditions were chosen as standard for the most of the further steady state and kinetic experiments.

The measured steady state fluorescence spectra showed, that binding of the Bcd homeodomain to the fluorescently labeled oligonucleotides b5F and b5C induces small changes ($\sim 4-6\%$) in the fluorescence intensities of both fluorophores (Figure 5.4). A similar behavior was obtained for the oligonucleotides labeled at 3'-end (not shown). These changes are too small to obtain reliable information about the homeodomain-DNA interaction kinetics, and therefore other approaches were considered.

Fluorescent labeling of the protein

As an alternative approach to utilizing fluorescence signals of external fluorophores, fluorescent labeling of Bcd homeodomain with several fluorescent probes was tested. For this purpose a mutant of Bcd homeodomain (Bcd_{CHis}-Cys) containing a cysteine aminoacid at the C-terminus was constructed, expressed and purified. Fluorescent labels Alexa 488, Alexa 546, IAEDANS and monobromobimane (mBBr) were used (Figure 5.5) for labeling this protein under the conditions described in (materials and methods).

Bcd_{CHis}-Cys protein labeled with Alexa 488 and Alexa 546 could not be used for spectroscopic measurements because the labeled protein precipitated during the labeling reaction. In order to avoid these problems smaller fluorophores like IAEDANS and mBBr were also tested. The protein was efficiently labeled and did not precipitate. However the dilution of the stock solution of thus labeled protein resulted in strong changes (5-8 fold) of its fluorescence intensity, suggesting the presence of possible aggregates of the labeled Bcd_{CHis}-Cys with an increased quantum yield of the fluorescent labels. Because of these

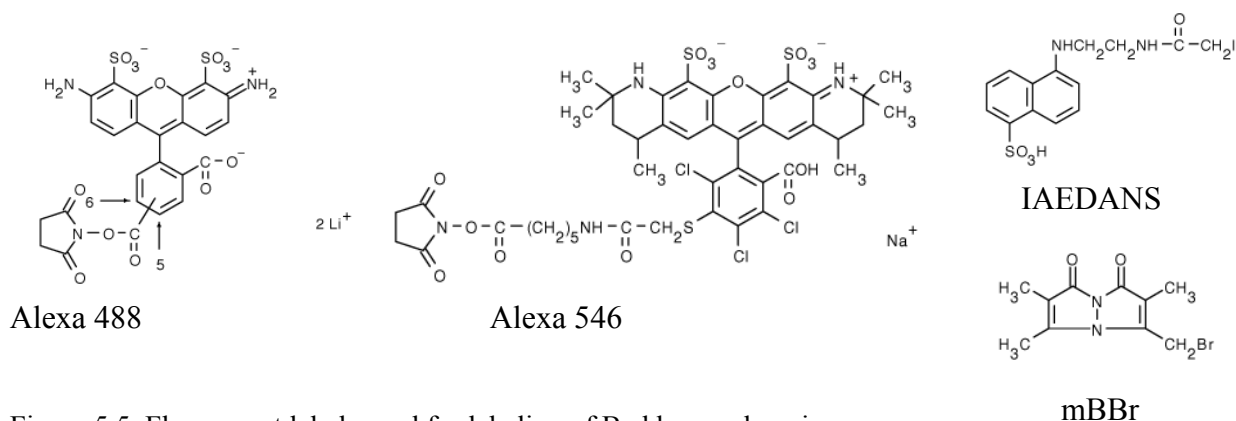


Figure 5.5. Fluorescent labels used for labeling of Bcd homeodomain.

difficulties the use of extrinsic fluorescent labels on the protein to study homeodomain-DNA interactions had to be abandoned.

Intramolecular FRET

Intramolecular FRET between fluorescent labels on the DNA has been shown to be sensitive to the interactions between DNA and proteins that bend it upon binding [Parkhurst et al. 2001]. The applicability of this approach to study homeodomain-DNA interactions was tested using a double-stranded 15-mer oligonucleotide (bFC), labeled with fluorescein (as a fluorescence energy donor) at 5'-end of one strand and with Cy3 (as a fluorescence energy acceptor) at 3'-end of the same strand. The characteristic Förster distance R_0 of this FRET pair is 56 Å [Jares-Erijman 1996], which approximately corresponds to the length of this oligonucleotide. The fluorescence spectra of the donor and acceptor were measured exciting the donor (fluorescein) at 490 nm in absence and presence of Bcd homeodomain. The results (Figure 5.6 A) show that binding of Bcd to the oligonucleotide bFC results in a decrease of fluorescence intensity of the donor (at 525 nm) by ~20% and an increase of fluorescence intensity of the acceptor (at 560 nm) by ~32%, indicating an increase of FRET efficiency between fluorescein and Cy3. These data support a conjecture that the conformation of the specific oligonucleotide is changed due to binding of Bcd. Figure 5.6 B shows that the intramolecular FRET on the DNA (particularly changes of acceptor fluorescence) can be

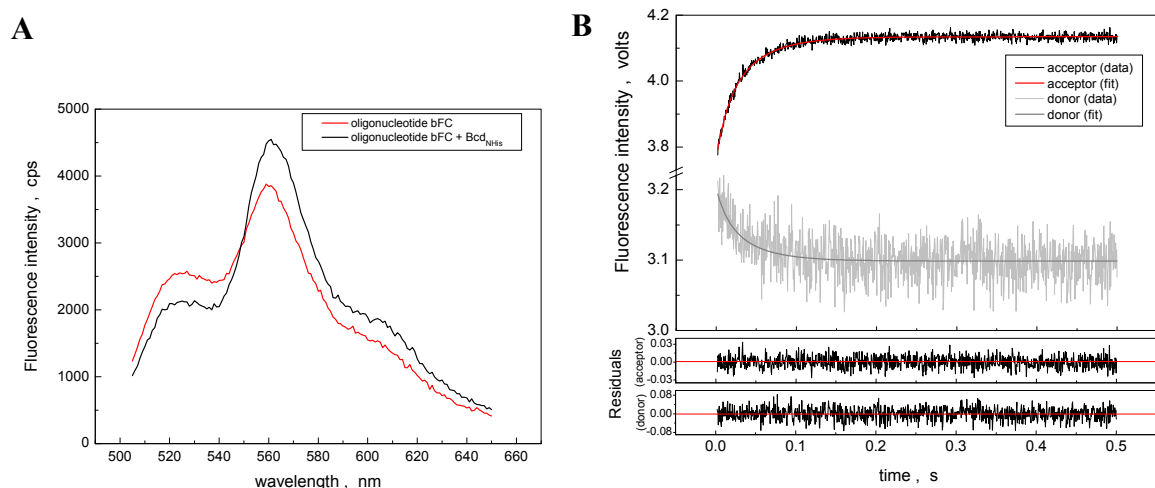


Figure 5.6. Changes of fluorescence intensity of donor (fluorescein) and acceptor (Cy3) due to FRET efficiency increase during the binding of Bcd_{NHis} to the oligonucleotide bFC at equilibrium (panel A) and in course of time (panel B). Data were recorded at 5.0 °C in a solution, containing 25 mM Na-HEPES, 100 mM NaCl, 0.001% surfactant P20 pH 7.0. Fluorescence was excited at 490 nm. The concentrations of the protein and the oligonucleotide were 20 nM.

effectively employed for the analysis of the homeodomain-DNA interaction kinetics.

Because of large fluorescence intensity changes and good signal to noise ratio, this approach was chosen as a main spectroscopic technique for further experiments to monitor the homeodomain-DNA binding at equilibrium and its transient kinetics.

5.2 Analysis of homeodomain-DNA binding at equilibrium conditions

A further step in the analysis of homeodomain-DNA interaction at equilibrium was the elucidation of macroscopic parameters of protein-DNA binding such as affinity constant and the stoichiometry of binding. In order to obtain this information, titration measurements were performed as described in chapter 4.7 of the methods section using double-labeled (with fluorescein and Cy3) oligonucleotides bFC, uFC, nsFC and homeodomains Bcd, Ubx and En.

At each titration step the fluorescence emission spectra of donor and acceptor (fluorescein and Cy3) were measured by exciting the donor at 493 nm. The fluorescence excitation spectra of the acceptor were also recorded (detecting emission at 600 nm) in order

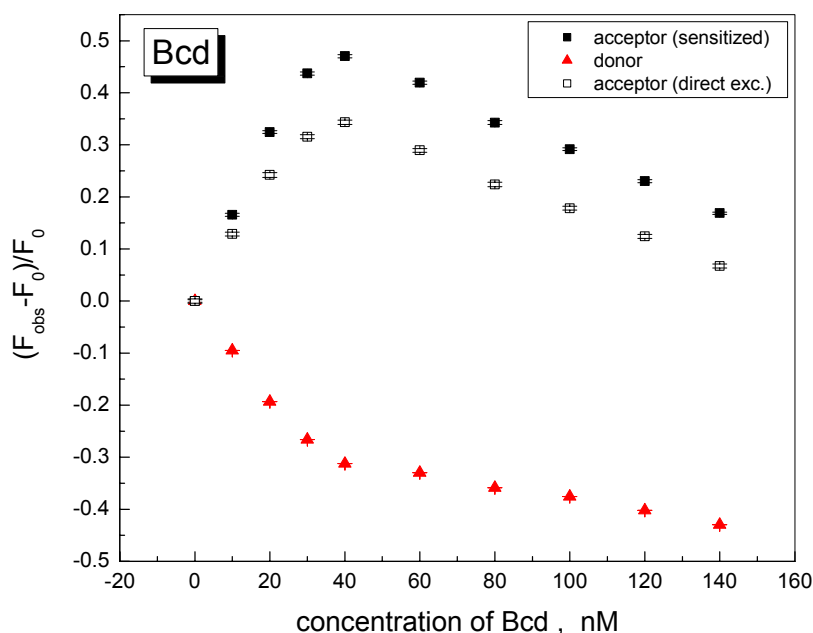


Figure 5.7. Binding curves of Bcd to the specific oligonucleotide bFC obtained at 5 °C in a buffer containing 25 mM Na-HEPES, pH 7.0, 100 mM NaCl, 0.001% surfactant P20. Fluorescence of donor was recorded at 525 nm (\blacktriangle) and sensitized emission of acceptor was recorded at 560 nm (\blacksquare) by excitation at 493 nm. Fluorescence of directly excited acceptor at 552 nm (\square) was recorded at 600 nm.

to calculate the FRET efficiency as well to evaluate the fluorescence properties of acceptor. Titration curves were calculated using the fluorescence intensity values of the donor at 525 nm and the values of sensitized (excited via donor at 493 nm) and directly excited (at 552 nm) emission intensity of the acceptor. The profiles of the titration curves of specific binding of Bcd and Ubx to bFC and uFC oligonucleotides (Figures 5.7 and 5.8), obtained from sensitized and directly excited emission of acceptor, are very similar. In both cases the curves have a bell shaped profile with a maximum at the protein/oligonucleotide ratio of ~2:1 (Figures 5.7 and 5.8). A linear increase in fluorescence signal at the protein/oligonucleotide molar ratios below 2:1 indicates that the value of equilibrium dissociation constant (K_d), which reflects the affinity of binding, is considerably lower than the total concentration (D_T) of the oligonucleotide. Under these conditions (called “stoichiometric binding”) virtually every protein molecule added to the solution is bound to the DNA, giving rise to the linear response in the titration curve until all sites are occupied. An accurate determination of K_d from this binding curve is not possible. In order to determine the dissociation constant from a titration curve it is required that $D_T \leq K_d$. Under these conditions the binding curve has a large nonlinear region allowing the determination of the equilibrium dissociation constant by fitting a theoretical model to the experimental data. Such an analysis was applied to the leading part

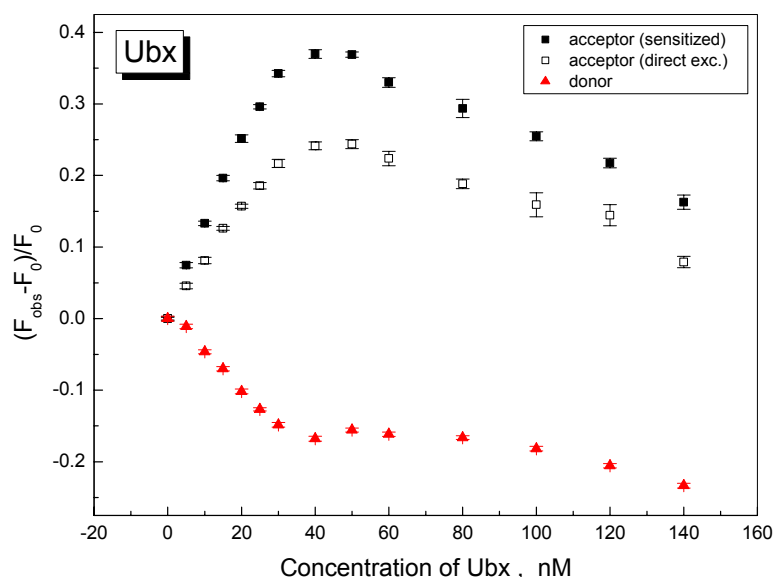


Figure 5.8. Binding curves of Ubx to the specific oligonucleotide uFC obtained at 5 °C in a buffer containing 25 mM Na-HEPES, pH 7.0, 100 mM NaCl, 0.001% surfactant P20. Fluorescence of donor (recorded at 525 nm, \blacktriangle) and sensitized emission of acceptor (recorded at 560 nm, \blacksquare) was obtained by excitation at 493 nm. The fluorescence of directly excited acceptor (at 552 nm, \square) was recorded at 600 nm.

of the titration curve of En binding to the uFC (Figure 5.9), using a single site binding model (Eq. 4.5). The estimated equilibrium dissociation constant in this case was found to be 5 ± 3 nM. K_d values of specific Bcd and Ubx binding to DNA could be evaluated only from the titration curves acquired at DNA concentrations significantly lower than those used for the titration experiments, i.e. at 10-100 pM. However, at these DNA concentrations the formation of homeodomain-DNA complex could not be accurately measured due to the low fluorescence signal, and therefore binding curves were not obtained. Nevertheless, the titration data of Bcd and Ubx presented above (Figures 5.7 and 5.8) can be used for the determination of the stoichiometry of binding. In general, this could be determined from the break-point position on the binding curve, i.e. where saturation was achieved. In the titration curves of Bcd and Ubx (Figures 5.7 and 5.8) the maximum of acceptor fluorescence (or the breakpoint of the donor emission decay curve) corresponds to a 2:1 molar ratio which can be taken as a stoichiometric point. This, however, assumes that the decline in fluorescence intensity above this point is not related to the formation of new protein-DNA complexes. This conjecture is supported by several facts indicating that the decrease in fluorescence intensity is a consequence of aggregation of homeodomain-DNA molar complexes. Indeed, at protein/DNA ratios larger than 2:1 the fluorescence intensity of both donor and acceptor

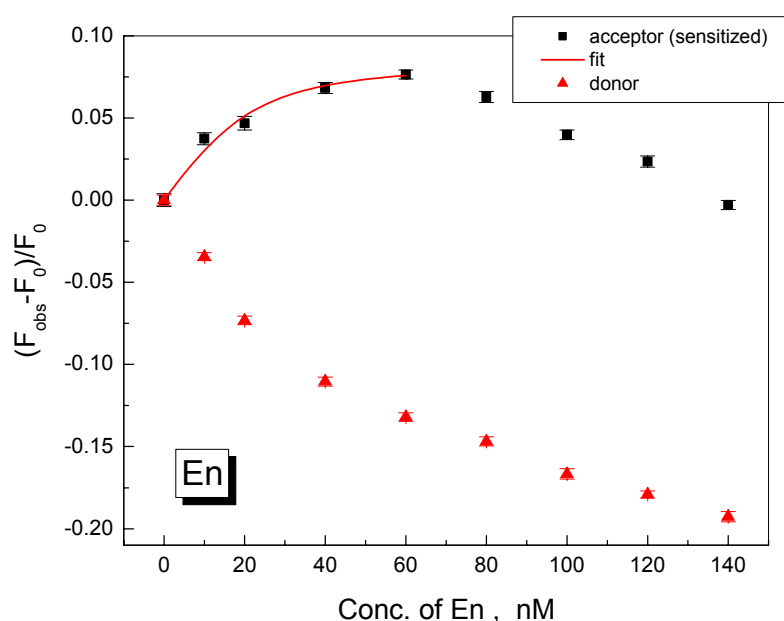


Figure 5.9. Binding curves of En to the specific oligonucleotide uFC obtained at 5 °C in a buffer containing 25 mM Na-HEPES, pH 7.0, 100 mM NaCl, 0.001% surfactant P20. Fluorescence of donor (recorded at 525 nm, ▲) and sensitized emission of acceptor (recorded at 560 nm, ■) was obtained by excitation at 493 nm.

starts to decrease gradually over time. Larger excess of protein resulted in higher rates of fluorescence intensity decay. Second, the results of analysis of FCS data of Bcd-bFC binding at molar 3:1 ratio (see Figure 5.13) show that the number of bFC molecules in solution decreases in time. The rate of this decrease was also proportional to the excess of protein over DNA. Third, the absorption spectra of the specific non-labeled oligonucleotides “b” and “u” decreased during several hours after the addition of 3-fold excess of homeodomains Bcd and Ubx (Figure 5.10); centrifugation of the sample led to a dramatical decrease in the optical density of the sample, indicating the formation of aggregates.

Binding of Bcd to the nonspecific oligonucleotide nsFC did not produce any significant changes in the fluorescence in the donor and acceptor with the exception of the decrease in the fluorescence intensity of both fluorophores at high protein concentrations, indicative of aggregation and microprecipitation of the Bcd-DNA complexes. In contrast, the

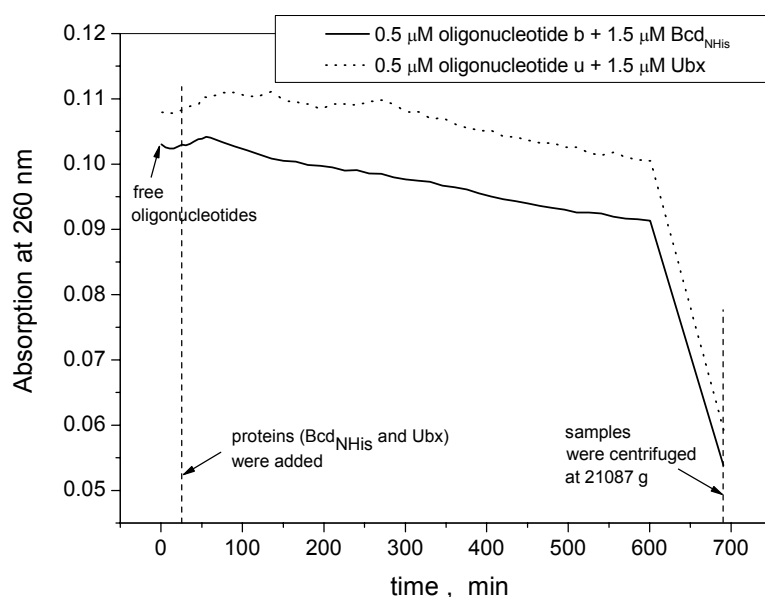


Figure 5.10. Decrease in absorption at 260 nm of the oligonucleotides “b” and “u” in the course of time after adding a 3-fold excess of the homeodomains Bcd or Ubx. Measurements were carried out at 5 °C in a solution containing 25 mM Na-HEPES, pH 7.0, 100 mM NaCl, 0.001% surfactant P20.

binding of Ubx to the nonspecific oligonucleotide nsFC led only to slight changes in the donor and acceptor fluorescence. However these changes were significantly smaller than to those obtained for the specific oligonucleotides, implying that nonspecific interactions of these proteins with DNA do not contribute to the alterations in the fluorescence signals detected for the specific oligonucleotides. It is worth noting that the length of the oligonucleotides used was limited to 15 bp. Taking into account that the length of the specific

recognition site is 6 bp (located in the center of the oligonucleotide), the probability of formation of nonspecific complexes is very low. These considerations justify my approach for the treatment of experimental data, in which the reaction of the nonspecific interactions was omitted.

5.3 Quantitative analysis of FRET between the fluorescein and Cy3 labels linked to the DNA

As it can be seen from the titration data, the efficiency of FRET between the fluorescein and the Cy3 labels attached at 3' and 5' ends of the specific oligonucleotides bFC and uFC increases when Bcd/Ubx homeodomains are added. The FRET efficiency and therefore the distance between the fluorescent labels can be estimated from the recorded emission and excitation spectra of the donor-acceptor pair. The determination of energy transfer (E) in this work is based on the approach described in the methods chapter 4.8. For the calculation of donor-acceptor distances of free oligonucleotides, the characteristic Förster distance $R_0 = 56 \text{ \AA}$ was used for fluorescein-Cy3 pair, assuming rapid orientational randomization of donor and acceptor dipoles orientations i.e. $\kappa^2 = 2/3$. The low measured anisotropy value of 0.125 of the fluorescein label covalently attached to the 5'-end of the DNA molecule via C₆-linker indicates a high rotational mobility. Therefore, the efficiency of energy transfer (E) is

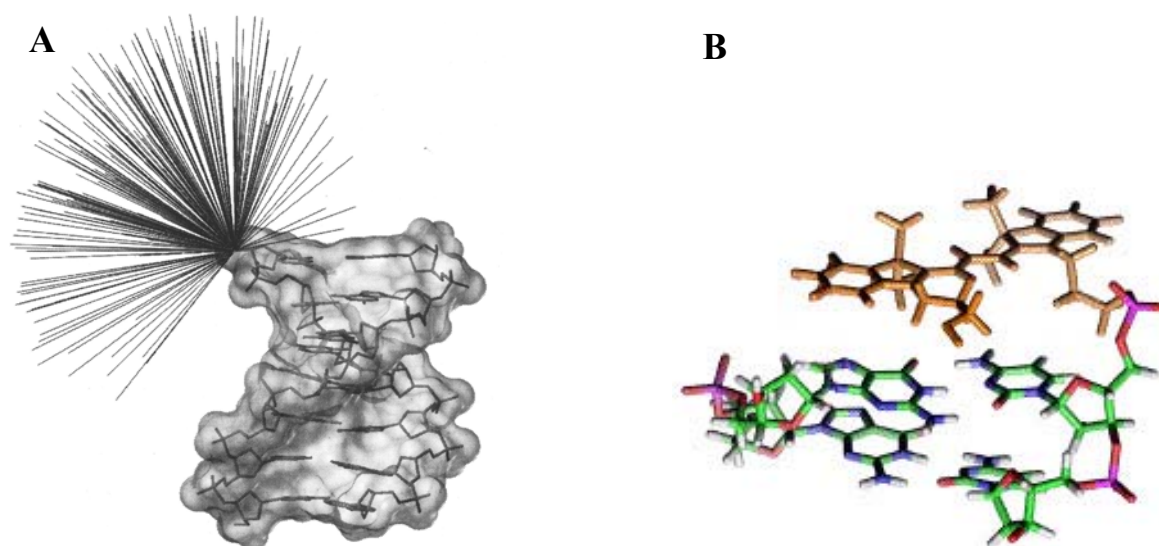


Figure 5.11. Panel A: Distribution of the 200 energetically most favourable conformers for 6-carboxy-fluorescein at the DNA helix end. A central atom in the 200 superimposed conformations is connected with atom C5' of the DNA through straight lines [Stühmeier et al. 2000]. Panel B: Conformation of Cy3 fluorophore and the two end base pairs of the DNA decamer [Norman et al. 2000].

only sensitive to the donor-acceptor distance, R . Thus donor-acceptor distances obtained for the free oligonucleotides bFC and uFC are in good agreement with the predicted value of $R \approx 65 \text{ \AA}$, using the molecular modeling [Hillisch 1998; Stühmeier et al. 2000] and NMR data [Norman et al. 2000], according to which fluorescein conformers point away from the DNA with an extended linker conformation (Figure 5.11 A) due to electrostatic repulsion between the two-fold negatively charged dye (two charges) and the polyanionic DNA. The calculated position of fluorescein corresponds to $\approx 10 \text{ \AA}$ from the 5' DNA termini in a virtual elongation of the DNA backbone [Hillisch 1998]. NMR data show that Cy3 molecule is stacked on the top of the DNA helix end and distanced from the last base pair by $\approx 5 \text{ \AA}$ [Norman et al. 2000] (Figure 5.11 B). This model is consistent with the anisotropy value of 0.3 measured for the acceptor (Cy3) which is covalently bound at the 3'-end of the DNA, indicating for limited mobility of Cy3. The estimation of energy transfer efficiency, E , and FRET distance, R , in homeodomain-DNA complexes is considerably more complicated due to presence of several molecular species. If one can assume that each species can be characterized by its energy transfer efficiency E_i , molar absorption coefficient ε_i and its emission spectrum shape function ϕ_i^A introduced in chapter 4.8, then the method of enhanced acceptor emission normalization (see chapter 4.8) can be applied by using the following expression:

$$\frac{F_{em}^A(\lambda, \lambda')}{F_{exc}^A(\lambda, \lambda'')} = \frac{\sum_i (\varepsilon_i^D(\lambda') d^+ E_i + \varepsilon_i^A(\lambda')) \phi_i^A(\lambda) C_i}{\sum_i \varepsilon_i^A(\lambda'') \phi_i^A(\lambda) C_i} \quad (5.1)$$

where C_i is the concentration of species i in solution. It is obvious that this expression cannot be used to calculate the energy transfer efficiency E_j for particular species j , unless all the parameters for all species, including all $E_{i \neq j}$ are known. However, when the molar absorption coefficients of donor and acceptor and the acceptor's quantum yield do not vary between the species, an averaged E value can be obtained from the Equation (5.1). Unfortunately, the titration profile of directly excited acceptor emission and measurements of acceptor absorption in homeodomain-DNA complexes showed that this assumption does not hold. Thus, energy transfer efficiency E can be evaluated if only single species is present at equilibrium after the homeodomain-DNA binding. At this point the information about the homeodomain-DNA binding reaction mechanism, obtained from the stopped-flow experiments, was combined with the data from appropriate titration curves in order to determine the conditions under which this requirement can be met. The results of analysis of specific Bcd binding kinetics to the oligonucleotide bFC show that the complex is formed at

the stoichiometric ratio of 2:1 (2Bcd-DNA), and that the second Bcd protein cooperatively binds almost irreversibly to the preformed Bcd-DNA complex (see chapter 5.5). Therefore, in less than 10 seconds after the mixing of two-fold excess of Bcd with oligonucleotide bFC only 2Bcd-DNA complexes are present in the reaction solution (Figure 5.16 A). This result enables one to apply the expression (Eq. 4.10) for the evaluation of the energy transfer efficiency, E , in the 2Bcd-DNA complex using acceptor emission and excitation spectra recorded at Bcd/DNA ratio of 2:1. Data of Ubx-DNA binding kinetics suggest that two Ubx molecules bind to the specific oligonucleotide uFC, forming a 2Ubx-DNA complex. However, in contrast to the case of Bcd the binding of first Ubx protein to the oligonucleotide is almost irreversible (the dissociation rate constant is very small), whereas the second protein binds with a considerably smaller affinity. This results in a linear response of the titration curve (Figure 5.15) in the region of the Ubx/DNA molar ratio 0/1-1/1 followed by the non-linear region, where free Ubx protein is in equilibrium with Ubx-DNA and 2Ubx-DNA complexes. According to this model the acceptor fluorescence emission and excitation spectra acquired at equimolar ratio of Ubx and DNA can be used for the calculations the energy transfer efficiency, E , in the Ubx-DNA complex, because under these conditions only the Ubx-DNA complex is present in solution at equilibrium (Figure 5.16 B) and the formalism (Eq. 4.10) can be applied.

A comparison of the energy transfer efficiency values obtained for 2Bcd-DNA and Ubx-DNA complexes with respective free oligonucleotides (bFC and uFC respectively) show that formation of 2Bcd-DNA complex induces an increase in E by 12 %, whereas the binding of Ubx to a specific oligonucleotide uFC leads to an increase of E by 18 % (Table 5.1).

Table 5.1. FRET efficiency and donor-acceptor distance values obtained for free oligonucleotides and oligonucleotide-homeodomain complexes.

Sample	FRET efficiency E	FRET distance R , Å
bFC	0.33 ± 0.007	63.0 ± 0.4
bFC+2Bcd	0.37 ± 0.006	61.1 ± 0.5
uFC	0.34 ± 0.006	62.6 ± 0.3
uFC+Ubx	0.40 ± 0.005	60.0 ± 0.3

Assuming that the changes in the fluorescence spectra of donor and absorption spectra of acceptor are negligible in course of binding and using R_0 equal to 56 Å, the reduction of the donor acceptor distance for 2Bcd-bFC complex corresponds to ~ 2 Å and that for Ubx-uFC complex ~ 2.5 Å (Table 5.1). These results suggest that the binding of Bcd and Ubx to their

specific sequences induces conformation changes (e.g. bend) of the DNA molecule leading to reduced donor-acceptor distances and therefore increased FRET efficiencies.

5.4 Analysis of the translational diffusion properties of the double-labeled oligonucleotide bFC and the Bcd-oligonucleotide complex using fluorescence correlation spectroscopy (FCS)

In order to obtain the information about the diffusion characteristics of the fluorescently labeled oligonucleotide bFC and analyze the slow processes occurring in course of Bcd binding to the DNA at the molar protein/DNA ratios higher than 2:1, FCS experiments were performed at the oligonucleotide concentration of 20 nM (see chapter 4.10).

A cross-correlation function of the fluorescein and Cy3 fluorescence signals (Figure 5.12) from the free DNA molecule in solution was calculated using a model of a single component, in which the presence of the triplet state was taken into account (see chapter 4.10, [Widengren et al. 1995]). Thus obtained value of the translational diffusion constant D_t of the bFC oligonucleotide is $1.5 \pm 0.2 \times 10^{-6} \text{ cm}^2/\text{s}$ (average of 20 measurements). This value is in agreement with the published values of the translational diffusion constants

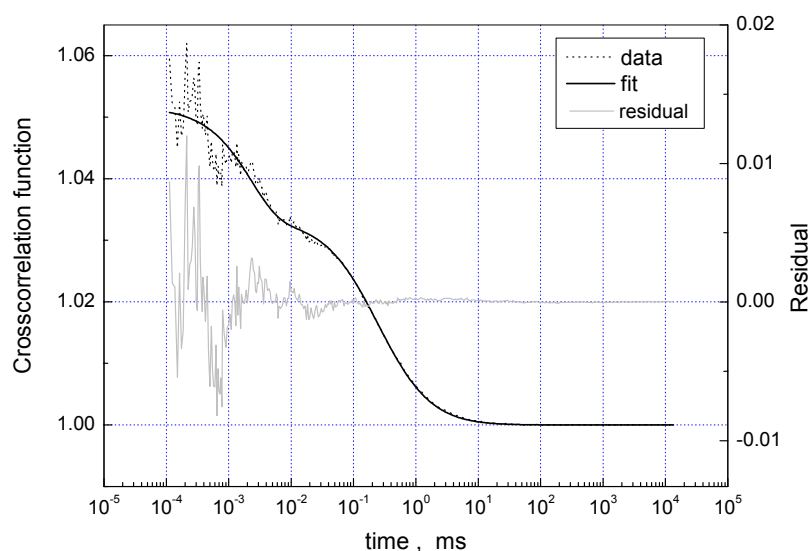


Figure 5.12. Example of the crosscorrelation function fit of the free oligonucleotide bFC using a single component model (including the triplet state). The data correspond to the D_t value of $1.4 \pm 0.2 \times 10^{-6} \text{ cm}^2/\text{s}$. Data were recorded at room temperature in the buffer containing 25 mM Na-HEPES, pH 7.0, 100 mM NaCl, 0.001% surfactant P20. Fluorescence was excited at 488 nm. Concentration of the oligonucleotide in solution was 20 nM.

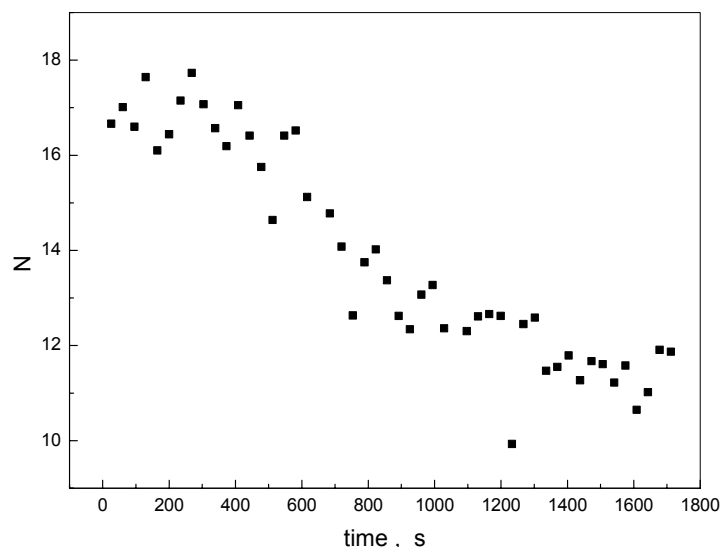


Figure 5.13. Change of the number of molecules in the detection volume, obtained from the analysis of the crosscorrelation curves recorded at each time point. Data were recorded after the adding of a 3-fold molar excess of Bcd to the solution containing 20 nM oligonucleotide bFC. Sample was prepared in the buffer containing 25 mM Na-HEPES, pH 7.0, 100 mM NaCl, 0.001% surfactant P20. Fluorescence was excited at 488 nm.

obtained experimentally and theoretically for oligonucleotides of comparable length (e.g. 14 bp) under similar conditions [Eimer and Pecora 1991; Fernandes et al. 2002].

In another FCS experiment a 3-fold excess of Bcd was added to the solution containing the oligonucleotide bFC and the crosscorrelation function was calculated every 30 s in a time course of ~30 min. The results of analysis performed on the obtained correlation curves show a ~35 % decrease in the number of molecules in the detection volume (~1.5 fL) over the entire experiment (Figure 5.13). These data correlate with the results of other spectroscopic experiments (see previous section) suggesting that the aggregation of the homeodomain-DNA complexes at the molar ratio higher than 2:1 is likely.

5.5 Investigation of the homeodomain-DNA interaction kinetics

In order to elucidate the mechanisms of homeodomain-DNA interaction kinetics and to obtain the information about the influence of Lys50 (specific to the K-50 family of homeodomains) and Arg54 (specific to Bcd) aminoacids on the affinity of Bcd homeodomain binding to a specific DNA, kinetic measurements of homeodomain binding to the specific

oligonucleotides were performed using the stopped-flow technique described in methods chapter 4.9). The progress of reaction was monitored by detecting the change in fluorescence intensity of the acceptor label (Cy3) on a DNA molecule. The binding of each homeodomain to the respective specific oligonucleotide (bFC and uFC) was measured at different homeodomain/oligonucleotide molar ratios ranging from 1/2 to 3/1 (an example is shown in the Figure 5.14). Global analyses were performed on the sets of the kinetic traces obtained for the binding of each individual homeodomain. Preliminary analysis showed, that the binding of Bcd and Ubx to the respective specific oligonucleotide may involve a kinetic step, with a very low dissociation rate constant which cannot be accurately evaluated using a stopped-flow technique. Therefore measurements of the homeodomain-oligonucleotide complex dissociation kinetics (an example is shown in the Figure 5.15) were carried out on a conventional spectrofluorometer using 200-fold excess on a non-labeled specific oligonucleotide (see chapter 4.6). The values obtained for the dissociation rate constant were used for further steps in global analysis of the datasets of the homeodomain-oligonucleotide association kinetics. For the analysis of kinetic data of the studied homeodomain proteins (Bcd_{NHIS} , Bcd_{CHIS} , $Bcd_{NHIS}K50A$, $Bcd_{NHIS}R54A$ and Ubx) a large number of sequential and/or parallel reaction models were considered including possible intermediate states of

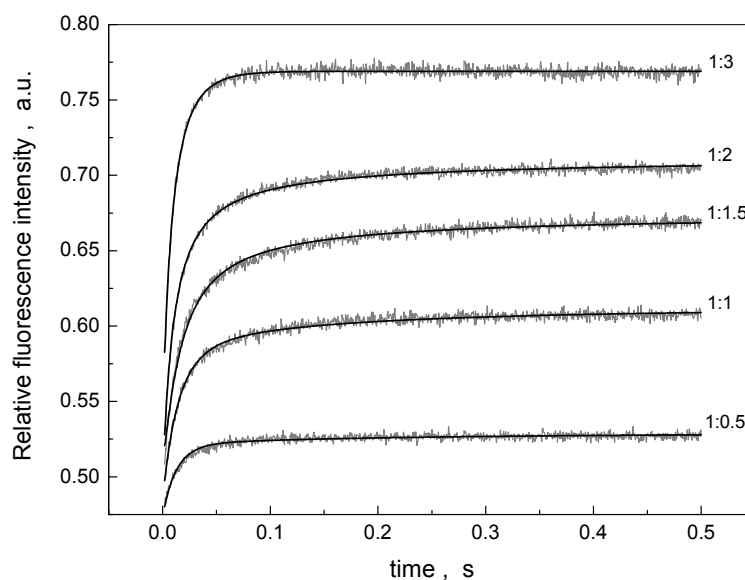


Figure 5.14. Global fit of kinetic profiles of Bcd_{NHIS} association to the oligonucleotide bFC recorded at different Bcd_{NHIS} /oligonucleotide molar ratios. Kinetic traces in gray represent fluorescence change of acceptor (Cy3), black solid curves represent fitting curves. Fluorescence was excited at 490 nm. Data were acquired at 5 °C in a buffer containing containing 25 mM Na-HEPES, pH 7.0, 100 mM NaCl, 0.001% surfactant P20. Concentration of the oligonucleotide bFC was 20 nM. Concentration of the Bcd_{NHIS} was varied from 10 nM to 60 nM.

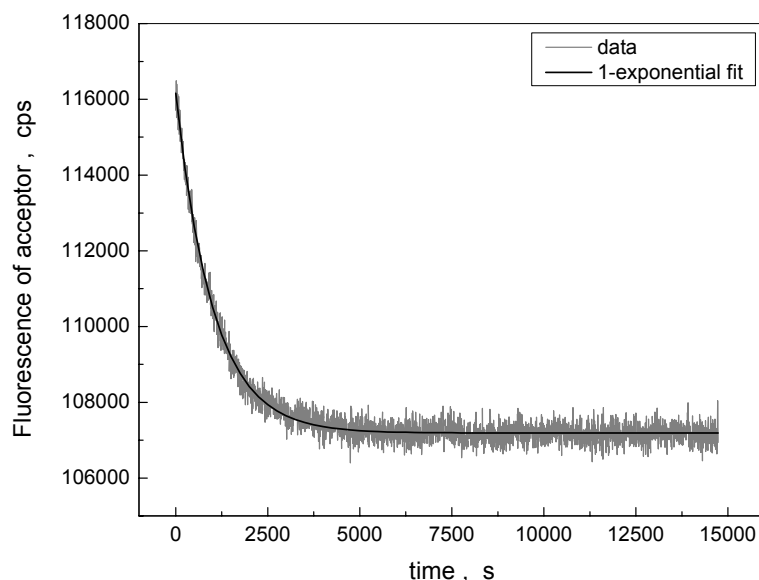


Figure 5.15. Dissociation kinetics of the 2Bcd_{NHis}-bFC complex. Signal trace was recorded at 5 °C after adding of a 200-fold excess of the non-labeled specific oligonucleotide “b” to the solution containing the 2Bcd_{NHis}-bFC complex preformed at 20 nM Bcd_{NHis} and 10 nM bFC. The sample was prepared in a buffer containing 25 mM Na-HEPES, pH 7.0, 100 mM NaCl, 0.001% surfactant P20.

homeodomain-DNA complex, homeodomain dimerisation and/or transition into an active state, binding of multiple homeodomain molecules to the DNA and complex-complex interaction. However, most of these models were excluded since at least one of the following criteria were not met during the fitting procedure:

- convergence of fitting parameters
- goodness of the global fit to multiple datasets recorded at different protein concentrations (i.e. residuals and sum of squares)
- correlation between the specific fluorescence amplitudes of molecular species present in reaction calculated during the fitting process (see appendix A3) and the experimental data obtained from the steady state fluorescence measurements.

The choice of the model, which met all these criteria, was also driven by the fact that the stoichiometry of the homeodomain-DNA complex was 2:1 as determined from the titration experiments (chapter 5.2). The best results for the homeodomain proteins analysed were obtained using the following reaction scheme:



Table 5.2. Results global analysis of kinetic data

protein+DNA	k_{1+} , $\times 10^9 \text{ M}^{-1}\cdot\text{s}^{-1}$	k_{1-} , s^{-1}	k_{2+} , $\times 10^9 \text{ M}^{-1}\cdot\text{s}^{-1}$	k_{2-} , s^{-1}
Bcd _{NHis} +bFC	3.9 ± 0.1	9.1 ± 0.8	3.2 ± 0.1	0.001 ± 0.0001
Bcd _{CHis} +bFC	2.0 ± 0.1	9 ± 0.9	1.0 ± 0.04	0.003 ± 0.0002
Bcd _{NHis} K50A+bFC	1.9 ± 0.2	5.9 ± 1.4	0.9 ± 0.06	0.004 ± 0.0003
Bcd _{NHis} R54A+bFC	2.2 ± 0.1	2.8 ± 0.8	0.9 ± 0.03	0.004 ± 0.0002
Ubx+uFC	2.7 ± 0.1	0.002 ± 0.0003	1.8 ± 0.2	15 ± 2

involving the sequential formation of a specific monomer complex (PD) between the homeodomain (P) and the oligonucleotide (D) and subsequent binding of a second molecule of the homeodomain to the preformed specific complex. The results of the kinetic data analysis are presented in the Table 5.2. The values of the bimolecular association rate constant, k_{1+} , of the studied homeodomains vary in the range $1.9 - 3.9 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$, suggesting that the binding process might be diffusion-controlled. A ~ 2 -fold decrease in the association rate constant, k_{1+} , of Bcd_{CHis} compared with that for Bcd_{NHis} shows that the C-terminal location of the His-tag in Bcd_{CHis} significantly affects the DNA association properties of this homeodomain. In contrast, the dissociation rate constant k_{1-} of the specific monomeric complex does not seem to be influenced by the position of the His-tag in the protein. Note that though the Bcd mutants K50A and R54A have the His tag at the N-terminus (same location as Bcd_{NHis}), their DNA binding abilities (k_{1+} , k_{2+}) do not differ from those of Bcd_{CHis}. Overall, a characteristic feature of the Bcd interaction kinetics with a short

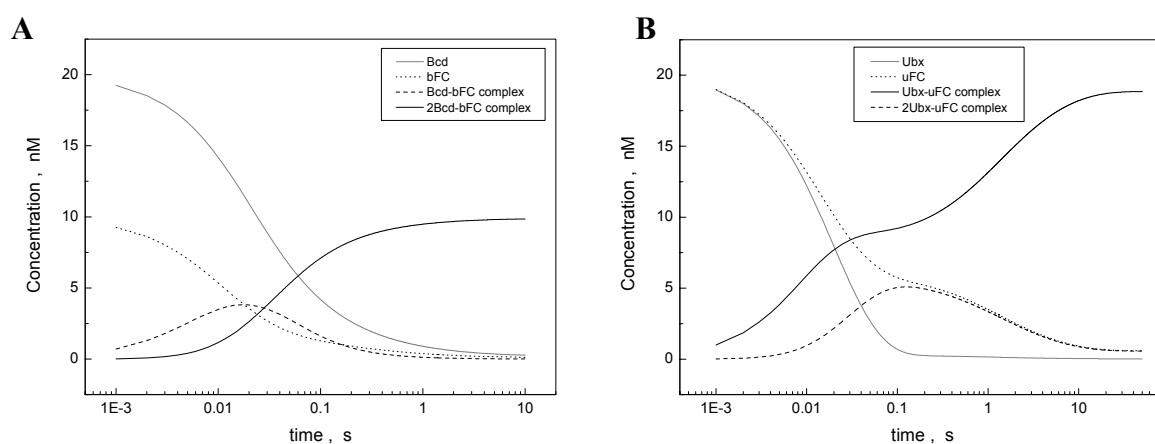


Figure 5.16. Concentration dynamics of species during the reaction of Bcd_{NHis} binding to the oligonucleotide bFC (panel A) and Ubx binding to the oligonucleotide uFC (panel B). Data were simulated using the values of the reaction rate constants obtained from the global analysis of the respective kinetic datasets. Total concentration of Bcd_{NHis} was equal to 40 nM (panel A). That of Ubx was equal to 20 nM (panel B). The concentration of oligonucleotide (bFC, uFC) was 20 nM in both cases.

specific DNA molecule (oligonucleotide bFC) is the formation of a relatively instable ($k_{1-} \approx 9 \text{ s}^{-1}$) specific Bcd-DNA complex at extremely high association rate ($k_{1+} \approx 4 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$) followed by the association step of a second Bcd molecule which increases the stability of the final complex by a factor of $\sim 10^4$. In contrast, Ubx seems to form a stable ($k_{1-} = 2 \pm 0.3 \times 10^{-3} \text{ s}^{-1}$) monomeric complex with the specific oligonucleotide uFC at slightly slower association rate constant k_{1+} relative to Bcd_{NHis}, succeeded by the association of the second Ubx molecule forming an instable ($k_1 = 15 \text{ s}^{-1}$) dimeric complex. These differences in the binding kinetics of the Bcd and Ubx homeodomains result in distinct distributions of species at equilibrium (Figure 5.16).

It is worth noting that for all homeodomains analysed the calculated specific fluorescence amplitude of the respective dimer complex (P₂D) was by 5 – 27 % higher than the one of the monomer complex (PD), which suggests a successive increase in FRET efficiency upon binding of the two homeodomain molecules to the DNA.

Effects of solution viscosity and ion concentration on the bimolecular association rate

In order to test whether the bimolecular association of the Bcd and Ubx to the respective specific oligonucleotide is controlled by diffusion the viscosity of the reaction solution was increased by adding 20 % of glycerol and the kinetics of homeodomain binding to the DNA was measured as described above. The obtained bimolecular association rate constants are shown in the Figure 5.17. The results show that the increase in the viscosity of solution from 1.53 to 2.88 cP significantly reduces the association rate constant for both homeodomains (down to $\approx 1 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$) providing a clear evidence for the diffusion control of the association rate.

In another set of experiments the dependence of the bimolecular association rate constant versus the salt concentration in solution was analyzed. In this case, the kinetics of Bcd binding to the oligonucleotide bFC was measured (as described above) at the NaCl concentrations of 100 mM, 200 mM and 300 mM. The obtained values of the association rate k_{1+} exhibit a linear dependence of $\log k_{1+}$ versus $\log[\text{NaCl}]$ (Figure 5.18). Since the number of positively charged residues in the Bcd molecule interacting with DNA is 10, the data are in a good agreement with a model (Equation (2.11)) proposed by [Lohman et al. 1978]. It is important to note that according to this model the association reaction between the protein

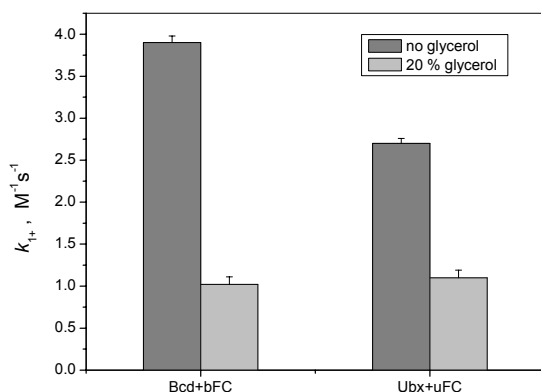


Figure 5.17. Effect of solution viscosity on the rate of bimolecular association of Bcd and Ubx homeodomains with the respective specific oligonucleotides

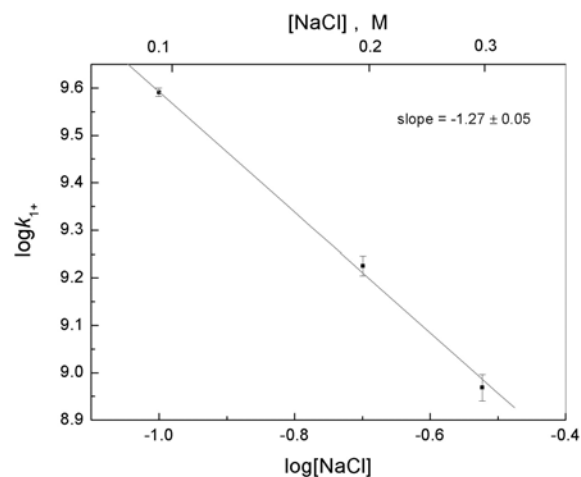


Figure 5.18. Effect of the salt concentration on the bimolecular association of Bcd to the specific oligonucleotide bFC.

and DNA occurs without formation of intermediates, and that the ion effects on the rate constant results from the screening of the charged macromolecules by low molecular weight ions.

Estimation of the diffusion-controlled reaction rate limits

In order to evaluate the limit of the Bcd homeodomain-DNA association rate controlled by diffusion, the hydrodynamic properties of the Bcd homeodomain and the oligonucleotide “b” were analyzed. First, a model of the Bcd_{NHis} tertiary structure was simulated on the basis of the primary sequence homology with other homeodomains, for which crystal or NMR structures exist, using the ProModII program on the Swiss-Model Server and molecular dynamics programs. A tertiary structure of the double-stranded 15 bp oligonucleotide “b” was generated using a molecular modeling program SYBYL. From the obtained tertiary structures of Bcd_{NHis} and oligonucleotide “b” the values of translational diffusion coefficient D_t and the hydration radius R_h were calculated using the programs of molecular hydrodynamics developed by Garcia de la Torre: HYDROPRO [Garcia de la Torre et al. 2000] and HYDRONMR [Garcia de la Torre et al. 2000]. The obtained values are represented in the Table 5.3. Further, the Smoluchowski equation (2.10) was applied to estimate the limit of the diffusion-controlled association reaction rate. The orientation coefficient κ was evaluated to be ≈ 0.3 by calculating a fraction of the surface of Bcd_{NHis} molecule potentially interacting

Table 5.3. Simulated hydrodynamic parameters of the macromolecules.
(Values were obtained for $T = 278$ K and $\eta_{\text{H}_2\text{O}} = 1.519$ cP)

Molecule	$D_t, \times 10^{-7} \text{ cm}^2/\text{s}$	$R_h, \text{ \AA}$
Bcd _{NHis} homeodomain	8.1	16.5
Oligonucleotide “b”	7.6	17.5

with DNA. For this the amino acids in the positions from 2 to 8 in the N-terminal arm and in the positions from 40 to 60 in the α -helix III were selected based on known structural information of homeodomain-DNA complexes (e.g. Engrailed/DNA: 3HDD, Ubx/DNA: 1B8I). The effective reaction radius was taken as a sum of hydration radiuses of the protein and DNA. It is interesting to note, that for a hypothetical case when the electrostatic interactions between the homeodomain and DNA are absent (i.e. $f = 1$) and the surface of the macromolecules is isoreactive ($\kappa = 1$), the estimation of diffusion-controlled association rate limit yields a value of $4.0 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$, which is very close to the experimentally obtained value. This implies that in a real situation the electrostatic interactions that contribute to the homeodomain-DNA binding possibly compensate the effect of orientational constraints, i.e. $f \cdot \kappa \approx 1$. These results along with the fact that Bcd homeodomain has a highly positively charged α -helix III and the N-terminal arm, that interact with DNA, suggest that these regions of the protein have a strong electrostatic potential acting as an energetic and orientational sink in such a way that virtually every collision between Bcd and DNA leads to the formation of specific complex.

6. Discussion

Conformational changes of the homeodomains during the DNA binding

It has been previously shown by [Nanda and Brand 2000; Subramaniam et al. 2001] that fluorescence of Trp48 in several homeodomains (including Bcd and Ubx) is highly quenched. The fluorescence (excited at 280 nm) of Bcd and Ubx homeodomains (Figures 5.1, 5.2) is largely dominated by the tyrosines (Tyr25 in Bcd homeodomain; Tyr8, Tyr11 and Tyr25 in Ubx homeodomain) as can be seen from the position of fluorescence intensity maxima at 313 nm and 305 nm (Figures 5.1 A, 5.2 A). The crystal structures of Ubx and En homeodomains in complex with DNA [Kissinger et al. 1990; Passner et al. 1999] (Figure 6.1) suggest that the observed decrease of fluorescence intensity (excited at 280 nm) upon homeodomain-DNA complex formation (Figures 5.1 A, 5.1 C, 5.2 A, 5.2 C) can be attributed to the quenching of Tyr25 fluorescence resulting from its interaction with the phosphate groups of the DNA backbone. Similar magnitudes of this quenching observed for the specific and non-specific DNA support this model. The increase in fluorescence intensity of Trp48 upon binding of Bcd and Ubx homeodomains to DNA reflects the rearrangements in the protein structure in the close vicinity to this residue. According to the model of Trp48 fluorescence quenching in homeodomains proposed by [Nanda and Brand 2000] and [Subramaniam et al. 2001], the increase in fluorescence of Trp48 in Bcd and Ubx homeodo-

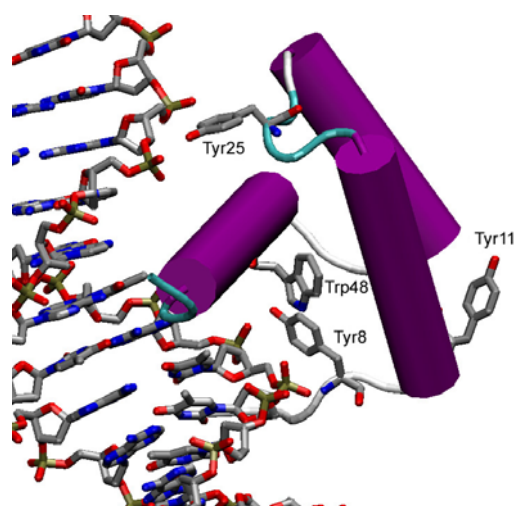


Figure 6.1 Position of the aromatic residues in Ubx homeodomain-DNA complex [Passner 1999].

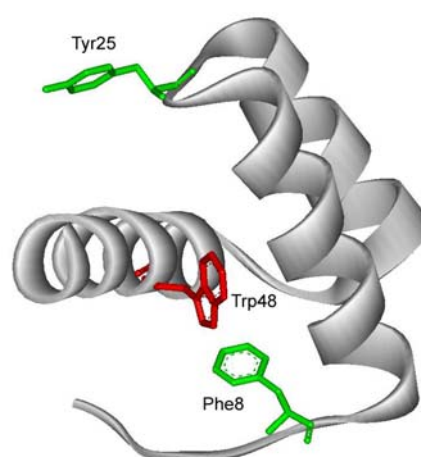


Figure 6.2 Orientation of the aromatic residues in the simulated structure of Bcd homeodomain.

mains due to binding to DNA may indicate an increase of the distance between the indole NH group of Trp48 and the π electron system of benzene of the aromatic amino acid at the position 8 in the homeodomain amino acid sequence (Figure 6.2). This possibility is plausible because the N-terminal arm (containing Tyr8) of homeodomains is known to dock into the minor groove of the DNA molecule [Pabo and Sauer 1992], [Gehring et al. 1994], which might require some rearrangements in the tertiary structure of homeodomain leading to an increase of the Tyr8-Trp48 distance. In contrast to the fluorescence changes of Tyr25, the alterations of Trp48 fluorescence arising upon homeodomain-DNA complex formation seem to be sequence-specific. This could be attributed to possibly distinct conformations of the N-terminal arm of the homeodomain in the free state and in complex with DNA, where it makes specific contacts with nucleotides in the minor groove of DNA [Pabo and Sauer 1992], [Gehring et al. 1994]. The blue shift of the Trp48 fluorescence spectrum observed in the case of Bcd homeodomain binding to DNA (Figures 5.1 B, 5.1 D) suggests that the Trp48 residue in Bcd homeodomain bound to DNA is exposed to a more hydrophobic environment than in the unbound state.

A large number of DNA-binding proteins change their secondary structure upon DNA-binding [see chapter 2.1.4]. In many cases these changes are related to an increase in size of α -helical structures interacting with DNA. These structural changes have been observed, for example, upon DNA binding of λ repressor [Clarke et al. 1991], of Fos Jun transcription factor [Patel et al. 1990] and of NK-2 [Tsao et al. 1994], MAT α 2 [Carra and Privalov 1997] and Mesx-1 [de la Mata et al. 2002] homeodomains. Ultraviolet circular dichroism spectroscopy has been applied in this work to analyse possible changes in the secondary structure of Bcd, Ubx and En homeodomains during binding to the specific and nonspecific DNA sequences. The obtained results (Figure 5.3) suggest that similarly to MAT α 2 and NK-2 homeodomains binding of Bcd and Ubx homeodomains to the specific DNA sequences induces an extension of the α -helical structures by claiming the parts of unstructured regions of these homeodomains. However from the obtained spectra of circular dichroism is not possible to assess in which regions of the protein the additional α -helical structures do form during the binding. In contrast to the results obtained by [Carra and Privalov 1997] for the MAT α 2 homeodomain, binding of Bcd and Ubx homeodomains to the nonspecific DNA sequence indicate significant changes in their α -helical structures, which are different from those obtained with the specific DNA. Comparison of the CD spectra obtained for binding of Bcd to the specific and nonspecific oligonucleotides (Figure 5.3 A)

suggests that the nonspecific interactions play an important role in the induction of the changes in the secondary structure of this homeodomain upon binding to DNA. Specific contacts between Bcd homeodomain and DNA seem to have a modulation effect on the secondary structures induced by the nonspecific interactions. The dramatic decrease of the α -helical structure in the case of the Ubx homeodomain binding to the nonspecific DNA (Figure 5.3. B) could be attributed to an aggregation of the protein-DNA complexes possibly occurring at the concentrations used during the experiment.

DNA conformation changes upon homeodomain binding

The consensus recognition sequence⁷ of homeodomains contains a TAAT motif. As has been mentioned in chapter 2.1.4, the pyrimidine-purine step TpA is intrinsically less stable than all other dinucleotides and therefore better suited for the initialization of bending and/or untwisting of DNA molecule [Drew et al. 1985; Kim et al. 1993]. In this connection Nelson and coworkers [Nelson and Laughon 1990] proposed that the TAAT motif could be required for possible bending of DNA by the homeodomains, perhaps by facilitating some protein-DNA contacts. A few years later, the crystal structures of the Paired (prd) homeodomain homodimer-DNA complex and MAT α 1/MAT α 2 heterodimer-DNA complex emerged showing a significant bending of the DNA molecule [Li et al. 1995; Wilson et al. 1995]. It has been shown recently by means of molecular dynamic simulations that binding of the even-skipped (eve) homeodomain to a specific DNA sequence can induce a notable bending of the DNA backbone [Flader et al. 2003].

Using a sensitive spectroscopic technique based on intramolecular FRET occurring between the fluorophores on the DNA molecule, it was possible in this work to detect the changes in DNA conformation upon binding of Bcd, Ubx and En homeodomains. The reduction of the interfluorophore distance upon binding of Bcd and Ubx homeodomains calculated from the FRET efficiency increase corresponds to an apparent bending of DNA axis by $\sim 30^\circ$ for both proteins. It is important to note that these results refer to a Bcd homeodomain homodimer-DNA complex and Ubx homeodomain monomer-DNA complex. The angle of putative DNA bending in the complex with Bcd homeodomain homodimer is comparable to the value of $\sim 20^\circ$ found in the crystal structure for the prd homeodomain

⁷ Consensus sequence – a sequence common for members of a certain group or family (e.g. binding/recognition sequence of DNA common for a group of DNA-binding proteins)

homodimer-DNA complex [Wilson et al. 1995]. However, this similarity should be taken with care, because the alignment of the homeodomain molecules in the homodimer-DNA complex and therefore possible mechanisms of DNA bending may differ substantially. In the prd homeodomain homodimer-DNA complex the prd homeodomain molecules are bound to the two neighbouring recognition sites ensuring binding cooperativity [Wilson et al. 1993], which possibly facilitates or even induces the bending of DNA molecule. In contrast, the Bcd homeodomain homodimer complex is formed on a DNA molecule containing only one recognition site such that the position and relative orientation at least of one of the two homeodomain molecules cannot be predicted. The available crystal structures of homeodomain-DNA complexes showing the DNA bending [Li et al. 1995; Wilson et al. 1995] cannot provide a sufficient insight into the abilities of a single homeodomain molecule to bend a specific DNA sequence, since the DNA bending found in the crystal structures could be an effect of a cooperative binding and/or protein-protein interactions. In this context it is worth noting, that the results of quantitative analysis of DNA conformation change upon Ubx homeodomain binding, obtained in this work, provide an important experimental evidence (to my knowledge for the first time) of the ability of a single molecule of this homeodomain to induce significant alterations in the structure of a DNA molecule.

Intramolecular FRET on a DNA is a sensitive spectroscopic method, which can be effectively applied to study the homeodomain-DNA interaction kinetics

Studies of HD-DNA interactions have been performed so far only under equilibrium conditions, typically using such classical techniques as the electrophoretic mobility retardation assay. The application of the transient kinetic methods, which can provide a deeper insight into the kinetic mechanisms of interaction, was complicated because of number of problems related to reliable detectability of the spectroscopic signal change required by these methods. Although the changes in the intrinsic fluorescence or CD of homeodomains induced upon binding to DNA can be detected during steady-state measurements, the concentration of protein required for detection ($\sim 10 \mu\text{M}$) leads to a very fast reaction rate, which impedes its detection by stopped-flow technique.

External fluorophores residing on the protein or DNA have been tested in this work for their applicability for measurements of homeodomain-DNA interaction kinetics. The results of the fluorescent labeling of Bcd_{NHis}-Cys show that the presence of fluorescent labels

residing on the protein dramatically affects its solubility and promotes the aggregation of the labeled protein. Due to this fact the application of such labeled protein for kinetic studies is not possible. Fluorescent labels residing on the DNA helped to circumvent this problem; however the fluorescence intensity changes of the labels, induced upon binding of homeodomain to the DNA molecule, were too small for reliable kinetic measurements.

The detection of intramolecular FRET between two fluorophores on the DNA was found to be the most sensitive spectroscopic method for monitoring the homeodomain-DNA binding in both steady state and kinetic measurements. The fluorescence intensity changes of the fluorophores, recorded using this technique during the homeodomain-DNA complex formation, were considerably higher than those measured with single-labeled oligonucleotides. Consequently, a significant improvement in the signal/noise ratio of the stopped-flow kinetic traces was observed using this method. However it is important to note that this technique has some limitations. First, a change in the 5'-3' distance of the DNA molecule is required, therefore only DNA bending proteins can be studied using this method. A second limitation is the distance between the donor and acceptor fluorophores required for FRET. Its maximum value is ~ 100 Å, which corresponds to a DNA molecule of ~ 30 bases.

Kinetic mechanisms of homeodomain-DNA interaction

The results obtained from the fluorescence titration experiments (chapter 5.2) indicate that the Bcd and Ubx homeodomains bind to DNA with a 2:1 stoichiometry. It is known that many eukaryotic and some prokaryotic transcription factors increase their DNA-binding specificities dramatically by forming dimers [Ptashne 1986; Frankel and Kim 1991]. This also appears to be the case for at least some homeodomain proteins. As has already been mentioned, the *Drosophila* prd homeodomain binds to a dimeric site on DNA forming a homodimer [Wilson et al. 1993]. Homeodomain proteins are also known to form heterodimers on the DNA, e.g. MAT α 1/MAT α 2 from yeast [Mak and Johnson 1993; Phillips et al. 1994], Extradenticle/Hom C proteins [Wilson and Desplan 1995] from *Drosophila* or Pbx/Hox from human [van Dijk et al. 1995]. However it is important to note that all these complexes are formed on DNA molecules containing two homeodomain recognition sites (identical in the case of homodimers and different in the case of heterodimers) that are arranged either in direct ($\rightarrow\rightarrow$) or inverse ($\rightarrow\leftarrow$) orientations. In contrast, the results

presented in this work show that Bcd and Ubx homeodomains can form homodimers on relatively short oligonucleotides containing only a single recognition site.

The kinetic data (Table 5.2) of the Bcd and Ubx homeodomains binding to DNA are consistent with a reaction mechanism involving the formation of a specific monomer complex between the homeodomain and the oligonucleotide and subsequent binding of a second homeodomain molecule to the preformed specific complex (Figure 6.3). Since the

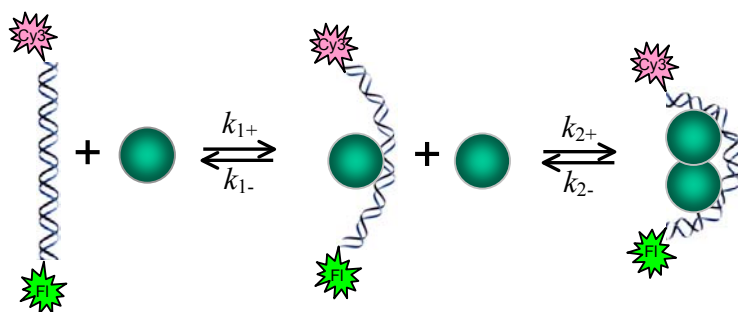


Figure 6.3. Schematic model of the homeodomain-DNA binding reaction. Dark spheres represent the homeodomain.

oligonucleotides used in the experiments contain only one specific recognition site, the interaction between the second homeodomain molecule and DNA is likely to be nonspecific. However the comparison of the dissociation rate constants obtained for Bcd shows that the binding of the second homeodomain molecule to a relatively unstable monomeric Bcd homeodomain-DNA complex dramatically increased (by a factor of $\sim 10^4$) the stability of the final complex. These data are in accord with the results obtained by [Burz et al. 1998] using biochemical techniques, that suggest a pairwise cooperativity of Bcd homeodomain binding to a strong (containing a high-affinity Bcd homeodomain recognition sequence TCTAATCCC) and weak (non-Bcd-specific binding sequence TCTAATTCC) binding sites located close to each other. A similar cooperative effect has been shown to occur between DNA-bound dimers of cI repressor from bacteriophage λ and HK022 [Ackers et al. 1982; Carlson and Little 1993]. DNA binding cooperativity is also observed between the homeodomains in the homodimer and heterodimer complexes mentioned above. In most examples of cooperative DNA binding by homeodomain-containing proteins, regions of the protein outside the homeodomain are important for protein-protein interactions. However, it has been shown [Burz and Hanes 2001] that the amino acids which are critical for the cooperative binding of Bcd, are within the homeodomain. In this context it is interesting that contrary to Bcd, the Ubx homeodomain does not show any detectable cooperativity upon

binding to DNA. This could be attributed to the fact that the Ubx homeodomain protein used in this work did not contain the flanking regions that are necessary for interaction with Extradenticle (Exd) and perhaps for cooperative binding of the Ubx homeodomain to the DNA [Beachy et al. 1993].

The very long lifetime of the dimeric Bcd homeodomain-DNA complex, compared with the duration of the processes during the development cycles, suggests that *in vivo* the Bcd protein may require external factors in order to dissociate from such dimeric (or possibly multimeric) complexes on the DNA. These factors could be the interaction with other proteins or a structural modification (e.g. phosphorylation) affecting either the cooperative interactions between the homeodomains or their interaction with DNA. It has been shown [Ronchi et al. 1993; Janody et al. 2000] that the activity of Bcd at the anterior pole of the embryo is indeed repressed by phosphorylation.

A comparison of the reaction rate constants obtained for the Bcd homeodomain proteins Bcd_{CHis} and Bcd_{NHis} shows that the position of His-tag at C-terminus in Bcd_{CHis} substantially impacts the characteristics of Bcd homeodomain association to DNA, whereas the stability of the monomeric Bcd homeodomain-DNA complex (represented by the dissociation rate constant k_{-1}) remains unaffected. These results suggest that a close proximity of the His-tag to the recognition helix of the homeodomain does not impair the interaction of the protein with DNA, although it has been observed that the α -helical structure of the helix III is largely destabilized by the His-tag at C-terminus [Wütrich K., personal communication], which may influence the efficiency of association with DNA.

The significant decrease of the association rate constants of Bcd homeodomain mutants K50A and R54A compared with those obtained for Bcd_{NHis} suggests that these two positively charged amino acids (Lys50 and Arg54) that face to DNA, may make an important contribution to long range electrostatic interactions between the homeodomain and DNA, thereby increasing the efficiency of association by directing the molecules into a proper orientations.

Diffusion-control of homeodomain-DNA association

The values obtained ($1.9 - 3.9 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$) of the bimolecular association rate constants of the studied Bcd and Ubx homeodomains suggest that the homeodomain-DNA association process may be diffusion-controlled. This assumption was confirmed by the results of kinetic

experiments, which showed that an increase of solution viscosity results in a substantial decrease of the association rate constant for both Bcd and Ubx homeodomains, providing (for the first time to my knowledge) a direct evidence of diffusion-controlled homeodomain-DNA association. Furthermore, the diffusion-controlled reaction rate limit estimated for Bcd homeodomain association to a 15 bp oligonucleotide neglecting the electrostatic corrections is smaller by a factor of ~ 3 than the experimentally obtained value.

Up to date only a few DNA-binding proteins are known that associate with DNA at the reaction rates equal or higher than diffusion-controlled [Riggs et al. 1970; Shimamoto and Utiyama 1983; Berger et al. 1998; Patrick and Turchi 2001; Dhavan et al. 2002]. The most prominent example is the *lac* repressor-operator interaction which has been an object of numerous studies [Riggs et al. 1970; Barkley et al. 1981; Winter et al. 1981; Ha 1990; Lewis et al. 1996]. As it was already mentioned in the chapter 2.1.5, the ~ 500 -fold difference found between the estimated diffusion-controlled limit and the measured rate of the *lac* repressor association to the operator sequence within a long DNA molecule was explained theoretically [Richter and Eigen 1974], assuming nonspecific DNA-binding of the repressor and the reduction of diffusion dimensionality by protein sliding along DNA. In contrast to this situation the extremely high rates of Bcd homeodomain association to DNA were observed using the oligonucleotides only 15 bp long. In this case the effects of the nonspecific binding and sliding are negligibly small; therefore the observed discrepancy between the estimated diffusion-controlled limit and the measured rate of Bcd homeodomain-DNA association may be attributed to the long-range electrostatic interactions between the positively charged regions of the homeodomain molecule and negative charges on the DNA. This assumption is supported by the dependence of bimolecular association rate constant on salt concentration. These data are in a good agreement with the model for the electrolyte screening [Lohman et al. 1978], which implies that the rate-limiting step for association is dictated by the long-range electrostatic interactions between the charged regions of the molecules shielded by the ions of electrolyte. Moreover, the analysis of the electrostatic field about the Bcd homeodomain molecule (Figure 6.4) shows that the strongest electrostatic potential is distributed around the α -helix III (the recognition helix which docks into the major groove of DNA). All these data suggest that the strong electrostatic field of this region may act as an energetic and orientational sink ensuring that virtually every collision between Bcd homeodomain and DNA leads to a successful formation of specific complex. This observation correlates with the results of theoretical analysis (based on application of

Poisson-Boltzmann equation [Davis and McCammon 1990; Sharp and Honig 1990]) of electrostatic interactions between the En homeodomain and DNA molecules [Fogolari et al. 1997], which indicate that the electrostatic field of the homeodomain can play a significant role in driving this molecule

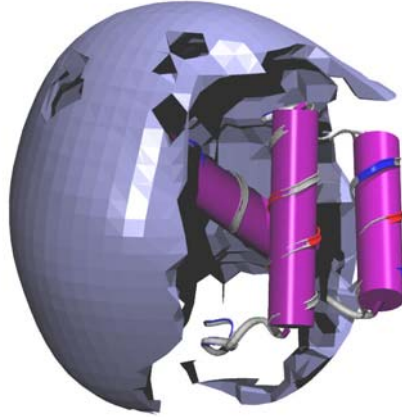


Figure 6.4. Isosurface of the electrostatic field (equal 5 V/m) of Bcd homeodomain molecule.

to a correct orientation relative to DNA. A similar mechanism leading to an enhancement of the association rate could be involved in the binding of the ribonuclease barnase to its inhibitor barstar, where the nonspecific electrostatic interactions between the complementary charged areas of the molecules have been reported to play a crucial role in the diffusion controlled formation of the complex [Schreiber and Fersht 1996].

7. Summary

Spectroscopic and transient-kinetic methods have been applied to study homeodomain-DNA interactions. The analysis of intrinsic fluorescence and circular dichroism spectra of Bcd, Ubx and En homeodomains has shown that the binding to DNA induces changes in their structure (increase of α -helical content and change in conformation of N-terminal arm), which are different for interaction with specific and non-specific sequences.

Significant changes in conformation of oligonucleotides containing a specific recognition site of Bcd or Ubx homeodomains were observed upon association with these homeodomains using the technique of intramolecular FRET between two fluorescent labels on the DNA. The results imply a bending of DNA axis by $\sim 30^\circ$ in a Bcd homeodomain homodimer-DNA complex and Ubx homeodomain monomer-DNA complex.

The detection of intramolecular FRET between the fluorophores on the DNA was found to be the most sensitive spectroscopic method for monitoring the homeodomain-DNA binding in both steady state and kinetic measurements. The fluorescence titration results obtained using this method indicate that Bcd and Ubx homeodomains bind to the specific oligonucleotides with a 2:1 stoichiometry. At protein concentrations exceeding this ratio, both complexes exhibited aggregation. The effect was independent of whether or not DNA was labeled with fluorescent dyes.

The kinetic data of the Bcd and Ubx homeodomains specific binding to DNA are consistent with a two-step reaction mechanism. In the first step a specific complex between one protein molecule and the oligonucleotide is formed. In the next step the second protein molecule binds the specific complex interacting with both the first bound protein molecule and DNA. The binding affinities in these two steps are very different for Bcd and Ubx homeodomains. For the Bcd binding, the complex formed at the first step is relatively unstable ($k_{\text{dis}} \approx 9 \text{ s}^{-1}$), however binding of the second protein molecule increases the stability of the final complex by a factor of $\sim 10^4$. In contrast, at the first step Ubx forms very stable complex ($k_{\text{dis}} = 2 \pm 0.3 \times 10^{-3} \text{ s}^{-1}$), whereas the subsequent association of the second Ubx molecule leads to the destabilization of the final complex ($k_{\text{dis}} = 15 \text{ s}^{-1}$).

It has been shown that the process of Bcd and Ubx homeodomain association with DNA is diffusion-controlled. The analysis of the electrostatic field distribution around the Bcd homeodomain molecule and the comparison of the estimated limit of Bcd homeodomain-DNA association rate with the experimentally obtained value suggests that the strong

electrostatic field of the α -helix III plays a significant role driving this molecule to a correct orientation relative to DNA thereby enhancing the association efficiency.

Appendix

A1. Function of Ultrabithorax and Engrailed homeodomain proteins

Ultrabithorax

Ultrabithorax (Ubx) homeodomain is a DNA-binding domain of the respective protein encoded by a homeotic gene within the lowest tier of the genetic control hierarchy that directs early *Drosophila* development. Ubx operates within the segmented framework established in the embryo by earlier-acting maternal effect and segmentation genes: its function is to specify the unique features of parasegments 5 and 6, which together constitute a contiguous region including the posterior thorax and a portion of the first abdominal segment [Beachy 1990]. This homeoprotein binds its target sequences cooperatively with another homeoprotein Extradenticle.

Engrailed

Engrailed (En) protein is a product of a segment polarity gene, a class of genes required for proper intrasegmental patterning during embryogenesis [Pankratz and Jäckle 1993]. This protein interacts cooperatively with the homeoprotein cofactor Extradenticle to bind DNA target sequences [Peltenburg and Murre 1996].

A2. Calculation of the kinetic fluorescence intensity profiles

The fluorescence intensity (in number of photons) of the sample during the reaction can be expressed as:

$$F_S(t) = I_{\text{exc}}(t)S(t) \quad (\text{A2.1})$$

where I_{exc} is the excitation light intensity (in number of photons), $S(t)$ is a time dependent fluorescence factor containing the information about the spectroscopic properties (extinction coefficient, quantum yield) and the concentration dynamics of the fluorescent species during the reaction. Excitation light leakage through the long-pass filter in the fluorescence detection path is proportional to the excitation light intensity:

$$F_L(t) = LI_{\text{exc}}(t) \quad (\text{A2.2})$$

where L is a constant. The measured fluorescence signal (in volts) is then:

$$F_m(t) = \alpha[S(t) + L]I_{exc}(t) + F_{offs} \quad (A2.3)$$

where α is a conversion constant and F_{offs} is an offset factor used to optimally adjust its change to the digitization scale of the ADC card. Any fluctuations of the excitation light intensity $I_{exc} = I_0 + \delta I(t)$, where I_0 is an approximate averaged light intensity value and δI is a momentary deviation from I_0 , will be reflected in the measured fluorescence intensity F_m . In order to correct the F_m for these fluctuations the excitation light intensity is measured:

$$X_m(t) = \beta I_{exc}(t) + X_{offs} \quad (A2.4)$$

β is a conversion constant; X_{offs} has the same purpose for the excitation light intensity measurements using absorption PMT as F_{offs} does for the fluorescence intensity measurements using the fluorescence PMT.

The correction of the measured fluorescence intensity F_m for the fluctuations and leakage of the excitation light using the equations (A2.3) and (A2.4) gives the following formula for $S(t)$:

$$S(t) = \frac{\alpha}{\beta} \cdot \frac{F_m(t) - F_{offs}}{X_{exc}(t) - X_{offs}} - L \quad (A2.5)$$

The factor L can be determined from a separate experiment without a fluorescent sample (i.e. $S(t) = 0$) using a rearrangement of the same equation:

$$L = \frac{\alpha}{\beta} \cdot \frac{F_m(t) - F_{offs}}{X_m - X_{offs}} \quad (A2.6)$$

A3. Numerical methods of kinetic data analysis

The measured fluorescence intensity of the sample at any time t during a reaction is

$$s(t) = \sum_{i=1}^m a_i C_i(t) \quad (A3.1)$$

where m is the number of species, a_i is the specific fluorescence amplitude which is proportional to the extinction coefficient and quantum yield of the i -th species in the reaction,

and C_i is the concentration of the i -th species. The expression for the entire kinetic trace containing n data points can be written as:

$$\begin{pmatrix} s(t_1) \\ s(t_2) \\ \vdots \\ s(t_n) \end{pmatrix} = \begin{pmatrix} C_1(t_1) & C_2(t_1) & \cdots & C_m(t_1) \\ C_1(t_2) & C_2(t_2) & \cdots & C_m(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ C_1(t_n) & C_2(t_n) & \cdots & C_m(t_n) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} \quad (\text{A3.2})$$

or in more compact form

$$\mathbf{s} = \mathbf{C}\mathbf{a} \quad (\text{A3.3})$$

During the fitting process a signal curve \mathbf{s}_{calc} is calculated from the Equation (A3.3) and fitted to the measured kinetic trace \mathbf{s} . The flow of the fitting program is represented in the Figure A3.1. First, the reaction mechanism is provided in form of differential equations containing the reaction rate constants as parameters and initial estimates are supplied. Then, the concentration profiles of species (represented as the columns of the matrix \mathbf{C}) are calculated using the Newton-Raphson and numerical integration methods [Press et al. 1992]. Next, the vector \mathbf{a} containing the specific fluorescence amplitudes of species is calculated from the Equation (A3.3) using linear regression and singular value decomposition (SVD)

methods [Press et al. 1992]. Subsequently the \mathbf{s}_{ca} , $\mathbf{R} = \mathbf{s} - \mathbf{s}_{\text{ca}}$ and $ssq = \sum_{i=1}^n R_i^2$ are calculated. If

the sum of squares, ssq , varies between two consecutive iterations by more than the value of predefined tolerance parameter, then the rate constants are varied using the non-linear regression (Marquardt-Levenberg) method [Press et al. 1992] and the cycle of \mathbf{C} , \mathbf{s}_{ca} , \mathbf{R} and ssq calculation is repeated till the change in ssq is smaller than the tolerance limit. At this point the fitting program generates the output including rate constants, specific fluorescence amplitudes, concentration profiles and the calculated kinetic signal trace, and then finishes. In the case of global analysis of multiple datasets the fitting procedure is the same as described above for a single curve with only exception that the vectors \mathbf{s}_i and matrices \mathbf{C}_i of k separate datasets are concatenated into one dataset during the fitting routines.

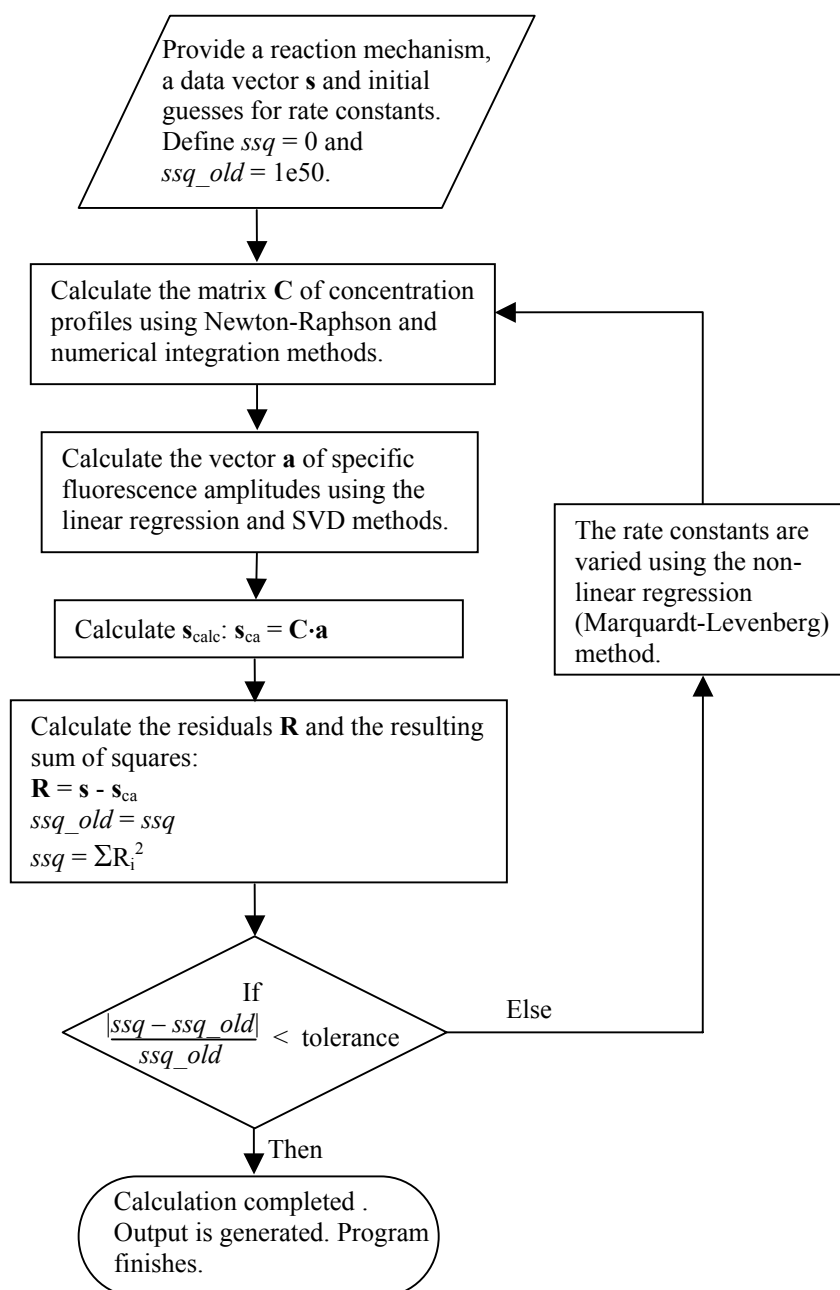


Figure A3.1. A schematic representation of program flow during fitting.

References

- Ackers G. K., Johnson A. D. and Shea M. A. (1982). "Quantitative model for gene regulation by lambda phage repressor." *Proc Natl Acad Sci U S A* **79**(4): 1129-33.
- Adam G. and Delbrück M. (1968). Reduction of Dimensionality in Biological Diffusion Processes. *Structural Chemistry and Molecular Biology*. D. N. Rich A. San Francisco, Freeman: 198-215.
- Ades S. E. and Sauer R. T. (1994). "Differential DNA-binding specificity of the engrailed homeodomain: the role of residue 50." *Biochemistry* **33**(31): 9187-9194.
- Affolter M., Percival-Smith A., Muller M., Leupin W. and Gehring W. J. (1990). "DNA binding properties of the purified Antennapedia homeodomain." *Proc. Natl. Acad. Sci. U S A* **87**(11): 4093-4097.
- Aggarwal A. K., Rodgers D. W., Drottar M., Ptashne M. and Harrison S. C. (1988). "Recognition of a DNA operator by the repressor of phage 434: a view at high resolution." *Science* **242**(4880): 899-907.
- Aiyar S. E., Helmann J. D. and deHaseth P. L. (1994). "A mismatch bubble in double-stranded DNA suffices to direct precise transcription initiation by Escherichia coli RNA polymerase." *Journal of Biological Chemistry* **269**(18): 13179-13184.
- Anderson C. F. and Record M. T. J. (1982). "Polyelectrolyte theories and their applications to DNA." *Annual Reviews in Physical Chemistry* **33**: 191-222.
- Anderson J. E., Ptashne M. and Harrison S. C. (1987). "Structure of the repressor-operator complex of bacteriophage 434." *Nature* **326**(6116): 846-852.
- Barkley M. D. (1981). "Salt dependence of the kinetics of the lac repressor-operator interaction: role of nonoperator deoxyribonucleic acid in the association reaction." *Biochemistry* **20**(13): 3833-3842.
- Barkley M. D., Lewis P. A. and Sullivan G. E. (1981). "Ion effects on the lac repressor-operator equilibrium." *Biochemistry* **20**(13): 3842-3851.
- Bass S., Sorrells V. and Youderian P. (1988). "Mutant Trp repressors with new DNA-binding specificities." *Science* **242**(4876): 240-245.
- Beachy P. A. (1990). "A molecular view of the Ultrabithorax homeotic gene of Drosophila." *Trends Genet* **6**(2): 46-51.
- Beachy P. A., Varkey J., Young K. E., von Kessler D. P., Sun B. I. and Ekker S. C. (1993). "Cooperative binding of an Ultrabithorax homeodomain protein to nearby and distant DNA sites." *Mol Cell Biol* **13**(11): 6941-56.

- Berg O. G. and von Hippel P. H. (1985). "Diffusion-controlled macromolecular interactions." *Ann. Rev. Biophys. Biophys. Chem.* **14**: 131-160.
- Berg O. G. and von Hippel P. H. (1988). "Selection of DNA binding sites by regulatory proteins." *Trends In Biochemical Sciences* **13**(6): 207-211.
- Berg O. G., Winter R. B. and von Hippel P. H. (1981). "Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory." *Biochemistry* **20**(24): 6929-6948.
- Berger C., Piubelli L., Haditsch U. and Bosshard H. R. (1998). "Diffusion-controlled DNA recognition by an unfolded, monomeric bZIP transcription factor." *FEBS Lett* **425**(1): 14-8.
- Berglund H., Baumann H., Knapp S., Ladenstein R. and Härd T. (1995). "Flexibility of an Arginine Side Chain at a DNA-Protein Interface." *J. Am. Chem. Soc.* **117**(51): 12883-12884.
- Billeter M. (1996). "Homeodomain-type DNA recognition." *Prog Biophys Mol Biol* **66**(3): 211-25.
- Billeter M., Guntert P., Luginbuhl P. and Wuthrich K. (1993). "Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex." *JMB* **234**(4): 1084-1093.
- Billeter M., Guntert P., Luginbuhl P. and Wuthrich K. (1996). "Hydration and DNA recognition by homeodomains." *Cell* **85**(7): 1057-1065.
- Bram S. and Beeman W. W. (1971). "On the cross-section structure of deoxyribonucleic acid in solution." *Journal of Molecular Biology* **55**(3): 311-324.
- Burz D. S. and Hanes S. D. (2001). "Isolation of mutations that disrupt cooperative DNA binding by the Drosophila bicoid protein." *J Mol Biol* **305**(2): 219-30.
- Burz D. S., Rivera-Pomar R., Jackle H. and Hanes S. D. (1998). "Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the Drosophila embryo." *EMBO J.* **17**(20): 5998-6009.
- Calladine C. R. and Drew H. R. (1992). *Understanding DNA*. New York, Academic Press.
- Carlson N. G. and Little J. W. (1993). "Highly cooperative DNA binding by the coliphage HK022 repressor." *J Mol Biol* **230**(4): 1108-30.
- Carra J. H. and Privalov P. L. (1997). "Energetics of Folding and DNA Binding of the MATalpha2 Homeodomain." *Biochemistry* **36**(3): 526-535.
- Carrasco A. E., McGinnis W., Gehring W. J. and De Robertis E. M. (1984). "Cloning of an *X. laevis* gene expressed during early embryogenesis coding for a peptide region homologous to Drosophila homeotic genes." *Cell* **37**(2): 409-14.

- Chasman D. I., Flaherty K. M., Sharp P. A. and Kornberg R. D. (1993). "Crystal structure of TATA-binding protein and model for interaction with DNA." *Proc. Natl. Acad. Sci. U S A* **90**(17): 8174-8178.
- Choo Y. and Klug A. (1997). "Physical basis of a protein-DNA recognition code." *Curr Opin Struct Biol.* **7**(1): 117-125.
- Clarke N. D., Beamer L. J., Goldberg H. R., Berkower C. and Pabo C. O. (1991). "The DNA binding arm of lambda repressor: critical contacts from a flexible region." *Science* **254**(5029): 267-270.
- Clegg R. (1992). Fluorescence Resonance Energy Transfer and Nucleic Acids. *Methods in Enzymology*. San Diego, Academic Press. **211**: 353-388.
- Damante G., Tell G., Leonardi A., Fogolari F., Bortolotti N., Di Lauro R. and Formisano S. (1994). "Analysis of the conformation and stability of rat TTF-1 homeodomain by circular dichroism." *FEBS Letters* **354**: 293-296.
- Davis M. E. and McCammon J. A. (1990). "Electrostatics in biomolecular structure and dynamics." *Chem. Rev.* **90**: 509-521.
- de la Mata I., J.L. G., Gonzalez C., Menendez M., Canada J., Jimenez-Barbero J. and Asensio J. L. (2002). "The Impact of R53C Mutation on the Three-Dimensional Structure, Stability, and DNA-Binding Properties of the Human Hesx-1 Homeodomain." *ChemBioChem* **3**: 726-740.
- deHaseth P. H., Lohman T. M. and Record M. T. J. (1977). "Nonspecific interaction of lac repressor with DNA: an association reaction driven by counterion release." *Biochemistry* **16**(22): 4783-4790.
- Deibert M., Grazulis S., Janulaitis A., Siksnys V. and Huber R. (1999). "Crystal structure of MunI restriction endonuclease in complex with cognate DNA at 1.7 Å resolution." *EMBO Journal* **18**(21): 5805-5816.
- Delcourt S. G. and Blake R. D. (1991). "Stacking energies in DNA." *Journal of Biological Chemistry* **266**(23): 15160-15169.
- Dhavan G. M., Crothers D. M., Chance M. R. and Brenowitz M. (2002). "Concerted binding and bending of DNA by Escherichia coli integration host factor1." *Journal of Molecular Biology* **315**(5): 1027-1037.
- Diebold R. J., Rajaram N., Leonard D. A. and Kerppola T. K. (1998). "Molecular basis of cooperative DNA bending and oriented heterodimer binding in the NFAT1-Fos-Jun-ARRE2 complex." *Proc. Natl. Acad. Sci. U S A* **95**(14): 7915-7920.
- Diekmann S. (1987). "Temperature and salt dependence of the gel migration anomaly of curved DNA fragments." *Nucleic Acids Research* **15**(1): 247-265.
- Drew H. R., Weeks J. R. and Travers A. A. (1985). "Negative supercoiling induces spontaneous unwinding of a bacterial promoter." *EMBO Journal* **4**(4): 1025-1032.

- Driever W., Thoma G. and Nusslein-Volhard C. (1989). "Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the bicoid morphogen." *Nature* **340**(6232): 363-7.
- Duboule D. (1994). *Guidebook to Homeobox genes*. Oxford, U.K., Oxford University Press.
- Eimer W. and Pecora R. (1991). "Rotational and translational diffusion of short rodlike molecules in solution: oligonucleotides." *J. Chem. Phys.* **94**(3): 2324-2329.
- Eis P. S. and Millar D. P. (1993). "Conformational distributions of a four-way DNA junction revealed by time-resolved fluorescence resonance energy transfer." *Biochemistry* **32**(50): 13852-13860.
- Ekker S. C., Young K. E., von Kessler D. P. and Beachy P. A. (1991). "Optimal DNA sequence recognition by the Ultrabithorax homeodomain of *Drosophila*." *Embo J.* **10**(5): 1179-86.
- Ellenberger T. E., Brandl C. J., Struhl K. and Harrison S. C. (1992). "The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex." *Cell* **71**(7): 1223-1237.
- Ephrussi A. and St Johnston D. (2004). "Seeing is believing: the bicoid morphogen gradient matures." *Cell* **116**(2): 143-52.
- Fernandes M. X., Ortega A., Lopez Martinez M. C. and Garcia de la Torre J. (2002). "Calculation of hydrodynamic properties of small nucleic acids from their atomic structure." *Nucleic Acids Res* **30**(8): 1782-8.
- Flader W., Wellenzohn B., Winger R. H., Hallbrucker A., Mayer E. and Liedl K. R. (2003). "Stepwise induced fit in the pico- to nanosecond time scale governs the complexation of the even-skipped transcriptional repressor homeodomain to DNA." *Biopolymers* **68**(2): 139-49.
- Fogolari F., Elcock A. H., Esposito G., Viglino P., Briggs J. M. and McCammon J. A. (1997). "Electrostatic effects in homeodomain-DNA interactions." *J Mol Biol* **267**(2): 368-81.
- Förster T. (1946). "Energiewanderung und Fluoreszenz." *Naturwissenschaften* **6**: 166-175.
- Fraenkel E., Rould M. A., Chambers K. A. and Pabo C. O. (1998). "Engrailed homeodomain-DNA complex at 2.2 Å resolution: a detailed view of the interface and comparison with other engrailed structures." *J Mol Biol* **284**(2): 351-61.
- Frankel A. D. and Kim P. S. (1991). "Modular Structure of Transcription Factors: Implications For Gene Regulation." *Cell* **65**: 717-719.
- Garcia de la Torre J., Huertas M. L. and Carrasco B. (2000). "Calculation of Hydrodynamic Properties of Globular Proteins from Their Atomic-Level Structure." *Biophys. J.* **78**(2): 719-730.

- Garcia de la Torre J., Huertas M. L. and Carrasco B. (2000). "HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations." *Journal of Magnetic Resonance* **147**(1): 138-146.
- Gehring W. J. (1987). "Homeo boxes in the study of development." *Science* **236**(4806): 1245-52.
- Gehring W. J., Affolter M. and Burglin T. (1994). "Homeodomain proteins." *Annu Rev Biochem* **63**: 487-526.
- Gill S. C. and von Hippel P. H. (1989). "Calculation of protein extinction coefficients from amino acid sequence data." *Anal Biochem* **182**(2): 319-26.
- Göhler T., Reimann M., Cherny D., Walter K., Warnecke G., Kim E. and Deppert W. (2002). "Specific Interaction of p53 with Target Binding Sites Is Determined by DNA Conformation and Is Regulated by the C-terminal Domain." *Journal of Biological Chemistry* **277**(43): 41192-41203.
- Gray D. M. (1996). Circular Dichroism of Protein-Nucleic Acid Interactions. *Circular Dichroism and the Conformational Analysis of Biomolecules*. F. G.D. New York, Plenum Press: 469-500.
- Ha J. H. (1990). Thermodynamics of the Interaction of Lactose Repressor with Symmetric Operator (Hydrophobic Effect). Madison, University of Wisconsin-Madison.
- Ha J. H., Spolar R. S. and Record M. T. J. (1989). "Role of the hydrophobic effect in stability of site-specific protein-DNA complexes." *J. Mol. Biol.* **209**(4): 801-816.
- Hagerman P. J. (1988). "Flexibility of DNA." *Annual Reviews of Biophysics and Biophysical Chemistry* **17**: 265-286.
- Hagerman P. J. (1990). "Sequence-directed curvature of DNA." *Annual Reviews of Biochemistry* **59**: 755-781.
- Hard T., Kellenbach E., Boelens R., Maler B. A., Dahlman K., Freedman L. P., Carlstedt-Duke J., Yamamoto K. R., Gustafsson J. A. and Kaptein R. (1990). "Solution structure of the glucocorticoid receptor DNA-binding domain." *Science* **249**(4965): 157-160.
- Hillisch A. (1998). Computer aided design and structure verification of single- and multiple-bulge DNA molecules. Vienna, University of Vienna.
- Hirsch J. A. and Aggarwal A. K. (1995). "Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity." *EMBO Journal* **14**(24): 6280-6291.
- Huang R. Y. and Kowalski D. (1993). "DNA unwinding element and ARS consensus comprise a replication origin within a yeast chromosome." *EMBO Journal* **12**(12): 4521-4531.

- Hunter C. A. (1993). "Sequence-dependent DNA structure: the role of base stacking interactions." *Journal of Molecular Biology* **230**(3): 1025-1054.
- Jacob F. and Monod J. (1961). "Genetic regulatory mechanisms in the synthesis of proteins." *J Mol Biol* **3**: 318-56.
- Janody F., Sturny R., Catala F., Desplan C. and Dostatni N. (2000). "Phosphorylation of bicoid on MAP-kinase sites: contribution to its interaction with the torso pathway." *Development* **127**(2): 279-89.
- Jares-Erijman E. A. and Jovin T. M. (1996). "Determination of DNA helical handedness by fluorescence resonance energy transfer." *J Mol Biol* **257**(3): 597-617.
- Jares-Erijman E. A. and Jovin T. M. (2003). "FRET imaging." *Nat Biotechnol* **21**(11): 1387-95.
- Joachimiak A., Haran T. E. and Sigler P. B. (1994). "Mutagenesis supports water mediated recognition in the trp repressor-operator system." *EMBO Journal* **13**(2): 367-372.
- Kamadurai H., B., Subramaniam S., Jones R. B., Green-Church K. B. and Foster M. P. (2003). "Protein folding coupled to DNA binding in the catalytic domain of bacteriophage lambda integrase detected by mass spectrometry." *Protein Science* **12**(3): 620-626.
- Karlsruhe C., Botuyan M. V. and Gorenstein D. G. (1992). "31P NMR spectra of oligodeoxyribonucleotide duplex lac operator-repressor headpiece complexes: importance of phosphate ester backbone flexibility in protein-DNA recognition." *Biochemistry* **31**(6): 1849-1858.
- Kerppola T. K. (2002). Protein-DNA Interactions: Structure and Energetics. *Encyclopedia of Life Sciences*, Macmillan Publishers Ltd.
- Kerppola T. K. and Curran T. (1991). "DNA bending by Fos and Jun: the flexible hinge model." *Science* **254**(5035): 1210-1214.
- Kim E., Albrechtsen N. and Deppert W. (1997). "DNA-conformation is an important determinant of sequence-specific DNA binding by tumor suppressor p53." *Oncogene* **15**(7): 857-869.
- Kim E., Rohaly G., Heinrichs S., Gimnopoulos D., Meißner H. and Deppert W. (1999). "Influence of promoter DNA topology on sequence-specific DNA binding and transactivation by tumor suppressor p53." *Oncogene* **18**(51): 7310-7318.
- Kim J. L., Nikolov D. B. and Burley S. K. (1993). "Co-crystal structure of TBP recognizing the minor groove of TATA element." *Nature* **365**(6446): 520-527.
- Kim Y., Geiger J. H., Hahn S. and Sigler P. B. (1993). "Crystal structure of yeast TBP/TATA-box complex." *Nature* **365**(6446): 512-520.

- Kissinger C. R., Liu B. S., Martin-Blanco E., Kornberg T. B. and Pabo C. O. (1990). "Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions." *Cell* **63**(3): 579-90.
- Kono H. and Sarai A. (1999). "Structure-based prediction of DNA target sites by regulatory proteins." *Proteins* **35**(1): 114-131.
- Kuhn A., Voit R., Stefanovsky V., Evers R., Bianchi M. and Grummt I. (1994). "Functional differences between the two splice variants of the nucleolar transcription factor UBF: the second HMG box determines specificity of DNA binding and transcriptional activity." *EMBO Journal* **13**(2): 416-424.
- Ladbury J. E., Wright J. G., Sturtevant J. M. and Sigler P. B. (1994). "A thermodynamic study of the trp repressor-operator interaction." *Journal of Molecular Biology* **238**(5): 669-681.
- Lagunavicius A., Grazulis S., Balciunaite E., Vainius D. and Siksnyys V. (1997). "DNA binding specificity of MunI restriction endonuclease is controlled by pH and calcium ions: involvement of active site carboxylate residues." *Biochemistry* **36**(37): 11093-11099.
- Lee S., Cavallo L. and Griffith J. (1997). "Human p53 binds Holliday junctions strongly and facilitates their cleavage." *Journal of Biological Chemistry* **272**(11): 7532-7539.
- Leonard D. A. and Kerppola T. K. (1998). "DNA bending determines Fos-Jun heterodimer orientation." *Nature Structural Biology* **5**(10): 877-881.
- Leonard D. A., Rjaram N. and Kerppola T. K. (1997). "Structural basis of DNA bending and oriented heterodimer binding by Fos and Jun." *Proc. Natl. Acad. Sci. U S A* **94**(10): 4913-4918.
- Lewis M., Chang G., Horton N. C., Kercher M. A., Pace H. C., Schumacher M. A., Brennan R. G. and Lu P. (1996). "Crystal structure of the lactose operon repressor and its complexes with DNA and inducer." *Science* **271**(5253): 1247-1254.
- Li T., Stark M. R., Johnson A. D. and Wolberger C. (1995). "Crystal structure of the MATA1/MAT alpha 2 homeodomain heterodimer bound to DNA." *Science* **270**(5234): 262-9.
- Lohman T. M. (1986). "Kinetics of protein-nucleic acid interactions: use of salt effects to probe mechanisms of interaction." *CRC Crit Rev Biochem* **19**(3): 191-245.
- Lohman T. M. and Bujalowski W. (1991). Thermodynamic methods for model-independent determination of equilibrium binding isotherms for protein-DNA interactions: spectroscopic approaches to monitor binding. *Methods in Enzymology*. R. T. Sauer. San Diego, Academic Press, Inc. **208**: 259-290.
- Lohman T. M., deHaseth P. H. and Record M. T. J. (1980). "Pentalysine-deoxyribonucleic acid interactions: a model for the general effects of ion concentrations on the interactions of proteins with nucleic acids." *Biochemistry* **19**(15): 3522-3530.

- Lohman T. M., deHaseth P. L. and Record M. T. J. (1978). "Analysis of ion concentration effects of the kinetics of protein-nucleic acid interactions. Application to lac repressor-operator interactions." *Biophys. Chem.* **8**(4): 281-294.
- Lohman T. M. and Mascotti D. P. (1992). "Thermodynamics of ligand-nucleic acid interactions." *Methods in Enzymology* **212**: 400-424.
- Luscombe N. M., Austin S. E., Berman H. M. and Thornton J. M. (2000). "An overview of the structures of protein-DNA complexes." *Genome Biology* **1**(1): REVIEWS001.1-REVIEWS001.37.
- Luscombe N. M., Laskowski R. A. and Thornton J. M. (2001). "Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level." *Nucleic Acids Res.* **29**(13): 2860-2874.
- Mak A. and Johnson A. D. (1993). "The carboxy-terminal tail of the homeodomain protein alpha 2 is required for function with a second homeodomain protein." *Genes Dev* **7**(10): 1862-70.
- Mandel-Gutfreund Y., Schueler O. and Margalit H. (1995). "Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles." *J. Mol. Biol.* **253**(2): 370-382.
- Manning G. S. (1969). "Limiting laws and counterion condensation in polyelectrolyte solutions. I. Colligative properties." *J. Chem. Phys.* **51**(3): 924-933.
- Manning G. S. (1978). "The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides." *Quart. Rev. Biophys.* **11**(2): 179-246.
- Mascotti D. P. and Lohman T. M. (1990). "Thermodynamic extent of counterion release upon binding oligolysines to single-stranded nucleic acids." *Proc. Natl. Acad. Sci. U S A* **87**(8): 3142-3146.
- McCammon J. A. and Harvey S. C. (1987). *Dynamics of proteins and nucleic acids*. Cambridge, Cambridge University Press.
- McGinnis W., Levine M. S., Hafen E., Kuroiwa A. and Gehring W. J. (1984). "A conserved DNA sequence in homoeotic genes of the *Drosophila* Antennapedia and bithorax complexes." *Nature* **308**(5958): 428-33.
- Nanda V. and Brand L. (2000). "Aromatic interactions in homeodomains contribute to the low quantum yield of a conserved, buried tryptophan." *Proteins* **40**(1): 112-125.
- Nelson H. B. and Laughon A. (1990). "The DNA binding specificity of the *Drosophila* fushi tarazu protein: a possible role for DNA bending in homeodomain recognition." *New Biol* **2**(2): 171-8.

- Newman M., Strzelecka T., Dorner L. F., Schildkraut I. and Aggarwal A. K. (1995). "Structure of Bam HI endonuclease bound to DNA: partial folding and unfolding on DNA binding." *Science* **269**(5224): 656-663.
- Niessing D., Driever W., Sprenger F., Taubert H., Jackle H. and Rivera-Pomar R. (2000). "Homeodomain position 54 specifies transcriptional versus translational control by Bicoid." *Mol Cell* **5**(2): 395-401.
- Norman D. G., Grainger R., Uhrin D. and Lilley D. M. (2000). "Location of cyanine-3 on double-stranded DNA: importance for fluorescence resonance energy transfer studies." *Biochemistry* **39**(21): 6317-6324.
- Noyes R. M. (1961). Effects of diffusion rates on chemical kinetics. *Progress in Reaction Kinetics*. G. Porter. New York, Pergamon Press. **1**: 129-160.
- O'Shea E. K., Klemm J. D., Kim P. S. and Alber T. (1991). "X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil." *Science* **254**(5031): 539-544.
- Otting G., Qian Y. Q., Billeter M., Müller M., Affolter M., Gehring W. J. and Wüthrich K. (1990). "Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution." *EMBO Journal* **9**(10): 3085-3092.
- Otwinowski Z., Schevitz R. W., Zhang R. G., Lawson C. L., Joachimiak A., Marmorstein R. Q., Luisi B. F. and Sigler P. B. (1988). "Crystal structure of trp repressor/operator complex at atomic resolution. [erratum appears in Nature 1988 Oct 27;335(6193):837.]." *Nature* **242**(6188): 321-329.
- Pabo C. O., Aggarwal A. K., Jordan S. R., Beamer L. J., Obeyesekere U. R. and Harrison S. C. (1990). "Conserved residues make similar contacts in two repressor-operator complexes." *Science* **247**(4947): 1210-1213.
- Pabo C. O. and Nekludova L. (2000). "Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition?" *J. Mol. Biol.* **301**(3): 597-624.
- Pabo C. O. and Sauer R. T. (1992). "Transcription factors: structural families and principles of DNA recognition." *Annu. Rev. Biochem.* **61**: 1053-1095.
- Pankratz M. J. and Jackle H. (1993). Blastoderm segmentation. *The development of Drosophila melanogaster*. M. Bate and A. Martinez Arias. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press: 467-516.
- Parkhurst L. J., Parkhurst K. M., Powell R., Wu J. and Williams S. (2001). "Time-resolved fluorescence resonance energy transfer studies of DNA bending in double-stranded oligonucleotides and in DNA-protein complexes." *Biopolymers* **61**(3): 180-200.
- Passner J. M., Ryoo H. D., Shen L., Mann R. S. and Aggarwal A. K. (1999). "Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex." *Nature* **397**(6721): 714-9.

- Patel L., Abate C. and Curran T. (1990). "Altered protein conformation on DNA binding by Fos and Jun." *Nature* **347**(6293): 572-575.
- Patrick S. M. and Turchi J. J. (2001). "Stopped-flow Kinetic Analysis of Replication Protein A-binding DNA. Damage recognition and affinity for single-stranded DNA reveal differential contributions of kon and koff rate constants." *J. Biol. Chem.* **276**(25): 22630-22637.
- Pavletich N. P. and Pabo C. O. (1991). "Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å." *Science* **252**(5007): 809-817.
- Payet D., Hillisch A., Lowe N., Diekmann S. and Travers A. (1999). "The Recognition of Distorted DNA Structures by HMG-D: A Footprinting and Molecular Modelling Study." *Journal of Molecular Biology* **294**(1): 79-91.
- Peitsch M. C. (1996). "ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling." *Biochem Soc Trans* **24**(1): 274-9.
- Peltenburg L. T. and Murre C. (1996). "Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx." *Embo J* **15**(13): 3385-93.
- Percival-Smith A., Muller M., Affolter M. and Gehring W. J. (1990). "The interaction with DNA of wild-type and mutant fushi tarazu homeodomains." *Embo J* **9**(12): 3967-74.
- Perutz M. F. (1990). *Mechanism of Cooperativity and Allosteric Regulation in Proteins*. Cambridge, Cambridge University Press.
- Petersen J. M., Skalicky J. J., Donaldson L. W., McIntosh L. P., Alber T. and Graves B. J. (1995). "Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an alpha helix." *Science* **269**(5232): 1866-1869.
- Petri V., Hsieh M. and Brenowitz M. (1995). "Thermodynamic and kinetic characterization of the binding of the TATA binding protein to the adenovirus E4 promoter." *Biochemistry* **34**(31): 9977-9984.
- Phillips C. L., Stark M. R., Johnson A. D. and Dahlquist F. W. (1994). "Heterodimerization of the yeast homeodomain transcriptional regulators alpha 2 and a1 induces an interfacial helix in alpha 2." *Biochemistry* **33**(31): 9294-302.
- Pick L., Schier A., Affolter M., Schmidt-Glenewinkel T. and Gehring W. J. (1990). "Analysis of the ftz upstream element: germ layer-specific enhancers are independently autoregulated." *Genes Dev* **4**(7): 1224-39.
- Pil P. and Lippard S. J. (1992). "Specific Binding of Chromosomal Protein HMG1 to DNA Damaged by the Anticancer Drug Cisplatin." *Science* **256**(5054): 234-237.
- Ponder J. W. and Case D. A. (2003). "Force fields for protein simulations." *Adv. Prot. Chem.* **66**: 27-85.

- Press W. H., Flannery B. P., Teukolsky S. A. and Vetterling W. T. (1992). *Numerical recipes in C. The art of scientific computing*. Cambridge, Cambridge University Press.
- Ptashne M. (1986). *A Genetic Switch. Phage lambda and Higher Organisms*. Cambridge, MA, Blackwell Scientific and Cell Press.
- Puglisi J. D. and Tinoco I., Jr. (1989). Absorbance melting curves of RNA. *Methods in Enzymology*. J. E. Dahlberg and J. N. Abelson. San Diego, Academic Press, Inc. **180**: 304-325.
- Qian Y. Q., Billeter M., Otting G., Muller M. and Gehring W. J. (1989). "The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors." *Cell* **59**(3): 573-580.
- Record M. T., Jr., Anderson C. F. and Lohman T. M. (1978). "Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity." *Quarterly Reviews of Biophysics* **11**(2): 103-178.
- Record M. T., Jr., Ha J. H. and Fisher M. A. (1991). "Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and DNA." *Methods in Enzymology* **208**: 291-343.
- Record M. T., Jr., Lohman M. L. and De Haseth P. (1976). "Ion effects on ligand-nucleic acid interactions." *Journal of Molecular Biology* **107**(2): 145-158.
- Reddy C. K., Das A. and Jayaram B. (2001). "Do water molecules mediate protein-DNA recognition?" *JMB* **314**(3): 619-632.
- Richter P. and Eigen M. (1974). "Diffusion controlled reaction rates in spheroidal geometry. Application to repressor-operator association and membrane bound enzymes." *Biophys. Chem.* **2**(3): 255-263.
- Riggs A. D., Bourgeois S. and Cohn M. (1970). "The lac repressor-operator interaction. 3. Kinetic studies." *J. Mol. Biol.* **53**(3): 401-417.
- Rigler R., Mets Ü. and Widengren J. (1993). "Fluorescence correlation spectroscopy with high count rate and low background: Analysis of translational diffusion." *Eur. Biophys. J.* **22**: 169-175.
- Rivera-Pomar R., Lu X., Perrimon N., Taubert H. and Jackle H. (1995). "Activation of posterior gap gene expression in the Drosophila blastoderm." *Nature* **376**(6537): 253-6.
- Rivera-Pomar R., Niessing D., Schmidt-Ott U., Gehring W. J. and Jackle H. (1996). "RNA binding and translational suppression by bicoid." *Nature* **379**(6567): 746-9.

- Ronchi E., Treisman J., Dostatni N., Struhl G. and Desplan C. (1993). "Down-regulation of the *Drosophila* morphogen bicoid by the torso receptor-mediated signal transduction cascade." *Cell* **74**(2): 347-55.
- Sambrook J., Fritsch E. F. and Maniatis T. (1989). *Molecular cloning. A laboratory manual*. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.
- Sarai A., Mazur J., Nussinov R. and Jernigan R. L. (1989). "Sequence dependence of DNA conformational flexibility." *Biochemistry* **28**(19): 7842-7849.
- Sauer R. T., Smith D. L. and Johnson A. D. (1988). "Flexibility of the yeast alpha 2 repressor enables it to occupy the ends of its operator, leaving the center free." *Genes Dev.* **2**(7): 807-816.
- Schier A. F. and Gehring W. J. (1993). "Functional specificity of the homeodomain protein fushi tarazu: the role of DNA-binding specificity in vivo." *Proc. Natl. Acad. Sci. U S A* **90**(4): 1450-4.
- Schreiber G. and Fersht A. R. (1996). "Rapid, electrostatically assisted association of proteins." *Nat Struct Biol* **3**(5): 427-31.
- Schwabe J., Neuhaus D. and Rhodes D. (1990). "Solution structure of the DNA-binding domain of the oestrogen receptor." *Nature* **348**(6300): 458-461.
- Schwabe J. W. R., Chapman L., Finch J. T., Rhodes D. and Neuhaus D. (1993). "DNA recognition by the estrogen receptor: from solution to the crystal." *Structure* **1**: 187-204.
- Schwede T., Kopp J., Guex N. and Peitsch M. C. (2003). "SWISS-MODEL: an automated protein homology-modeling server." *Nucl. Acids. Res.* **31**(13): 3381-3385.
- Schwille P., Meyer-Almes F. J. and Rigler R. (1997). "Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution." *Biophys J.*, **72**(4): 1878-86.
- Scott M. P., Tamkun J. W. and Hartzell G. W., 3rd (1989). "The structure and function of the homeodomain." *Biochim Biophys Acta* **989**(1): 25-48.
- Seeman N. C., Rosenberg J. M. and Rich A. (1976). "Sequence-specific recognition of double helical nucleic acids by proteins." *Proc. Natl. Acad. Sci. U S A* **3**(73): 804-808.
- Shakkeed Z., Guzikevich-Guerstein G., Frolow F., Rabinovich D., Joachimiak A. and Sigler P. B. (1994). "Determinants of repressor/operator recognition from the structure of the trp operator binding site." *Nature* **368**(6470): 459-473.
- Sharp K. A. and Honig B. (1990). "Calculating total electrostatic energies with the non-linear Poisson-Boltzmann equation." *J. Phys. Chem.* **94**: 7684-7692.

- Shimamoto N. and Utiyama H. (1983). "Mechanism and role of cooperative binding of bacteriophage fd gene 5 protein to single-stranded deoxyribonucleic acid." *Biochemistry* **22**(25): 5869-78.
- Shliakhtenko L. S., Lyubchenko I. L., Chernov B. K. and Zhurkin V. B. (1990). "Influence of temperature and ionic strength on electrophoretic mobility of synthetic DNA fragments." *Molecular Biology USSR* **24**(1): 79-95.
- Siksny V., Skirgaila R., Sasnauskas G., Urbanke C., Cherny D., Grazulis S. and Huber R. (1999). "The Cfr10I restriction enzyme is functional as a tetramer." *Journal of Molecular Biology* **291**(5): 1105-1118.
- Spolar R. S. and Record M. T., Jr. (1994). "Coupling of local folding to site-specific binding of proteins to DNA." *Science* **263**(5148): 777-784.
- Struhl G., Struhl K. and Macdonald P. M. (1989). "The gradient morphogen bicoid is a concentration-dependent transcriptional activator." *Cell* **57**(7): 1259-73.
- Stühmeier F., Hillisch A., Clegg R. M. and Diekmann S. (2000). Practical aspects of fluorescence resonance energy transfer (FRET) and its applications in nucleic acid biochemistry. *DNA-Protein Interactions*. B. M. Travers A. Oxford, Oxford University Press: 77-94.
- Sturtevant J. M. (1977). "Heat capacity and entropy changes in processes involving proteins." *Proc. Natl. Acad. Sci. U S A* **74**(6): 2236-2240.
- Subramaniam V., Jovin T. M. and Rivera-Pomar R. V. (2001). "Aromatic amino acids are critical for stability of the bicoid homeodomain." *J. Biol. Chem.* **6**(24): 21506-21511.
- Suzuki M. and Yagi N. (1994). "DNA recognition code of transcription factors in the helix-turn-helix, probe helix, hormone receptor, and zinc finger families." *Proc. Natl. Acad. Sci. U S A* **91**(26): 12357-12361.
- Talanian R. V., McKnight C. J. and Kim P. S. (1990). "Sequence-specific DNA binding by a short peptide dimer." *Science* **249**(4970): 769-771.
- Tonomura B., Nakatani H., Ohnishi M., Yamaguchi-Ito J. and Hiromi K. (1978). "Test reactions for a stopped-flow apparatus. Reduction of 2,6-dichlorophenolindophenol and potassium ferricyanide by L-ascorbic acid." *Anal Biochem* **84**(2): 370-83.
- Travers A. A. (1995). DNA bending by sequence and proteins. *DNA-Protein: Structural Interactions*. L. D.M.J. New York, Oxford University Press: 49-75.
- Tsao D. H., Gruschus J. M., Wang L. H., Nirenberg M. and Ferretti J. A. (1994). "Elongation of helix III of the NK-2 homeodomain upon binding to DNA: a secondary structure study by NMR." *Biochemistry* **33**(50): 15053-60.
- Tucker-Kellogg L., Rould M. A., Chambers K. A., Ades S. E., Sauer R. T. and Pabo C. O. (1997). "Engrailed (Gln50-->Lys) homeodomain-DNA complex at 1.9 Å resolution:

- structural basis for enhanced affinity and altered specificity." *Structure* **5**(8): 1047-54.
- van Dijk M. A., Peltenburg L. T. and Murre C. (1995). "Hox gene products modulate the DNA binding activity of Pbx1 and Pbx2." *Mech Dev* **52**(1): 99-108.
- von Hippel P. H. and Berg O. G. (1989). "Facilitated target location in biological systems." *J. Biol. Chem.* **264**(2): 675-678.
- von Hippel P. H. and McGhee J. D. (1972). "DNA-protein interactions." *Annu Rev Biochem.* **41**(10): 231-300.
- von Smoluchowski M. (1917). "Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen." *Z. Phys. Chem. Leipzig* **92**: 129-168.
- Watson J. D. and Crick F. H. C. (1953). "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid." *Nature* **171**(4356): 737-738.
- Weiss M. A., Ellenberger T., Wobbe C. R., Lee J. P., Harrison S. C. and Struhl K. (1990). "Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA." *Nature* **347**(6293): 575-578.
- Widengren J., Mets Ü. and Rigler R. (1995). "Fluorescence correlation spectroscopy of triplet states in solution: A theoretical and experimental study." *J. Phys. Chem.* **99**: 13368-13379.
- Wilson D., Sheng G., Lecuit T., Dostatni N. and Desplan C. (1993). "Cooperative dimerization of paired class homeo domains on DNA." *Genes Dev* **7**(11): 2120-34.
- Wilson D. S. and Desplan C. (1995). "Homeodomain proteins. Cooperating to be different." *Curr Biol* **5**(1): 32-34.
- Wilson D. S., Guenther B., Desplan C. and Kuriyan J. (1995). "High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA." *Cell* **82**(5): 709-19.
- Wilson D. S., Sheng G., S. J. and C. D. (1996). "Conservation and diversification in homeodomain-DNA interactions: a comparative genetic analysis." *Proc. Natl. Acad. Sci. USA* **93**(14): 6886-6891.
- Winter R. B., Berg O. G. and von Hippel P. H. (1981). "Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor-operator interaction: kinetic measurements and conclusions." *Biochemistry* **20**(24): 6961-6977.
- Wolberger C. (1996). "Homeodomain interactions." *Curr Opin Struct Biol* **6**(1): 62-68.
- Wu P. and Brand L. (1994). "Resonance energy transfer: methods and applications." *Anal Biochem* **218**(1): 1-13.

Acknowledgements

I would like to express my deep gratitude to my supervisor and head of the Molecular Biology Department of Max Planck Institute for Biophysical Chemistry Dr. T.M. Jovin for providing excellent experimental conditions, for his care and many of fruitful discussions about this work, for the opportunity to learn the “kinetic way of thinking” and to gain an immense experience in experimental methods of molecular biophysics.

I give my sincere thanks to Prof. Dr. E. Neher for his supervision of this external dissertation on behalf of the Faculty of Physics of the University of Göttingen. I thank also to Prof. Dr. W. Lauterborn for serving on the dissertation committee.

I am deeply grateful to Dr. R. Rivera-Pomar for his guidance in developmental biology - the world of flies, gradients and homeodomains, for teaching me the secrets of work with RNA and for his ability to find and show the positive sides of results in difficult moments of my research.

My special thanks go to Dr. D. Cherny for his constructive criticism and lots of valuable suggestions writing this manuscript, as well for many interesting discussions which enriched my knowledge about protein-DNA interactions and helped me not to forget my Russian.

I wish to express my thanks to Dr. R. Heintzmann for his substantial contribution in the development of programs for kinetic data analysis. I thank also R. Klement for his help in molecular modeling and solving PC related problems.

I want to express my gratitude to S. Höppner for revising the English of this manuscript and for his help with many questions regarding English and German.

I am very grateful to all the technical assistants of our department for their invaluable help. In particular I wish to thank G. Heim for teaching me the tricks for working with DNA, A. Zechel for her help in elaborative work of protein expression and purification, A. Dichter for his expertise in methods of molecular biology, H. Richmann for his help in solving technical problems. I would also like to thank J. Jaitner from precision mechanics workshop, W.Sauermann from the optical workshop and S. Leineweber and K.-H. Knauber from the electronic workshop, without whose help the detection of some data presented in this work would otherwise not have been possible.

I wish to thank Dr. D. Jovin, Dr. A. Schyolkina, Prof. Dr. G. Barisas and Dr. K. Zechel for valuable suggestions and Dr. S. Pabst and Dr. D. Fasshauer for fruitful collaboration.

I would like to thank all the past and present members of the molecular biology department for warm and friendly working atmosphere. I thank to all my friends whom I met during my study years in Göttingen and shared a good time.

My deepest and warmest thanks go to my wife Ruta for her love, motivation and support sharing those happy and sad moments over these years.

Finally, I would like to thank my parents for their permanent support, patience and believing in me throughout those long years of my education.

This work was supported by Konrad Adenauer Stiftung and Max Planck Gesellschaft.

Lebenslauf

Persönliche Daten

Darius Vainius
geb. am 26.08.1972 in Sakiai, Litauen
Staatsangehörigkeit Litauisch

Schulbildung

1979 – 1982 Grundschole in Panoviai, Sakiu raj., Litauen
1982 – 1987 Hauptschole in Panoviai, Sakiu raj., Litauen
1987 – 1990 Mittelschole in Kudirkos Naumiestic, Litauen
16.06.1990 Abschluß: Abitur
Gesamtnote: sehr gut.

Studium

1990 – 1995 Studium der Physik an der Fakultät für Physik an der
Universität Vilnius, Litauen.

01.07.1995 Abschluß: Diplom. Gesamtnote: sehr gut.
Thema der Diplomarbeit: „Untersuchung der
Wechselwirkung zwischen der Restriktionsendonuklease
MunI und DNS mit der Methode der
Fluoreszenzspektroskopie“

1995 – 1997 Magisterstudium der angewandten Physik (Schwerpunkt
Biophysik) an der Fakultät für Physik an der Universität
Vilnius, Litauen.

01.07.1997 Abschluß: Magister artium. Gesamtnote: sehr gut.
Thema der Magisterarbeit: „Untersuchung der Mutanten der
Restriktionsendonuklease *MunI* mit Hilfe der Methode der
Fluoreszenzspektroskopie“

Promotion

1997 – 2004 In Biophysik an der Universität Göttingen, Fakultät für
Physik. Durchführung am Max-Planck-Institut für
biophysikalische Chemie Göttingen, Abt. Molekulare
Biologie.

Veröffentlichungen

Lagunavicius A., Grazulis S., Balciunaite E., Vainius D., and Siksnyis V.,
“DNA binding specificity of *MunI* restriction endonuclease is controlled
by pH and calcium ions: involvement of active site carboxylate
residues”, (1997), *Biochemistry*, 36: 11093-11099;

Pabst S., Margittai M., Vainius D., Langen R., Jahn R., and
Fasshauer D., „Rapid and selective binding to the synaptic SNARE
complex suggests a modulatory role of complexins in neuroexocytosis”,
(2002), *Journal of Biological Chemistry*, 277: 7838-7848.